Contents

Preface xv

Ezra Miller and Vic Reiner
What is Geometric Combinatorics?
An Overview of the Graduate Summer School 1

Bibliography 17

Alexander Barvinok
Lattice Points, Polyhedra, and Complexity 19

Introduction 21

Lecture 1. Inspirational Examples. Valuations 23
Valuations 26
Problems 27

Lecture 2. Identities in the Algebra of Polyhedra 29
Problems 34

Lecture 3. Generating Functions and Cones. Continued Fractions 39
Generating Functions and Cones 39
Continued Fractions 42
Computing $f(K, \mathbf{x})$ for 2-dimensional Cones 43
The Computational Complexity 44
Problems 45

Lecture 4. Rational Polyhedra and Rational Functions 47
Problems 51

Lecture 5. Computing Generating Functions Fast 53
Why Do We Need Generating Functions? 53
What "Fast" and "Short" Means 54
Problems 57
Concluding Remarks 58

Bibliography 61

Sergey Fomin and Nathan Reading
Root Systems and Generalized Associahedra 63

Lecture 1. Reflections and Roots 67
1.1. The Pentagon Recurrence 67
1.2. Reflection Groups 68
1.3. Symmetries of Regular Polytopes 70
1.4. Root Systems 73
1.5. Root Systems of Types A, B, C, and D 75

Lecture 2. Dynkin Diagrams and Coxeter Groups 77
2.1. Finite Type Classification 77
2.2. Coxeter Groups 79
2.3. Other "Finite Type" Classifications 80
2.4. Reduced Words and Permutohedra 82
2.5. Coxeter Element and Coxeter Number 84

Lecture 3. Associahedra and Mutations 87
3.1. Associahedron 87
3.2. Cyclohedron 93
3.3. Matrix Mutations 95
3.4. Exchange Relations 96

Lecture 4. Cluster Algebras 101
4.1. Seeds and Clusters 101
4.2. Finite Type Classification 104
4.3. Cluster Complexes and Generalized Associahedra 106
4.4. Polytopal Realizations of Generalized Associahedra 108
4.5. Double Wiring Diagrams and Double Bruhat Cells 111

Lecture 5. Enumerative Problems 115
5.1. Catalan Combinatorics of Arbitrary Type 115
5.2. Generalized Narayana Numbers 120
5.3. Non-crystallographic Types 124
5.4. Lattice Congruences and the Weak Order 125

Bibliography 129

Robin Forman
Topics in Combinatorial Differential Topology and Geometry 133

Lecture 1. Discrete Morse Theory 137
1. Introduction 137
2. Cell Complexes and CW Complexes 138
3. The Morse Theory 143
4. A More Combinatorial Language 146
5. Our First Example: The Real Projective Plane 147
6. Sphere Theorems 148
7. Our Second Example 148
8. Exercises for Lecture 1 151

Lecture 2. Discrete Morse Theory, continued 153
1. Suspensions and Discrete Morse Theory 153
2. Monotone Graph Properties 155
3. The Morse Complex 161
4. Canceling Critical Points 165
5. Exercises for Lecture 2 166

Lecture 3. Discrete Morse Theory and Evasiveness 169
1. The Main Results 169
2. Betti Numbers for General Sets of Faces 175
3. Exercises for Lecture 3 180

Lecture 4. The Charney-Davis Conjectures 181
1. Introduction 181
2. Exercises for Lecture 4 187

Lecture 5. From Analysis to Combinatorics 189
1. Hodge Theory and the Hopf-Charney-Davis Conjectures 189
2. The Charney-Davis Conjecture and the h-vector 194
3. Exercises for Lecture 5 198

Bibliography 201

Mark Haiman and Alexander Woo
Geometry of q and q,t-Analogs in Combinatorial Enumeration 207
Introduction 209

Lecture 1. Kostka Numbers and q-Analogs 211
1.1. Definition of Kostka Numbers 211
1.2. $K_{\lambda\mu}$ in Symmetric Functions 212
1.3. S_n Representations 212
1.4. GL_n Representations 214
1.5. The q-Analog $K_{\lambda\mu}(q)$ 215
1.6. Exercises 216

Lecture 2. Catalan Numbers, Trees, Lagrange Inversion, and their q-Analogs 217
2.1. Catalan Numbers 217
2.2. Rooted Trees 218
2.3. The Lagrange Inversion Formula 219
2.4. q-Analogs 220
2.5. q-Lagrange Inversion 222
2.6. Exercises 226

Lecture 3. Macdonald Polynomials 227
3.1. Symmetric Function Bases and the Involution ω 227
3.2. Plethystic Substitution 228
3.3. The Cauchy Kernel and Hall Inner Product 228
3.4. Dominance Ordering 229
3.5. Definition of Macdonald Polynomials 229
3.6. More Properties of Macdonald Polynomials 232
3.7. Exercises 234

Lecture 4. Connecting Macdonald Polynomials and q-Lagrange Inversion; (q,t)-Analogs 235
4.1. The Operator ∇ and a (q,t)-Analog of $k_n(q)$ 235
4.2. Proof of Theorem 7 236
4.3. First Remarks on Positivity 239
4.4. Exercises 240

Lecture 5. Positivity and Combinatorics? 241
5.1. Representation Theory of $\widetilde{H}_\mu(x;q,t)$ 241
5.2. Representation Theory of ∇e_n 243
5.3. Combinatorics of ∇e_n 243
5.4. Combinatorics of $\widetilde{H}_\mu(x;q,t)$ 245
5.5. Exercises 245

Bibliography 247

Dmitry N. Kozlov
Chromatic Numbers, Morphism Complexes, and Stiefel-Whitney Characteristic Classes 249

Preamble 251

Lecture 1. Introduction 253
1.1. The Chromatic Number of a Graph 253
1.2. The Category of Graphs 256

Lecture 2. The Functor $\mathtt{Hom}\,(-,-)$ 261
2.1. Complexes of Graph Homomorphisms 261
2.2. Morphism Complexes 264
2.3. Historic Detour 267
2.4. More about the $\mathtt{Hom}$-Complexes 269
2.5. Folds 273

Lecture 3. Stiefel-Whitney Classes and First Applications 277
3.1. Elements of the Principal Bundle Theory 277
3.2. Properties of Stiefel-Whitney Classes 279
3.3. First Applications of Stiefel-Whitney Classes to Lower Bounds of Chromatic Numbers of Graphs 281

Lecture 4. The Spectral Sequence Approach 285
4.1. $\mathtt{Hom}_+$-construction 285
4.2. Spectral Sequence Generalities 288
4.3. The Standard Spectral Sequence Converging to $H^*(\mathtt{Hom}_+(T,G))$ 293

Lecture 5. The Proof of the Lovász Conjecture 295
5.1. Formulation of the Conjecture and Sketch of the Proof 295
5.2. Completing the Sketch for the Case k is Odd 297
5.3. Completing the Sketch for the Case k is Even 301

Lecture 6. Summary and Outlook 305
6.1. Homotopy Tests, $\mathbb{Z}_2$-Tests, and Families of Test Graphs 305
6.2. Conclusion and Open Problems 308

Bibliography 311

Robert MacPherson
Equivariant Invariants and Linear Geometry 317

Introduction 319
0.1. Spaces with a Torus Action 320
0.2. Linear Graphs 322
0.3. Rings and Modules 323

Lecture 1. Equivariant Homology and Intersection Homology 327
1.1. Introduction 327
1.2. Simplicial Complexes 328
1.3. Pseudomanifolds 329
1.4. Ordinary Homology Theory 330
1.5. Basic Definitions of Equivariant Topology 332
1.6. Equivariant Homology 333
1.7. Formal Properties of Equivariant Homology 335
1.8. Torus Equivariant Cohomology of a Point 337
1.9. The Equivariant Cohomology of a 2-Sphere 338
1.10. Equivariant Intersection Cohomology 340

Lecture 2. Moment Graphs 343
2.1. Assumptions on the Action of T on X 343
2.2. The Moment Graph 344
2.3. Complex Projective Line and the Line Segment 346
2.4. Projective $(n-1)$-Space and the Simplex 347
2.5. Quadric Hypersurfaces and the Cross-Polytope 348
2.6. Grassmannians and Hypersimplices 351
2.7. The Flag Manifold and the Permutahedron 353
2.8. Toric Varieties and Convex Polyhedra 354
2.9. Moment Maps 357

Lecture 3. The Cohomology of a Linear Graph 359
3.1. The Definition of the Cohomology of a Linear Graph 359
3.2. Interpreting $\mathcal{H}^i(\mathcal{G})$ for Small i 360
3.3. Piecewise Polynomial Functions 361
3.4. Morse Theory 362
3.5. Perfect Morse Functions 364
3.6. Determining $\mathcal{H}^*(\mathcal{G})$ as a $\mathcal{O}(\mathbb{T})$ Module 366
3.7. Poincaré Duality 366
3.8. The Main Theorems 367

Lecture 4. Computing Intersection Homology 371
4.1. Graphs Arising from Reflection Groups 371
4.2. Upward Saturated Subgraphs 372

4.3. Sheaves on Graphs 373
4.4. A Criterion for Perfection 374
4.5. Definition of the Sheaf $\underline{M}$ 375
4.6. The Main Results 377
4.7. Flag Varieties and Generalized Schubert Varieties 377

Lecture 5. Cohomology as Functions on a Variety 379
5.1. The Fixed Point Arrangement 379
5.2. How to Compute the Fixed Point Arrangement 380
5.3. The Main Result 381
5.4. Springer Varieties 382
5.5. Relation with Lecture 3 385

Bibliography 387

Richard P. Stanley
An Introduction to Hyperplane Arrangements 389

Lecture 1. Basic Definitions, the Intersection Poset and the Characteristic Polynomial 391
1.1. Basic Definitions 391
1.2. The Intersection Poset 397
1.3. The Characteristic Polynomial 398
Exercises 400

Lecture 2. Properties of the Intersection Poset and Graphical Arrangements 403
2.1. Properties of the Intersection Poset 403
2.2. The Number of Regions 409
2.3. Graphical Arrangements 414
Exercises 419

Lecture 3. Matroids and Geometric Lattices 421
3.1. Matroids 421
3.2. The Lattice of Flats and Geometric Lattices 423
Exercises 428

Lecture 4. Broken Circuits, Modular Elements, and Supersolvability 431
4.1. Broken Circuits 431
4.2. Modular Elements 437
4.3. Supersolvable Lattices 442
Exercises 446

Lecture 5. Finite Fields 449
5.1. The Finite Field Method 449
5.2. The Shi Arrangement 452
5.3. Exponential Sequences of Arrangements 454
5.4. The Catalan Arrangement 456
5.5. Interval Orders 459
5.6. Intervals with Generic Lengths 466
5.7. Other Examples 467
Exercises 468

Lecture 6. Separating Hyperplanes 475
6.1. The Distance Enumerator 475
6.2. Parking Functions and Tree Inversions 478
6.3. The Distance Enumerator of the Shi Arrangement 483
6.4. The Distance Enumerator of a Supersolvable Arrangement 487
6.5. The Varchenko Matrix 490
Exercises 491

Bibliography 495

Michelle L. Wachs, Poset Topology: Tools and Applications 497

Introduction 499

Lecture 1. Basic Definitions, Results, and Examples 501
1.1. Order Complexes and Face Posets 501
1.2. The Möbius Function 505
1.3. Hyperplane and Subspace Arrangements 507
1.4. Some Connections with Graphs, Groups and Lattices 512
1.5. Poset Homology and Cohomology 513
1.6. Top Cohomology of the Partition Lattice 515

Lecture 2. Group Actions on Posets 519
2.1. Group Representations 519
2.2. Representations of the Symmetric Group 521
2.3. Group Actions on Poset (Co)homology 526
2.4. Symmetric Functions, Plethysm, and Wreath Product Modules 528

Lecture 3. Shellability and Edge Labelings 537
3.1. Shellable Simplicial Complexes 537
3.2. Lexicographic Shellability 541
3.3. CL-shellability and Coxeter Groups 553
3.4. Rank Selection 558

Lecture 4. Recursive Techniques 563
4.1. Cohen-Macaulay Complexes 563
4.2. Recursive Atom Orderings 567
4.3. More Examples 569
4.4. The Whitney Homology Technique 573
4.5. Bases for the Restricted Block Size Partition Posets 579
4.6. Fixed Point Möbius Invariant 586

Lecture 5. Poset Operations and Maps 587
5.1. Operations: Alexander Duality and Direct Product 587
5.2. Quillen Fiber Lemma 591
5.3. General Poset Fiber Theorems 596
5.4. Fiber Theorems and Subspace Arrangements 599
5.5. Inflations of Simplicial Complexes 601

Bibliography 605

Günter M. Ziegler
Convex Polytopes: Extremal Constructions and f-Vector Shapes 617

Introduction 619

Lecture 1. Constructing 3-Dimensional Polytopes 621
1.1. The Cone of f-vectors 623
1.2. The Steinitz Theorem 625
1.3. Steinitz' Theorem via Circle Packings 628

Lecture 2. Shapes of f-Vectors 643
2.1. Unimodality Conjectures 644
2.2. Basic Examples 644
2.3. Global Constructions 647
2.4. Local Constructions 649

Lecture 3. 2-Simple 2-Simplicial 4-Polytopes 653
3.1. Examples 654
3.2. 2-Simple 2-Simplicial 4-Polytopes 657
3.3. Deep Vertex Truncation 659
3.4. Constructing DVT(Stack$(n, 4)$) 661

Lecture 4. f-Vectors of 4-Polytopes 665
4.1. The f-Vector Cone 666
4.2. Fatness and the Upper Bound Problem 669
4.3. The Lower Bound Problem 671

Lecture 5. Projected Products of Polygons 673
5.1. Products and Deformed Products 673
5.2. Computing the f-Vector 674
5.3. Deformed Products 674
5.4. Surviving a Generic Projection 678
5.5. Construction 678

Appendix: A Short Introduction to `polymake`
(by Thilo Schröder and Nikolaus Witte) 681
A.1. Getting Started 681
A.2. The `polymake` System 684

Bibliography 687

Preface

The IAS/Park City Mathematics Institute (PCMI) was founded in 1991 as part of the "Regional Geometry Institute" initiative of the National Science Foundation. In mid 1993 the program found an institutional home at the Institute for Advanced Study (IAS) in Princeton, New Jersey.

The IAS/Park City Mathematics Institute encourages both research and education in mathematics and fosters interaction between the two. The three-week summer institute offers programs for researchers and postdoctoral scholars, graduate students, undergraduate students, high school teachers, undergraduate faculty, and researchers in mathematics education. One of PCMI's main goals is to make all of the participants aware of the total spectrum of activities that occur in mathematics education and research: we wish to involve professional mathematicians in education and to bring modern concepts in mathematics to the attention of educators. To that end the summer institute features general sessions designed to encourage interaction among the various groups. In-year activities at the sites around the country form an integral part of the High School Teachers Program.

Each summer a different topic is chosen as the focus of the Research Program and Graduate Summer School. Activities in the Undergraduate Summer School deal with this topic as well. Lecture notes from the Graduate Summer School are being published each year in this series. The first fourteen volumes are:

- Volume 1: *Geometry and Quantum Field Theory* (1991)
- Volume 2: *Nonlinear Partial Differential Equations in Differential Geometry* (1992)
- Volume 3: *Complex Algebraic Geometry* (1993)
- Volume 4: *Gauge Theory and the Topology of Four-Manifolds* (1994)
- Volume 5: *Hyperbolic Equations and Frequency Interactions* (1995)
- Volume 6: *Probability Theory and Applications* (1996)
- Volume 7: *Symplectic Geometry and Topology* (1997)
- Volume 8: *Representation Theory of Lie Groups* (1998)
- Volume 9: *Arithmetic Algebraic Geometry* (1999)
- Volume 10: *Computational Complexity Theory* (2000)
- Volume 11: *Quantum Field Theory, Supersymmetry, and Enumerative Geometry* (2001)
- Volume 12: *Automorphic Forms and their Applications* (2002)
- Volume 13: *Harmonic Analysis and Partial Differential Equations* (2003)
- Volume 14: *Geometric Combinatorics* (2004)

Volumes are in preparation for subsequent years.

Some material from the Undergraduate Summer School is published as part of the Student Mathematical Library series of the American Mathematical Society. We hope to publish material from other parts of the IAS/PCMI in the future. This will include material from the High School Teachers Program and publications documenting the interactive activities which are a primary focus of the PCMI. At the summer institute late afternoons are devoted to seminars of common interest to all participants. Many deal with current issues in education: others treat mathematical topics at a level which encourages broad participation. The PCMI has also spawned interactions between universities and high schools at a local level. We hope to share these activities with a wider audience in future volumes.

John C. Polking
Series Editor
April 2007

IAS/Park City Mathematics Series
Volume 14, 2004

What is Geometric Combinatorics? –An Overview of the Graduate Summer School

Ezra Miller and Victor Reiner

What is geometric combinatorics? This question is a bit controversial, but at least *in part*, it is the study of **geometric** objects and their **combinatorial** structure. Rather than trying to define this precisely at the outset, in this lecture we'll mainly give examples that appear in the 2004 PCMI graduate courses.

1. Polytopes

A popular class of examples are the *convex polytopes*, that is, convex hulls of finite point sets in $\mathbb{R}^d$. These form the main topic of the graduate course by Ziegler, but also play prominent roles in the undergraduate courses by Swartz and Thomas, and in the undergraduate faculty course by Su (as well as making cameo appearances in the graduate courses by Barvinok, Fomin, Forman, MacPherson, and Wachs!).

In $\mathbb{R}^2$, convex polytopes are *polygons* such as triangles, quadrilaterals, pentagons, hexagons, etc. In $\mathbb{R}^3$ they can be more interesting, such as the triangular prism depicted in Figure 1(a).

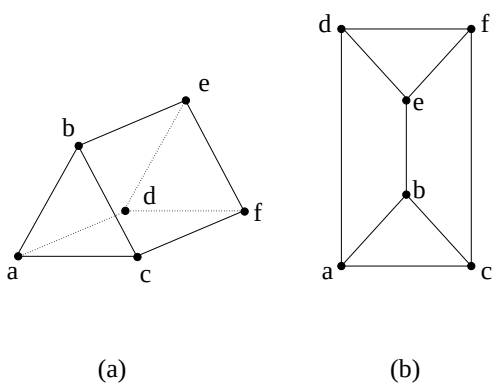


Figure 1. (a) The triangular prism P, with f-vector $f(P) = (f_0, f_1, f_2) = (6, 9, 5)$. (b) Its graph or 1-skeleton, drawn as a 2-dimensional Schlegel diagram.

[1]School of Mathematics, University of Minnesota, Minneapolis MN, 55455.
E-mail address: `ezra@math.umn.edu, reiner@math.umn.edu`.

©2007 American Mathematical Society

What do we mean by combinatorial structure for a convex polytope? An obvious combinatorial feature of a convex polytope is that it has *faces*, each being the intersection of the polytope with some hyperplane containing the polytope entirely in one of its two closed half-spaces. Zero-dimensional faces are called *vertices* (labelled a, b, c, d, e, f in Figure 1), one-dimensional faces are *edges*, and faces of codimension 1 within the polytope are called *facets*. One can record the combinatorial structure of the faces of a convex polytope P in varying ways and levels of detail.

- One might simply *count* the faces of various dimensions, and encode this data in the *f-vector*
$$f(P) = (f_0, f_1, \ldots, f_{d-1}),$$
where $f_i(P)$ is the number of i-dimensional faces of P. For example, the triangular prism in Figure 1 has $f(P) = (f_0, f_1, f_2) = (6, 9, 5)$.
- One might consider the *graph* or *1-skeleton* of P; this is the abstract graph whose node set is the set of vertices of P, and whose (undirected) arcs are the edges of P. For example, Figure 1(b) depicts this graph for the triangular prism. Here we have chosen to draw this graph in the plane by projecting the whole 3-dimensional polytope P to the plane inside one of its quadrangular facets, a visualization technique known as a 2-dimensional *Schlegel diagram* for P. The back of the 2004 PCMI T-shirt depicts the 3-dimensional Schlegel diagram of a four-dimensional polytope with an interesting property, related to work of Joswig and Ziegler [**4**]: it is *dimensionally ambiguous* in the sense that this same 1-skeleton appears also for a five-dimensional polytope.
- One might further consider the entire *partially ordered set* (or *poset*, for short) of all faces of P ordered by inclusion; see Figure 2.

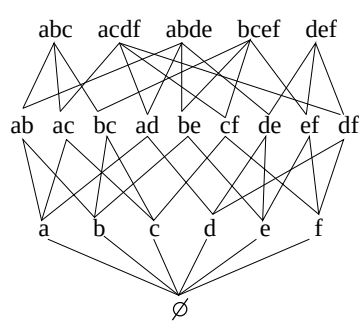


Figure 2. The *Hasse diagram* for the poset of faces of the prism in Figure 1.

2. Characterizing f-vectors

What kinds of combinatorial questions about convex polytopes might we ask? One that has been considered often is the following.

Question 1. *Which (non-negative) vectors $(f_0, f_1, \ldots, f_{d-1})$ in $\mathbb{Z}^d$ can actually arise as the f-vector of a d-dimensional convex polytope?*

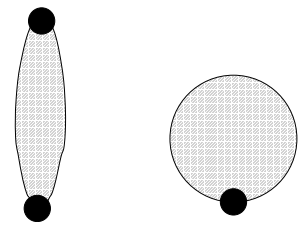


Figure 3. The digon and the monogon: two valid CW-balls, obeying the topological constraint $f_0 = f_1$. The digon has 2 vertices and 2 edges, while the monogon has 1 vertex and 1 edge.

From now on, when we speak of a "d-dimensional" polytope, we will assume that it is *fully* d-dimensional in the sense that its points affinely span a d-dimensional space. For $d = 2$, Question 1 has an obvious answer.

Proposition 2. *A vector $(f_0, f_1) \in \mathbb{Z}^2$ is the f-vector of a 2-dimensional convex polytope (polygon) if and only if*

(i) $f_0 = f_1$, *and*
(ii) $f_0, f_1 \geq 3$.

In spite of its simplicity, this answer foreshadows some important issues arising in higher dimensions. Note that the equation constraint (i) is really a consequence of *topology*: the boundary of a convex polygon is homeomorphic to a one-dimensional sphere. The same equation (i) would hold—without any *polytopality* assumption—for any CW-complex homeomorphic to a 2-dimensional ball, e.g. the *digon* or *monogon* depicted in Figure 3.

On the other hand, the inequality (ii) is really a consequence of polytopality. It highlights the importance of clarifying in which *category* we work when studying f-vectors (such as CW-spheres, regular CW-spheres, PL-spheres, polytopal spheres, etc.) as this can have a dramatic effect on the answers and the difficulty level for questions about f-vectors.

Question 1 for $d = 3$ is also not hard, and was answered by Steinitz roughly a century ago.

Theorem 3. *A vector $(f_0, f_1, f_2) \in \mathbb{Z}^3$ is the f-vector of a 3-dimensional convex polytope if and only if*

(i) $f_0 - f_1 + f_2 = 2$ *(Euler's relation),*
(ii) $f_0, f_2 \geq 4$, *and*
(iii) $2f_1 \geq 3f_2$, $2f_1 \geq 3f_0$.

Again, the equational constraint (i) is a familiar consequence of topology. Polytopality provides us with the first inequality $f_0 \geq 4$ in (ii), since we have assumed that our polytope affinely spans $\mathbb{R}^3$ and hence must have at least 4 affinely independent vertices.[2] The condition $f_2 \geq 4$ then follows from the important tool of *polar duality*: every convex polytope P in $\mathbb{R}^d$ has a (polar) dual polytope $P^\diamond$, whose faces correspond bijectively with those of P, but in an *inclusion-reversing* and *dimension-reversing* fashion. Thus for a 3-dimensional polytope P with f-vector

[2]Actually, both inequalities in (ii) already follow from (i) and (iii), and hence are redundant, but we have included them anyway.

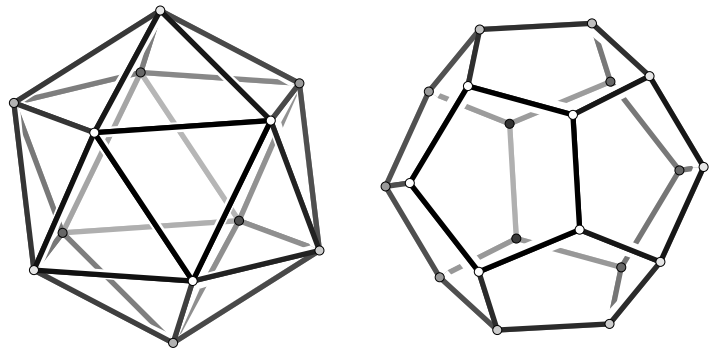


Figure 4. A pair of Platonic solids, which are polar dual to each other: the icosahedron and the dodecahedron. Their f-vectors (f_0, f_1, f_2) are related by reversal, namely $(12, 30, 20)$ and $(20, 30, 12)$, respectively.

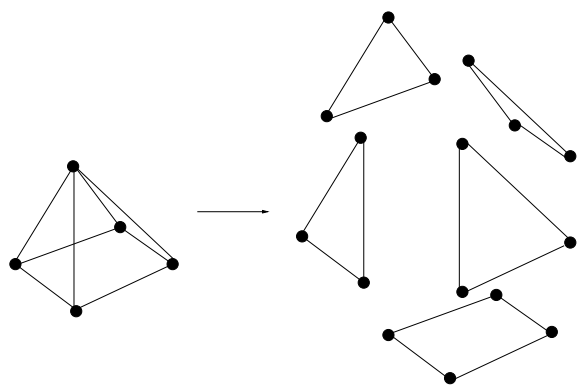


Figure 5. "Blowing apart" the facets of a 3-dimensional polytope and then counting edges in two ways shows that $2f_1 \geq 3f_2$.

(f_0, f_1, f_2), its polar dual $P^\diamondsuit$ will have f-vector (f_2, f_1, f_0). Two classic examples of dual Platonic solids, the icosahedron and dodecahedron are shown in Figure 4.

The remaining inequalities (iii) in the above theorem are another consequence of convexity that follows from counting the edges in the polytope after "blowing apart" the facets, as depicted in Figure 5. Combining the fact that every edge lies in exactly two facets with the fact that each facet has at least three boundary edges, one is led to the inequality $2f_1 \geq 3f_2$. The second inequality in (iii) then follows from polar duality. This shows the necessity of Steinitz's conditions; the sufficiency can be shown by constructing 3-dimensional polytopes with specified f-vectors via some relatively simple constructions (start with a pyramid having an arbitrary polygonal base, and iterate the operation of *shaving off* a vertex, or its polar dual operation of *stellarly subdividing* a facet).

What about Question 1 for $d \geq 4$? In dimension 4 there are only partial answers (see Ziegler's course), and in higher dimensions, the question is wide open.

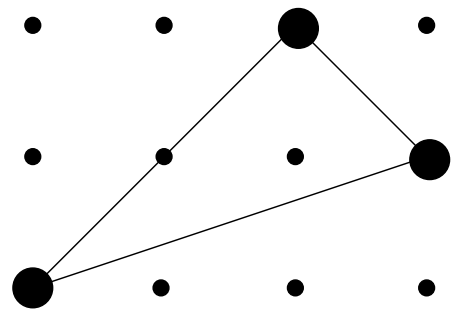


Figure 6. The area of a lattice triangle having $i = 1$ interior lattice point and $b = 4$ boundary lattice points is $i + \frac{1}{2}b - 1 = 2$.

3. Lattice points

There is even more combinatorial structure attached to *lattice polytopes*, the topic of the graduate course by Barvinok, appearing also in the undergraduate course by Thomas as well as the undergraduate faculty course by Su. A lattice polytope is a convex polytope whose vertices lie in $\mathbb{Z}^d$. Here there are non-trivial results even for $d = 2$, that is for lattice polygons! The most famous is probably *Pick's formula* for the area of a lattice polygon.

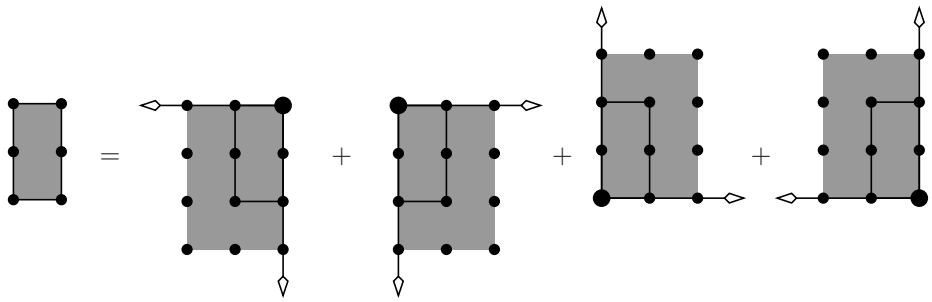
Theorem 4. *(Pick* [**6**]*) Let P be a lattice polygon with i lattice points in its interior and b lattice points on its boundary. Then the area of P is $i + \frac{1}{2}b - 1$.*

Figure 6 illustrates this result for a certain lattice triangle. In fact, Pick's Theorem holds even for lattice polygons which are not convex.

The theory of lattice polytopes becomes more interesting in higher dimensions, including the theory of *Ehrhart polynomials*. It is a subject that has seen many advances within the last decade that have greatly increased our ability for explicit computations. One such advance is *Brion's formula*, which says how to list the lattice points in a lattice polytope. More precisely, let P be a polytope in $\mathbb{R}^d$ with integer vertices. If $\mathbf{a} = (a_1, \ldots, a_d) \in \mathbb{Z}^d$ is a lattice point, then write $\mathbf{t}^{\mathbf{a}} = t_1^{a_1} \cdots t_d^{a_d}$ for the corresponding Laurent monomial. The generating function for the lattice points in P is the sum of all Laurent monomials $\mathbf{t}^{\mathbf{a}}$ for $\mathbf{a} \in \mathbb{Z}^d \cap P$. It is a rational function because it has only finitely many terms. In contrast, consider the *tangent cone* $T_{\mathbf{v}}$ to the polytope at the vertex $\mathbf{v}$, which is the translate by $\mathbf{v}$ of cone generated over the positive real numbers by $P - \mathbf{v}$. The generating function for the lattice points $\mathbb{Z}^d \cap T_{\mathbf{v}}$ in a tangent cone is not a finite sum, but it is still expressible as a rational function $C_{\mathbf{v}}(\mathbf{t})$. Brion's formula breaks the lattice point enumerator of P into a sum over the vertices of P:

$$\sum_{\mathbf{a} \in \mathbb{Z}^d \cap P} \mathbf{t}^{\mathbf{a}} = \sum_{\text{vertices } \mathbf{v} \text{ of } P} C_{\mathbf{v}}(t).$$

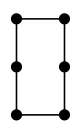
This counter-intuitive result looks like it counts each lattice point in P once for each vertex of P, and furthermore counts all of the lattice points outside of P some number of times, as well. But when that wild-looking generating function (supported on all of $\mathbb{Z}^d$) is expressed as a single rational function, the over-counting inside of P and parts outside of P vanish. Brion's formula is important for computation because it provides a "short" way to represent the set of lattice points in P.



$$
\begin{aligned}
&= \frac{st^2}{(1-s^{-1})(1-t^{-1})} + \frac{t^2}{(1-s)(1-t^{-1})} + \frac{1}{(1-s)(1-t)} + \frac{s}{(1-s^{-1})(1-t)} \\
&= \frac{s^2t^3 - t^3 + 1 - s^2}{(1-s)(1-t)} \\
&= \frac{(1-t^3)(1-s^2)}{(1-s)(1-t)} \\
&= (1+t+t^2)(1+s) \\
&= 1 + t + t^2 + s + st + st^2
\end{aligned}
$$

Figure 7. Brion's formula verified for the 2×1 lattice rectangle in $\mathbb{R}^2$

Example 5. Let $P \subset \mathbb{R}^2$ be the 2×1 lattice rectangle



with vertex set $\{(0,0),(1,0),(2,0),(1,2)\}$. The lattice point enumerator of P, written in variables $(s,t) = (t_1,t_2)$, is $1 + t + t^2 + s + st + st^2$. The lattice points in the tangent cone at (say) the vertex $(1,2)$ of P consist of all integer vectors (a,b) such that $a \leq 1$ and $b \leq 2$. The generating function for these lattice points is $st^2/(1-s^{-1})(1-t^{-1})$. The statement of Brion's formula in this case is verified in the calculation appearing in Figure 7.

4. Hyperplane arrangements

Another interesting example of geometric objects with combinatorial structure are arrangements of *hyperplanes*, the subject of Stanley's graduate course, and other (affine or) linear subspaces of a vector space, which form part of the subject of Wachs's graduate course. Figure 8 illustrates an affine arrangement of hyperplanes (lines) in $\mathbb{R}^2$, along with a *central* arrangement of hyperplanes in $\mathbb{R}^3$ depicted via their intersections with the unit sphere.

Hyperplanes dissect $\mathbb{R}^d$ into open *regions* (or *chambers*), which can be bounded or unbounded, and which one can attempt to count. When one complexifies real hyperplanes or subspaces by considering them inside $\mathbb{C}^d$, they "poke holes" in the space, creating non-trivial topology one can try to measure, e.g. by computing homotopy invariants such as homology or homotopy groups, or cohomology rings. When the hyperplanes or subspaces are defined over $\mathbb{Z}$, one can consider their

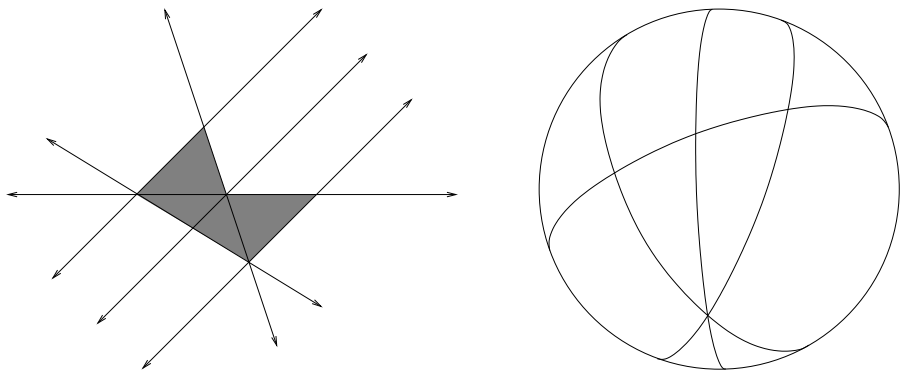


Figure 8. An arrangement of affine lines in $\mathbb{R}^2$ with the bounded regions shaded, and a central arrangement of hyperplanes in $\mathbb{R}^3$ depicted as great circles on a unit sphere.

reductions mod p as arrangements in vector spaces $\mathbb{F}_p^d$ over finite fields, and then count points lying on or off the arrangement. It turns out that almost all of this enumerative or topological analysis comes down to understanding the *topology* of another poset: the *lattice of intersections* of the subspaces, ordered by inclusion.

In particular, one learns that it is important to associate a simplicial complex (and hence a topological space) to this poset, via the ubiquitous *order complex* or *nerve* construction. We also find ourselves in need of a wide array of tools, provided in the graduate course on poset topology by Wachs, for understanding the homotopy or homeomorphism type of the various kinds of simplicial complexes that arise in this way.

5. Symmetry

Many of the examples of combinatorial geometric objects cropping up all over mathematics, such as in the geometry and representation theory of Lie groups and algebras, are those possessing a high degree of *symmetry*. Such objects are the subject of the graduate course by Fomin, and also play a prominent role in the part of Wachs's course that deals with the *equivariant theory* of poset topology.

To give some flavor of Fomin's course, let's look briefly at the classical topic of *regular polytopes*. A regular polytope is one in which every *maximal flag* of faces

$$\text{vertex} \subset \text{ edge} \subset \cdots \subset \text{ facet}$$

"looks" the same, meaning that the group of linear symmetries preserving the polytope acts transitively on all such flags. The 3-dimensional regular polytopes are exactly the Platonic solids, depicted in Figure 4. Classical results in the theory associate to every regular polytope P a certain well-studied and well-behaved hyperplane arrangement: the symmetry group of a regular polytope is always generated by *reflection symmetries*, and one simply takes the associated reflecting hyperplanes for all such symmetries. For the regular tetrahedron, the associated dissection by reflecting hyperplanes and the hyperplane arrangement are shown in Figure 9. Not only do these reflection arrangements play a central role in Fomin's course, but they show up as key motivating examples, along with some of their well-behaved deformations, in Stanley's course as well.

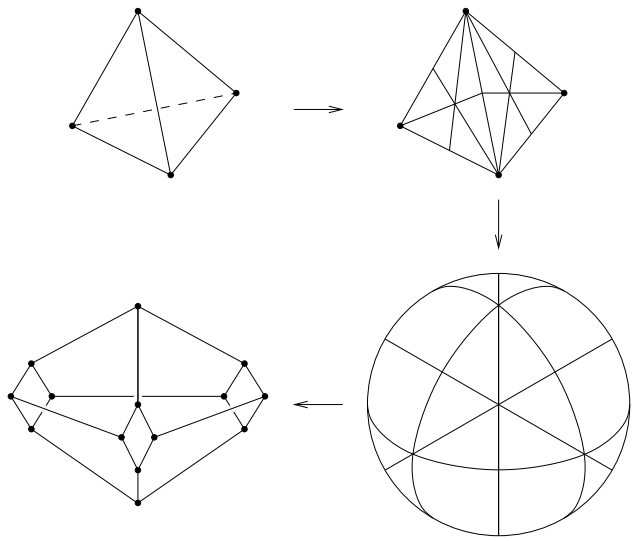


Figure 9. The reflection symmetries of the regular tetrahedron, dissecting its boundary. The associated reflection hyperplane arrangement and root system gives rise to the 3-dimensional associahedron, with f-vector $(f_0, f_1, f_2) = (14, 21, 9)$. Note that $f_0 = 14 = \frac{1}{5}\binom{2\cdot 4}{4} = C_4$ is a Catalan number.

A collection of vectors consisting of a pair of two opposite normal vectors for each of these reflecting hyperplanes gives rise to what is called a *root system*. It should be noted that not every root system comes from the reflection arrangement of a regular polytope, but three of the four infinite families of (finite, irreducible, crystallographic) root systems (types A, B, and C) do arise in this way from the higher-dimensional regular polytopes that generalize tetrahedra (simplices) and cubes/octahedra (hypercubes/hyperoctahedra).

Moving beyond the classical theory, an exciting development in 21st century geometric combinatorics (and a main focus of Fomin's graduate course) has been the discovery of what are called *cluster algebras*. The cluster algebras of finite type give rise to new and important convex polytopes associated to root systems, called *generalized associahedra*. For root systems of type A, these are the classical *associahedra* or *Stasheff polytopes* which have been known for decades in topology, geometry and algebra. The bottom part of Figure 9 depicts the type A associahedron arising from the reflection arrangement for the regular tetrahedron. In type B, one recovers the more recently discovered *cyclohedra* of Bott and Taubes.

These polytopes exhibit wondrous numerology, closely connected with *Catalan numbers* $C_n = \frac{1}{n+1}\binom{2n}{n}$ in type A, and more generally with the mysterious numerology of *exponents* for all root systems. A great deal of intriguing combinatorics awaits discovery in these polytopes.

6. Moment graphs

Geometric combinatorics does not only concern structures arising from spaces that feel discrete. Smooth spaces often have underlying combinatorics, as well. Many smooth spaces can be considered from the point of view in MacPherson's course,

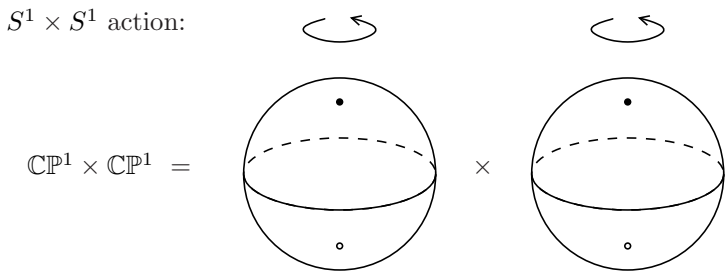


Figure 10. $\mathbb{CP}^1 \times \mathbb{CP}^1$ and the action of $S^1 \times S^1$

where the combinatorics takes the form of a graph drawn with straight edges in $\mathbb{R}^n$. The setup is as follows.

An *algebraic torus* is a group of the form $T = (\mathbb{C}^*)^n$, where $\mathbb{C}^* = \mathbb{C} \setminus \{0\}$ is the set of nonzero complex numbers, considered as a group under multiplication. Inside of the algebraic torus T is an honest compact torus $T_{\mathbb{R}} = (S^1)^n$, the product of n copies of the unit circle group. MacPherson's course concerns spaces X with an action of T. More precisely, let X be a smooth compact complex algebraic variety of dimension d; thus X is a real manifold of dimension $2d$ with some extra structure to make it a manifold over $\mathbb{C}$. We require that the action $T : X \to X$ has finitely many

- T-fixed points and
- complex 1-dimensional orbits.

An orbit of complex dimension 1 has real dimension 2, and is necessarily isomorphic to a copy of $\mathbb{C}^*$. Since X is compact, the closure of such an orbit is an isomorphic copy of the Riemann sphere (projective complex line) $\mathbb{P}^1$: add an origin 0 and a point ∞ at infinity (both of which will be T-fixed points) to the copy of $\mathbb{C}^*$. The union of the T-fixed points and the 1-dimensional orbits is a configuration, called a *balloon sculpture*, of finitely many Riemann spheres in X joined at some of their poles. The *moment graph* is a real 1-dimensional shadow of the complex 1-dimensional balloon sculpture. It is obtained from the balloon sculpture by identifying together all points in each orbit of the compact torus $T_{\mathbb{R}}$.

Example 6. Let $X = \mathbb{CP}^1 \times \mathbb{CP}^1$ be a product of two Riemann spheres. This space comes with an action of $T = \mathbb{C}^* \times \mathbb{C}^*$, so $n = 2$ in the preceding notation. The compact torus $T_{\mathbb{R}} = S^1 \times S^1$ is the familiar real 2-dimensional doughnut. The two copies of S^1 spin the corresponding spheres $\mathbb{CP}^1$ around their axes, each leaving the other sphere fixed pointwise, as depicted in Figure 10. The balloon sculpture in X consists of four spheres joined pole-to-pole in a cycle, as in Figure 11. The circles of latitude in the four balloons are $T_{\mathbb{R}}$ orbits, as are each of the poles. Collapsing each of these orbits to a point yields the moment graph of $\mathbb{CP}^1 \times \mathbb{CP}^1$: a square.

In the above example, the quotient of all of X by $T_{\mathbb{R}}$ is the entire square—including the interior, over which the $T_{\mathbb{R}}$ orbits are 2-dimensional tori. More generally, for every lattice polytope P there is a *toric variety* X_P whose moment graph is the edge graph of P, and whose quotient by $T_{\mathbb{R}}$ is all of P. Although toric varieties constitute a very important class of examples—they are the simplest spaces with moment graphs—they aren't the only spaces with moment graphs.

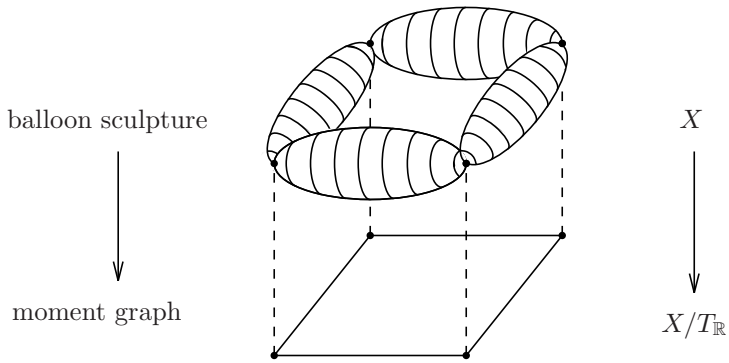


Figure 11. The balloon sculpture of $X = \mathbb{CP}^1 \times \mathbb{CP}^1$ and its map to the moment graph

Example 7. Let X be the quadric hypersurface in $\mathbb{CP}^6$ consisting of the solutions to the polynomial equation $z^2 + x_1y_1 + x_2y_2 + x_3y_3 = 0$. The algebraic torus $T = (\mathbb{C}^*)^3$, with coordinates (τ_1, τ_2, τ_3), acts by

$$(\tau_1, \tau_2, \tau_3) \cdot (z : x_1 : x_2 : x_3 : y_1 : y_2 : y_3) = (z : \tau_1 x_1 : \tau_2 x_2 : \tau_3 x_3 : \tau_1^{-1} y_1 : \tau_2^{-1} y_2 : \tau_3^{-1} y_3);$$

that is, the polynomial equation is invariant under T. The moment graph is ubiquitous when it comes to this summer school. In particular, there is geometric combinatorics on the front of the 2004 PCMI T-shirt as well as on the back! This example is straight from MacPherson's notes, where it is treated in more detail.

Figure 12. The moment graph of the quadric hypersurface X in $\mathbb{CP}^6$

The methods surrounding moment graphs are particularly well-suited to spaces like complete flag manifolds and their relatives, including Grassmannians, other quotients of compact Lie groups by parabolic subgroups, and loop Grassmannians. These spaces are crucial to interactions of combinatorics with representation theory and algebraic geometry. Moment graphs (and *moment maps*, when they are available) are vehicles by which smooth spaces give rise to more obviously discrete-geometric objects such as polytopes, graphs, and root systems. A hint of the consequences of this transition occurs in Fomin's graduate course.

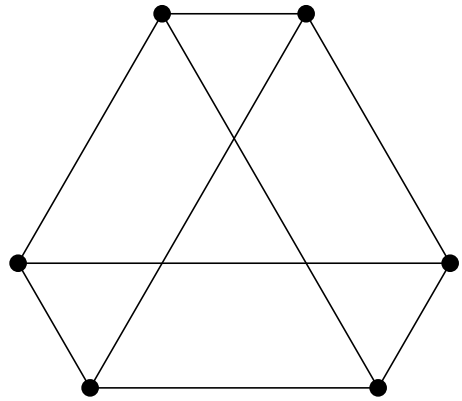


Figure 13. The convex hull of the moment graph of the flag manifold $\mathcal{F}\ell_3$ is a permutohedron

Example 8. Let $X = \mathcal{F}\ell_3$ be the manifold of flags in $\mathbb{C}^3$. Thus $\mathcal{F}\ell_3$ consists of the chains $\{0\} = V_0 \subset V_1 \subset V_2 \subset V_3 = \mathbb{C}^3$ of vector subspaces of $\mathbb{C}^3$ with $\dim V_i = i$. The algebraic torus $(\mathbb{C}^*)^3$ naturally acts on X by virtue of its action on $\mathbb{C}^3$. The moment graph of $\mathcal{F}\ell_3$ is depicted in Figure 13. The graph can be naturally embedded in a plane sitting in $\mathbb{R}^3$, and its convex hull, which is the image of the moment map, is a hexagon. More generally, the convex hull of the moment graph of the manifold $\mathcal{F}\ell_n$ of flags in $\mathbb{C}^n$ is a polytope called the *permutohedron*, whose vertices are the $n!$ permutations of $(1, \ldots, n)$. Unlike the moment graphs of toric varieties (but like the PCMI logo in Figure 12), the edges of the permutohedron constitute only part of the moment graph of $\mathcal{F}\ell_n$, which also has edges passing through the interior.

7. Fixed points of smooth symmetries

Moment graphs isolate combinatorial structures from a priori smooth geometric contexts. But what is this combinatorial data good for? Although it may not seem likely at first, the moment graph actually retains an enormous amount of information about a space X. In particular, much of the topology of X can be faithfully recovered from the discrete data of its moment graph.

This sort of claim reflects a phenomenon that is quite general. Without introducing too many hypotheses, the setup is that X should be a space with an action of some Lie group G such that the set X^G of fixed points is finite. Now suppose that ξ is some global topological invariant of X that is G-*equivariant*, meaning that it takes into account the G-action. The aim is to produce statements saying that

$$\xi = \sum_{x \in X^G} \xi_x$$

breaks up as a sum of local contributions ξ_x at the fixed points. Theorems of this form are called *localization theorems* or *fixed point formulas*, and often come attached to names such as Atiyah, Bott, or Lefschetz. The idea is that the residual action of G on the tangent spaces to the G-fixed points carries enough information about the action on X to reconstruct topological data.

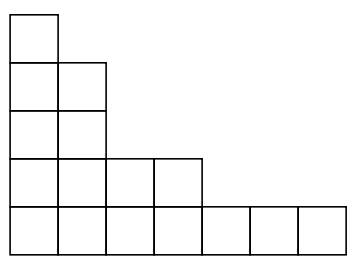
In MacPherson's course, ξ is an equivariant cohomology class or some related invariant, the point being that the entire equivariant cohomology ring of X is determined by its moment graph. In Barvinok's course, taking ξ to be the character of the global sections of a line bundle on a toric variety yields Brion's theorem as a statement in equivariant K-theory. For instance, the computation in Example 5 comes from localization applied to the line bundle $\mathcal{O}(1,2)$ on the toric variety from Example 6. Of course, this key to Barvinok's polynomial-time algorithms for lattice point enumeration does not, in a logical sense, require thinking about equivariant K-theory of toric varieties; but it is worthwhile to note that it was in such a context that Brion discovered the formula in the first place.

The notion that topology is encoded by local data near fixed points is a powerful one. Even forgetting temporarily about the two preceding examples, it has far-reaching consequences, combinatorial and otherwise, ranging from Okounkov and Pandharipande's proof of Witten's conjecture (Kontsevich's theorem) [**5**] to Deligne's proof of the Weil conjectures (see [**2**]).

Yet another example underlies a fundamental part of the geometry in Haiman's course. The smooth space there is the *Hilbert scheme* H_n of n points in the plane $\mathbb{C}^2$. As a set, H_n consists of the ideals $I \subseteq \mathbb{C}[x,y]$ in the two-variable polynomial ring such that $\mathbb{C}[x,y]/I$ has dimension n as a vector space over $\mathbb{C}$. The ideal of polynomials vanishing on n distinct points in $\mathbb{C}^2$ is an example of a point $I \in H_n$. However, there are other colength n ideals, such as the ideal generated by x^n and y; a $\mathbb{C}$-linear basis for the quotient $\mathbb{C}[x,y]/I$ is given by $\{1, x, x^2, \ldots, x^{n-1}\}$.

More generally, for every *partition* λ of n, meaning a weakly decreasing list of integers whose sum is n, there is an ideal I_λ generated by monomials. Think of λ as a (Young) shape, so that for example

$\lambda = (7,4,2,2,1) \quad \longleftrightarrow$



is a partition of $7+4+2+2+1 = 16$ in "French" notation. The nooks immediately outside of λ can be labeled naturally with monomials as in Figure 14. The ideal I_λ is then generated by these monomials. Thus, for $\lambda = (7,4,2,2,1)$ we get

$$I_\lambda = \langle x^7, x^4y, x^2y^2, xy^4, y^5 \rangle.$$

It is easy to verify that the boxes inside λ correspond to monomials that constitute a $\mathbb{C}$-linear basis for the quotient $\mathbb{C}[x,y]/I_\lambda$. In particular, if λ is a partition of n then $\mathbb{C}[x,y]/I_\lambda$ has dimension n as a vector space over $\mathbb{C}$.

The torus $T = (\mathbb{C}^*)^2$ acts on H_n because it acts on $\mathbb{C}^2$ by scaling the axes. More concretely, if $I = \langle f_1(x,y), \ldots, f_r(x,y) \rangle$ is an ideal, then $(\sigma,\tau) \in T$ acts on I by

$$(\sigma,\tau) \cdot I = \langle f_1(\sigma x, \tau y), \ldots, f_r(\sigma x, \tau y) \rangle.$$

If $f(x,y)$ is a monomial, then $f(\sigma x, \tau y)$ is a scalar multiple of $f(x,y)$. Therefore, if all of the polynomials $f_i(x,y)$ are monomials, then $(\sigma,\tau) \cdot I = I$. In other words, if

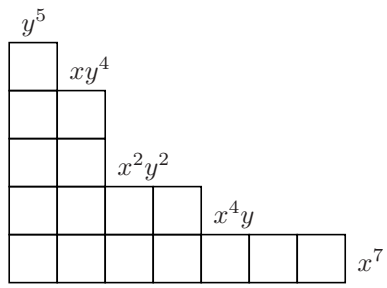


Figure 14. Monomials in the nooks immediately outside of the partition $\lambda = (7, 4, 2, 2, 1)$

$I = I_\lambda$ for some partition λ then I_λ is a T-fixed point of H_n. The converse holds as well: if I is a T-fixed point, then $I = I_\lambda$ is a monomial ideal for some partition λ.

For Hilbert schemes, therefore, combinatorics is evident already in the fixed points themselves, regardless of localization theorems. This makes fixed point formulas on H_n all the more interesting: any such formula will have a sum $\sum_\lambda \xi_\lambda$ over partitions λ of n on one side of the equation. What Haiman's geometric theory shows is that, for certain vector bundles and more general sheaves on H_n with interesting global section characters ξ, fixed point formulas result in extraodinarily interesting sums over λ.

The reason why the fixed point formulas are so interesting is that a certain particularly natural vector bundle on H_n yields summands ξ_λ that are essentially the *Macdonald polynomials* from Lecture 3 of Haiman's course. This statement is equivalent to the $n!$ theorem (see Theorem 8 in the notes by Haiman and Woo). One of the fixed point formulas it yields results in the $(n+1)^{n-1}$ theorem [**3**], a combinatorial statement that motivated the whole geometric story. It says that

$$R_n = \mathbb{C}[x_1, y_1, \ldots, x_n, y_n]/\langle x_1^r y_1^s + \cdots + x_n^r y_n^s \mid r, s \in \mathbb{N}\rangle,$$

which is known as the ring of *diagonal coinvariants*, has dimension $(n+1)^{n-1}$ as a vector space over $\mathbb{C}$. In fact, since the summands ξ_λ are torus-equivariant data at the fixed point I_λ, the fixed point formula is a doubly-graded version of this enumerative statement. Combinatorial methods for such *q, t-analogues* in general, and the Macdonald polynomials in particular, constitute Haiman's course.

8. Morse theory

Localization theorems are powerful ways to reconstruct topological invariants from knowledge of local data near fixed points. However, even to speak of fixed points we must have a group action. In the preceding situations, such actions were natural, in that they were fundamental to the smooth spaces under consideration. The flag manifold, for instance, is the quotient of a Lie group by a closed subgroup, and hence obviously has lots of Lie group actions on it; and a toric variety is (by some definitions) the closure of a dense torus orbit. But what if our smooth space doesn't come with a natural Lie group action? Make a group action from scratch!

Suppose that X is a real manifold with a Riemannian metric. Any real-valued function $f : X \to \mathbb{R}$ yields a *gradient flow* on X: each point goes in the direction of steepest descent. Gradient flow can be viewed as an action of the Lie group $\mathbb{R}$

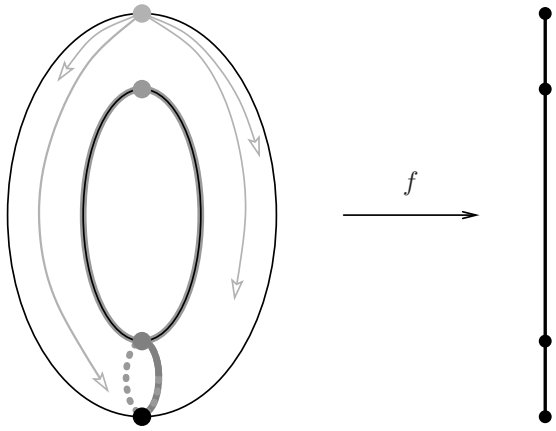


Figure 15. Four critical points on a torus, with the negative flow directions

(thought of as parametrizing time) on X. The fixed points of the flow are the critical points of f, where the derivative of f vanishes; these points are ambivalent about which direction to go, so they stay put. See Figure 15 for an example with four critical points on a torus. When f is generic, we can define the *index* of a critical point x to be the number of independent directions at x in which the flow points away from x—that is, the limit is x as time approaches $-\infty$.

Topological invariants are extracted from this (more or less) combinatorial data of critical points, indexes, and downward flow submanifolds by constructing a cell decomposition of X. There is one cell for each critical point, and the dimension of the cell is the index of the critical point. From Figure 15, we see that a torus can be constructed from a vertex (the bottom critical point), two edges (the middle two critical points), and one 2-cell (the top critical point). The manner in which the downward flow submanifold from one critical point approaches the other critical points determines how to glue the cells.

Gradient flow is all well and good if we're given a smooth manifold. But what if, in the spirit of how this Overview started, we're given a discrete geometric object, such as a collection of polytopes or a simplicial complex Δ? The answer lies in Forman's lectures: use discrete Morse theory. The idea is strikingly simple. Let P be the Hasse diagram of the face poset of Δ. Orient all of the edges of P downward. A *Morse flow* in this context is a (partial) matching on P such that reversing the edges in the matching does not ruin the directed acylicity property of the directed graph P. This mirrors the stipulation that our Morse function f mapped X to the real numbers, and not (for example) to the circle. The *critical simplices* of the Morse flow are the unmatched elements of P. In analogy with the smooth case, the critical simplices correspond to cells in a complex that is homotopy-equivalent to Δ, so the topological invariants have not changed. Discrete Morse theory is an extremely useful tool in making explicit calculations. It is also a key theoretical tool for poset homology, which leads to Wachs's course.

Going beyond Morse theory, it is possible to combinatorialize a number of other notions from differential geometry. Forman's fourth and fifth lectures, for example, discuss combinatorializations of curvature, and the purely combinatorial

questions that result as a consequence. Of course, combinatorialization often helps us understand the smooth setting better. Recent work of Biss [**1**], for example, shows that understanding metric tangent data in a purely combinatorial context loses no topological information whatsoever. The discrete analogues of smooth tangent bundles are, in that case, matroid bundles on combinatorial differentiable manifolds ("CD-manifolds").

9. Further topics

We have tried in this Overview to give an idea of what "geometric combinatorics" might mean, although (for obvious reasons) we have done so mostly in the context of the courses at PCMI 2004. But this summer's offerings are by no means comprehensive! There are vast numbers of ways combinatorial structure arises in geometry. Here, for example, is a small list of keywords.

- Tropicalization: polyhedral structures reflect the geometry of complex algebraic varieties.
- Degeneration: replace a manifold or variety, such as a Schubert variety, by a degenerate version that has several components, each of which is simpler.
- Stratification: different strata, as in moduli spaces of curves, can represent collections of geometric objects with identical combinatorial properties.
- Branch point data (Hurwitz schemes and ramified covers): counting methods rely on combinatorics of the symmetric group.
- Generating functions: for example, Gromov–Witten theory leads to multivariate hypergeometric series.
- Characteristic classes: for example, functorial approaches to graph coloring and Tverberg-type theorems.

Some of the above items were hot topics at the 2004 PCMI Research Program: the Clay lecture by Sturmfels was one of many talks about tropical geometry and its applications, and the research talk by (for example) Vakil concerned recent advances using degeneration. The last item on the list was expanded by Kozlov to a survey paper that is included in this volume. The survey concerns *graph complexes* and functorial approaches to graph coloring. More precisely, in 1978 Lovász proved a subtle conjecture of Kneser in graph theory using functoriality: a proper vertex-coloring of a graph is intepreted as a morphism in a certain category of graphs. This leads to a morphism between two topological spaces with free $\mathbb{Z}_2$-actions, to which the Borsuk–Ulam theorem can be applied. Recently these techniques of graph complexes and characteristic classes have been greatly extended, culminating in Babson and Kozlov's proof of a conjecture of Lovász.

Keeping in mind that the above list is incomplete, it should be clear that there would never be enough time to cover all of the relevant topics. The only remedy would be another summer school on Geometric Combinatorics.

BIBLIOGRAPHY

1. D. K. Biss, *The homotopy type of the matroid Grassmannian*, Ann. of Math. (2) **158** (2003), no. 3, 929–952.
2. E. Freitag and R. Kiehl, *Étale cohomology and the Weil conjecture*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)] Vol. 13, Springer-Verlag, Berlin, 1988.
3. M. Haiman, *Vanishing theorems and character formulas for the Hilbert scheme of points in the plane*, Invent. Math. **149** (2002), 371-407.
4. M. Joswig and G.M. Ziegler, *Neighborly cubical polytopes*, Discrete Comput. Geom. **24** (2000), 325–344.
5. A. Okounkov and R. Pandharipande, *Gromov–Witten theory, Hurwitz numbers, and Matrix models, I*, preprint. arXiv:math.AG/0101147
6. G. Pick, *Geometrisches zur Zahlenlehre*, Sitzungsberichte Lotos (Prag), Natur-med. Verein für Böhmen **19** (1899), 311–319.

Lattice Points, Polyhedra, and Complexity

Alexander Barvinok

IAS/Park City Mathematics Series
Volume 14, 2004

Lattice Points, Polyhedra, and Complexity

Alexander Barvinok

Introduction

The central topic of these lectures is efficient counting of integer points in polyhedra. Consequently, various structural results about polyhedra and integer points are ultimately discussed with an eye on computational complexity and algorithms. This approach is one of many possible and it suggests some new analogies and connections. For example, we consider unimodular decompositions of cones as a higher-dimensional generalization of the classical construction of continued fractions. There is a well recognized difference between the theoretical computational complexity of an algorithm and the performance of a computational procedure in practice. Recent computational advances [**L+04**], [**V+04**] demonstrate that many of the theoretical ideas described in these notes indeed work fine in practice. On the other hand, some other theoretically efficient algorithms look completely "unimplementable", a good example is given by some algorithms of [**BW03**]. Moreover, there are problems for which theoretically efficient algorithms are not available at the time. In our view, this indicates the current lack of understanding of some important structural issues in the theory of lattice points and polyhedra. It shows that the theory is very much alive and open for explorations.

Exercises constitute an important part of these notes. They are assembled at the end of each lecture and classified as review problems, supplementary problems, and preview problems.

Review problems ask the reader to complete a proof, to fill some gaps in a proof, or to establish some necessary technical prerequisites. Problems of this kind tend to be relatively straightforward. To be able to complete them is essential for understanding.

Supplementary problems explore various topics in more depth and breadth. Problems of this kind can be harder. They may use some general concepts which are not formally introduced in the text, but which, nevertheless, are likely to be familiar to the reader.

[1]Department of Mathematics, University of Michigan, Ann Arbor, MI 48109-1043.
E-mail address: barvinok@umich.edu.
This work is partially supported by the NSF grant DMS 0400617.

©2007 American Mathematical Society

Preview problems address what is going to appear in the following lectures or could be an object of further study. The purpose of these problems is to make the reader prepared, to the extent possible, for further developments.

Not every result mentioned in these lectures is accompanied by a complete proof. However, the reader can find the details in the references at the end of each chapter.

Acknowledgments. I am grateful to Ezra Miller, Vic Reiner, and Bernd Sturmfels, the organizers of the 2004 Graduate Summer School at Park City, for their invitation to give these lectures and for their support. I am grateful to students and researchers who attended the lectures, asked questions, and otherwise showed their interest in the material. It is my pleasure to thank Greg Blekherman for the excellent job of conducting review sessions where the lecture material was discussed and problems were solved. I am indebted to Greg Blekherman and Kevin Woods for reading the first, pre-event, version of the notes and suggesting corrections and improvements and to Kevin Woods for detailed comments and suggestions on the post-event version.

LECTURE 1
Inspirational Examples. Valuations

The theory we are about to describe is inspired by two simple well-known formulas. Our first inspiration comes from the formula for the sum of the finite geometric series.

Example 1.

$$\sum_{m=0}^{n} x^m = \frac{1 - x^{n+1}}{1 - x}.$$

We observe that the long polynomial on the left-hand-side of the equation sums up to a short rational function on the right-hand-side.

Geometrically, we do the following: we take the interval $[0, n]$, for every integer point m in the interval we write the monomial x^m, and then take the sum over the integer points in the interval, see Figure 1.

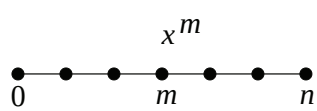


Figure 1. Integer points in the interval

We observe that the thus obtained "long" polynomial (it contains $n + 1$ monomials) can be written as a "short" rational function (it is expressed in terms of only 4 monomials).

Naturally, we ask what happens if we replace the interval by something higher-dimensional. Let us, for example, draw a big triangle in the plane, for each integer point $m = (m_1, m_2)$ in the triangle let us write the bivariate monomial $\mathbf{x}^m = x_1^{m_1} x_2^{m_2}$, and then let us try to write the sum over all integer points in the triangle as some simple rational function in x_1 and x_2, see Figure 2. If the triangle is really large, we get a really long polynomial this way. Later, we will see how to write it as a short rational function.

Our second inspiration comes from the formula for the sum of the infinite geometric series.

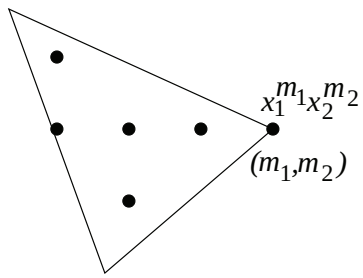


Figure 2. Integer points in the triangle

Example 2.

$$\sum_{m=0}^{+\infty} x^m = \frac{1}{1-x}.$$

This formula makes sense because the series on the left-hand-side converges for all $|x| < 1$ to the function on the right-hand-side. Similarly,

$$\sum_{m=-\infty}^{0} x^m = \frac{1}{1-x^{-1}} = \frac{-x}{1-x}$$

makes sense because the series converges for all $|x| > 1$.

How do we make sense of

$$\sum_{m=-\infty}^{+\infty} x^m \qquad ?$$

This sum does not converge for any x, so we take the easiest route and say that the sum is 0. This may look bizarre but there is some consistence in the way we define the sums: the inclusion-exclusion principle seems to be respected. Indeed, we get the set of all integers if we take all non-negative integers, add all non-positive integers, and subtract 0, as it was double-counted:

$$\sum_{m=-\infty}^{+\infty} x^m = \sum_{m=0}^{+\infty} x^m + \sum_{m=-\infty}^{0} x^m \quad - \quad x^0.$$

This suspiciously agrees with

$$0 = \frac{1}{1-x} + \frac{-x}{1-x} - 1.$$

Geometrically, the real line $\mathbb{R}^1$ is divided into two unbounded rays intersecting in a point. For every region (the two rays, the line, and the point), we construct a rational function so that the sum of x^m over the lattice points in the region converges to that rational function, if it converges at all, and the inclusion-exclusion principle is upheld, see Figure 3.

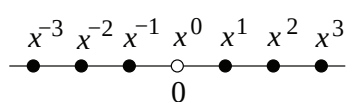


Figure 3. The real line divided into two rays

Naturally, we ask what happens in higher dimensions. Let us draw three lines in general position in the plane: each line splits the plane into two halfplanes, every two lines form four angles, and there are various other regions (one triangle, the whole plane, and some nameless unbounded polygonal regions), see Figure 4.

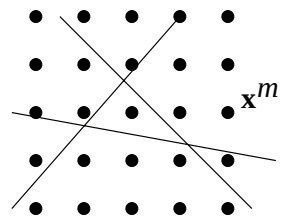


Figure 4. The plane divided into regions

Among those regions, there are regions R where the sum $\sum_{m\in R\cap\mathbb{Z}^2}\mathbf{x}^m$ converges for some $\mathbf{x}$, and there are regions where such a sum would never converge. Can we assign a rational function to every region simultaneously so that each series converges to the corresponding rational function, if it converges at all, and the inclusion-exclusion principle is observed?

Later, we will see that such an assignment is indeed possible.

We need some definitions.

Defintion 1. The action takes place in Euclidean space $\mathbb{R}^d$, with coordinates $x = (x_1, \dots, x_d)$, the scalar product

$$\langle x, y\rangle = \sum_{i=1}^{d} x_i y_i \quad \text{for} \quad x = (x_1, \dots, x_d) \quad \text{and} \quad y = (y_1, \dots, y_d),$$

and with the integer point lattice $\mathbb{Z}^d \subset \mathbb{R}^d$, consisting of the points x with integer coordinates. A *polyhedron* $P \subset \mathbb{R}^d$ is the set of solutions to finitely many linear inequalities,

$$P = \Big\{x \in \mathbb{R}^d : \quad \sum_{i=1}^{d} a_{ij}x_j \le b_i, \quad i = 1, \dots, m\Big\}.$$

If all a_{ij}, b_i are integers, the polyhedron is *rational*. The main object in these notes is the set $P \cap \mathbb{Z}^d$ of integer points in a rational polyhedron P.

What can we do with polyhedra? The intersection of finitely many (rational) polyhedra is a (rational) polyhedron. The union doesn't have to be but may happen to be a polyhedron. To account for all possible relations among polyhedra, we introduce the *algebra of polyhedra*.

Defintion 2. For a set $A \subset \mathbb{R}^d$, let $[A] : \mathbb{R}^d \longrightarrow \mathbb{R}$ be the indicator of A. Thus $[A]$ is the function on $\mathbb{R}^d$ defined by

$$[A](x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases}$$

The *algebra of polyhedra* $\mathcal{P}(\mathbb{R}^d)$ is the vector space spanned by the indicators $[P]$ for all polyhedra $P \subset \mathbb{R}^d$. The coefficient field does not matter much: it can be $\mathbb{Q}, \mathbb{R}$, or $\mathbb{C}$.

The *algebra of rational polyhedra* $\mathcal{P}(\mathbb{Q}^d) \subset \mathcal{P}(\mathbb{R}^d)$ is defined similarly as the subspace spanned by the indicators $[P]$ of rational polyhedra P.

Why do we call $\mathcal{P}(\mathbb{R}^d)$ and $\mathcal{P}(\mathbb{Q}^d)$ algebras? So far, we defined $\mathcal{P}(\mathbb{R}^d)$, $\mathcal{P}(\mathbb{Q}^d)$ as vector spaces. There is one obvious algebra structure on $\mathcal{P}(\mathbb{R}^d)$ and $\mathcal{P}(\mathbb{Q}^d)$. Namely, let $f, g : \mathbb{R}^d \longrightarrow \mathbb{R}$ be functions from the algebras. Then we can define their point-wise product $h = fg$ by $h(x) = f(x)g(x)$. It is immediate to check that h indeed lies in the corresponding algebra. There is a less obvious though more interesting algebra structure on $\mathcal{P}(\mathbb{R}^d)$ and $\mathcal{P}(\mathbb{Q}^d)$, see Supplementary Problem 2 in Lecture 2.

Another observation: as long as $d > 0$, the indicators $[P]$ of (rational) polyhedra $P \subset \mathbb{R}^d$ *do not* form a basis of $\mathcal{P}(\mathbb{R}^d)$, $\mathcal{P}(\mathbb{Q}^d)$, because they are linearly dependent. This is what makes the theory interesting.

Valuations

Let V be a vector space. A linear transformation $\mathcal{P}(\mathbb{R}^d), \mathcal{P}(\mathbb{Q}^d) \longrightarrow V$ is called a *valuation*. Basically, this course is about the existence and properties of one particular valuation $\mathcal{P}(\mathbb{Q}^d) \longrightarrow \mathbb{C}(x_1, \ldots, x_d)$, where $\mathbb{C}(x_1, \ldots, x_d)$ is the space of d-variate rational functions. We saw a glimpse of this valuation in Examples 1 and 2.

To warm up, we introduce one of the simplest and most useful valuations.

Theorem 1. *There exists a unique valuation $\chi : \mathcal{P}(\mathbb{R}^d) \longrightarrow \mathbb{R}$, called the Euler characteristic, such that $\chi([P]) = 1$ for any non-empty polyhedron $P \subset \mathbb{R}^d$.*

Sketch of proof. Uniqueness of χ, if it exists, is clear: there is at most one way to extend the definition $\chi([P]) = 1$ linearly on the whole algebra $\mathcal{P}(\mathbb{R}^d)$. Because the indicators $[P]$ are linearly dependent, it is not at all obvious that such an extension exists. To establish existence, we use induction on the dimension d. If $d = 0$, we define $\chi(f) = f(0)$ and it works.

Suppose that $d > 0$. First, we prove the existence of χ on the subspace of $\mathcal{P}(\mathbb{R}^d)$ spanned by the indicators of bounded polyhedra, also known as polytopes. Let us slice $\mathbb{R}^d$ into copies of $\mathbb{R}^{d-1}$ by the value of the last coordinate of a point. That is, we define H_t to be the hyperplane $x_d = t$. Then H_t looks like $\mathbb{R}^{d-1}$ and by the induction hypothesis there is the Euler characteristic χ_t there. Given a function $f \in \mathcal{P}(\mathbb{R}^d)$, we define its restriction f_t onto H_t. One can easily check that if f is a linear combination of indicators of bounded polyhedra in $\mathbb{R}^d$ then f_t is a linear combination of indicators of bounded polyhedra in H_t. Hence, we can define $\chi_t(f_t)$. Now, the key observation is that the one-sided limit

$$\lim_{\epsilon \longrightarrow +0} \chi_{t-\epsilon}(f_{t-\epsilon})$$

always exists and that for all but finitely many t's it is equal to $\chi_t(f_t)$. In fact, if

$$f = \sum_i \alpha_i [P_i],$$

then

$$\lim_{\epsilon \longrightarrow +0} \chi_{t-\epsilon}(f_{t-\epsilon}) = \chi_t(f_t)$$

unless t is the minimum value of the last coordinate on one of the polyhedra P_i in the support of f, see Figure 5.

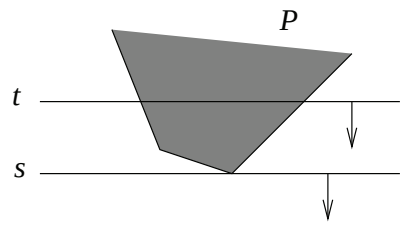


Figure 5. Example: for $f = [P]$, we have $\lim_{\epsilon \longrightarrow +0} \chi_{t-\epsilon}(f_{t-\epsilon}) = \chi_t(f_t) = 1$ and $0 = \lim_{\epsilon \longrightarrow +0} \chi_{s-\epsilon}(f_{s-\epsilon}) \neq 1 = \chi_s(f_s)$

This allows us to define

$$\chi(f) = \sum_{t \in \mathbb{R}} \Big(\chi_t(f_t) - \lim_{\epsilon \longrightarrow +0} \chi_{t-\epsilon}(f_{t-\epsilon}) \Big).$$

Although the sum is infinite, only finitely many terms are non-zero.

One can check that χ satisfies the required properties.

Now, we extend χ to the whole algebra $\mathcal{P}(\mathbb{R}^d)$. Let us take P_t to be the cube $|x_i| \leq t$ for $i = 1, \ldots, d$ and let us define

$$\chi(f) = \lim_{t \longrightarrow +\infty} \chi(f \cdot [P_t]) \quad \text{for} \quad f \in \mathcal{P}(\mathbb{R}^d).$$

□

Problems

Review problems.

1. Let $A_1, \ldots, A_n \subset \mathbb{R}^d$ be sets. Prove the inclusion-exclusion formula

$$\left[\bigcup_{i=1}^{n} A_i \right] = \sum_{I} (-1)^{|I|-1} \left[\bigcap_{i \in I} A_i \right],$$

   where the sum is taken over all non-empty subsets $I \subset \{1, \ldots, n\}$ and $|I|$ is the cardinality of I.

2. Fill in the gaps in the proof of Theorem 1.

3. Show that the Euler characteristic can be extended to the space spanned by the indicators $[A]$ of closed convex sets $A \subset \mathbb{R}^d$ so that $\chi([A]) = 1$ if A is a non-empty closed convex set (a set A is called *convex* if, for every pair of points $x, y \in A$ it contains the interval $[x, y] = \{\alpha x + (1-\alpha) y : \ 0 \leq \alpha \leq 1\}$).

A supplementary problem.

1. Let $P \subset \mathbb{R}^d$ be a bounded polyhedron with a non-empty interior $\operatorname{int} P$. Show that $[\operatorname{int} P] \in \mathcal{P}(\mathbb{R}^d)$ and that $\chi([\operatorname{int} P]) = (-1)^d$. Deduce the Euler-Poincaré formula: if P is a d-dimensional polytope (bounded polyhedron), then $\sum_{k=0}^{d} (-1)^k f_k = 1$, where f_k is the number of k-dimensional faces of P (including the polytope itself).

Preview problems.

1. Let $P \subset \mathbb{R}^d$ be a polyhedron and let $T : \mathbb{R}^d \longrightarrow \mathbb{R}^k$ be a linear transformation. Prove that $T(P)$ is a polyhedron.

2. We know that whenever there is an Euler characteristic, there must be an underlying cohomology theory. What is the underlying cohomology theory for the Euler characteristic in Theorem 1?

   One problem is that the Euler characteristic of Theorem 1 is not a topological invariant: we have $\chi([A]) = 1 \neq -1 = \chi([B])$, where A is a line and B is an open interval. Hence the underlying cohomology theory must somehow distinguish between bounded and unbounded sets.

Remarks: Theorem 1 and its proof is due to H. Hadwiger, see also Section I.7 of [**Ba02**] for more detail.

LECTURE 2
Identities in the Algebra of Polyhedra

What can we do with polyhedra? One important observation is that the image of a polyhedron under a linear transformation is a polyhedron.

Theorem 1. *Let $P \subset \mathbb{R}^d$ be a polyhedron and let $T : \mathbb{R}^d \longrightarrow \mathbb{R}^k$ be a linear transformation. Then $T(P) \subset \mathbb{R}^k$ is a polyhedron. Furthermore, if P is a rational polyhedron and T is a rational linear transformation (that is, the matrix of T is rational), then $T(P)$ is a rational polyhedron.*

The crucial step in the proof. Let us consider the following particular case: $k = d-1$ and T is the projection onto the first $(d-1)$ coordinates: $(x_1, \ldots, x_d) \longmapsto (x_1, \ldots, x_{d-1})$. Suppose that the polyhedron P is defined by a system of linear inequalities:

$$\sum_{j=1}^{d} a_{ij} x_j \le b_i \quad \text{for} \quad i = 1, \ldots, m.$$

Let us look at the coefficients of x_d.

Let $I_+ = \{i : \ a_{id} > 0\}$, $I_- = \{i : \ a_{id} < 0\}$, and $I_0 = \{i : \ a_{id} = 0\}$. Then a point $y = (x_1, \ldots, x_{d-1})$ belongs to $T(P)$ if and only if

$$\sum_{j=1}^{d-1} a_{ij} x_j \le b_j \quad \text{for} \quad i \in I_0, \tag{1}$$

and there exists x_d such that

$$\begin{aligned} x_d &\le \frac{b_i}{a_{id}} - \sum_{j=1}^{d-1} \frac{a_{ij}}{a_{id}} x_j \quad \text{for} \quad i \in I_+ \\ x_d &\ge \frac{b_i}{a_{id}} - \sum_{j=1}^{d-1} \frac{a_{ij}}{a_{id}} x_j \quad \text{for} \quad i \in I_- \end{aligned} \tag{2}$$

Conditions (1) are some linear inequalities needed to describe $T(P)$, but not all of them. We get the complete set of linear inequalities by majorizing every lower bound by every upper bound in (2), see Figure 6:

$$\frac{b_{i_1}}{a_{i_1 d}} - \sum_{j=1}^{d-1} \frac{a_{i_1 j}}{a_{i_1 d}} x_j \ge \frac{b_{i_2}}{a_{i_2 d}} - \sum_{j=1}^{d-1} \frac{a_{i_2 j}}{a_{i_2 d}} x_j \quad \text{for every pair} \quad i_1 \in I_+, i_2 \in I_-.$$

Thus we perform a step of the procedure known as the *Fourier-Motzkin elimination.* □

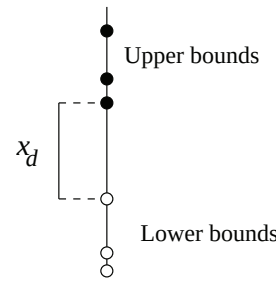


Figure 6. The interval for x_d is obtained by majorizing every lower bound by every upper bound.

Linear transformations preserve linear relations among indicators of polyhedra.

Theorem 2. *Let $T : \mathbb{R}^d \longrightarrow \mathbb{R}^k$ be a linear transformation. Then there exists a linear transformation $\mathcal{T} : \mathcal{P}(\mathbb{R}^d) \longrightarrow \mathcal{P}(\mathbb{R}^k)$ such that $\mathcal{T}[P] = [T(P)]$ for every polyhedron $P \subset \mathbb{R}^d$.*

Proof. Let us define the "kernel" $G : \mathbb{R}^d \times \mathbb{R}^k \longrightarrow \mathbb{R}$ by

$$G(x,y) = \begin{cases} 1 & \text{if } T(x) = y \\ 0 & \text{if } \ T(x) \neq y. \end{cases}$$

Let us choose $f \in \mathcal{P}(\mathbb{R}^d)$. We must define $h \in \mathcal{P}(\mathbb{R}^k)$ such that $\mathcal{T}(f) = h$. To this end, for every $y \in \mathbb{R}^k$, we define a function $g_y \in \mathcal{P}(\mathbb{R}^d)$ by $g_y(x) = G(x,y)f(x)$. One can check that $g_y \in \mathcal{P}(\mathbb{R}^d)$. Hence we can apply the Euler characteristic χ to g_y. We let $h(y) = \chi(g_y)$. Thus we got a function $h : \mathbb{R}^k \longrightarrow \mathbb{R}$. Next, one should check that if $f = [P]$ then $h = [T(P)]$. It follows that if $f \in \mathcal{P}(\mathbb{R}^d)$ then $h \in \mathcal{P}(\mathbb{R}^k)$. We conclude that $\mathcal{T}(f) = h$ defines the required linear transformation. □

We call $G(x,y)$ the "kernel" to underline a certain similarity between our construction and the standard construction of various integral operators between functional spaces in analysis. In analysis, we often construct a linear transformation which transforms a function $f : X \longrightarrow \mathbb{R}$ into a function $h : Y \longrightarrow \mathbb{R}$ by choosing an appropriate kernel $K(x,y) : X \times Y \longrightarrow \mathbb{R}$ and defining $h(y) = \int K(x,y)f(x)\ d\mu(x)$ for some measure μ on X. In polyhedral combinatorics, we can construct some interesting linear operators $\mathcal{A} : \mathcal{P}(\mathbb{R}^d) \longrightarrow \mathcal{P}(\mathbb{R}^k)$ by choosing an appropriate "kernel" $K(x,y) : \mathbb{R}^d \times \mathbb{R}^k \longrightarrow \mathbb{R}$ and letting $\mathcal{A}(f) = h$, where $h(y) = \chi(g_y)$ for $g_y(x) = K(x,y)f(x)$. The similarity is partially explained by the observation that one can think of the Euler characteristic as a finitely-additive measure on $\mathbb{R}^d$ that is a "combinatorialization" of the Lebesgue measure. In analysis, we want to know how large is a given set and the Lebesgue measure tells us that. In polyhedral combinatorics, we just want to know whether a given polyhedron is non-empty, and the Euler characteristic tells that.

It follows from Theorem 2 that whenever we have a linear relation $\sum_{i=1}^m \alpha_i[P_i] = 0$ among the indicator functions of polyhedra, the same relation $\sum_{i=1}^m \alpha_i[T(P_i)] = 0$ holds for their images under a linear transformation. This is obvious for invertible

transformations T but starting to look less obvious for projections, see Figure 7 for a simple example.

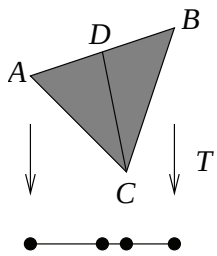


Figure 7. We have $[ABC] = [ACD] + [CBD] - [CD]$ and $[T(ABC)] = [T(ACD)] + [T(CBD)] - [T(CD)]$.

Now we need to take a closer look at polyhedra. Some polyhedra have *vertices*, some don't.

Defintion 1. Let $P \subset \mathbb{R}^d$ be a polyhedron. A point $v \in P$ is called a *vertex* of P if whenever $v = (x+y)/2$ for some $x, y \in P$, we must have $x = y = v$. If v is a point in P, we define the *tangent cone* of P at v as follows:

$$\mathrm{co}(P, v) = \Big\{ x \in \mathbb{R}^d : \quad \epsilon x + (1-\epsilon) v \in P \quad \text{for all sufficiently small} \quad \epsilon > 0 \Big\}.$$

Figure 8 shows what tangent cones may look like.

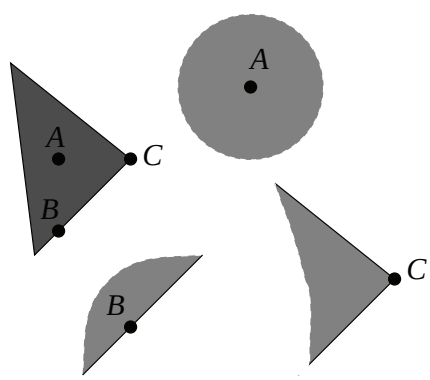


Figure 8. A polyhedron and its tangent cones.

Not all polyhedra have vertices. In fact, a non-empty polyhedron has a vertex if and only if it does not contain a line.

Defintion 2. We say that a polyhedron P contains a *line* if there are points x and y such that $y \neq 0$ and $x + ty \in P$ for all $t \in \mathbb{R}$. Finally, let $\mathcal{P}_0(\mathbb{R}^d) \subset \mathcal{P}(\mathbb{R}^d)$, $\mathcal{P}_0(\mathbb{Q}^d) \subset \mathcal{P}(\mathbb{Q}^d)$ be the subspace spanned by the indicators of (rational) polyhedra that contain lines.

It turns out that modulo polyhedra with lines, every polyhedron is just the sum of its tangent cones.

Theorem 3. *Let $P \subset \mathbb{R}^d$ be a polyhedron. Then there is a $g \in \mathcal{P}_0(\mathbb{R}^d)$ such that*

$$[P] = g + \sum_v [\mathrm{co}(P, v)],$$

where the sum is taken over all vertices v of P. If P is a rational polytope then we can choose $g \in \mathcal{P}_0(\mathbb{Q}^d)$.

A plausible argument. We don't really prove this important theorem, although we come very close. We start by showing that the theorem is not obviously false.

We notice that if P is non-empty and does not contain vertices then P contains a line and hence we can choose $g = [P]$.

Suppose we have been sloppy and included in the sum not only all vertices v of P but also some non-vertices $v \in P$. No harm done: if $v \in P$ is a non-vertex then $\mathrm{co}(P, v)$ contains a line and so we just have to adjust g. This shows that the formula is robust enough.

Suppose that the theorem holds for some polyhedron $P \subset \mathbb{R}^d$ and let $T : \mathbb{R}^d \longrightarrow \mathbb{R}^k$ be a sufficiently generic linear transformation. We claim that the theorem holds for the image $T(P)$. Indeed, by Theorem 2 the transformation T gives rise to the transformation $\mathcal{T}$ on the algebra of polyhedra. Let us apply $\mathcal{T}$ to both sides of the identity. We have $\mathcal{T}[P] = [T(P)]$ and $\mathcal{T}[\mathrm{co}(P, v)] = [T(\mathrm{co}(P, v))] = [\mathrm{co}(T(P), T(v))]$, cf. Review Problem 10.

We have to be somewhat careful with g: we know that g is a linear combination of indicators of polyhedra with lines. If we are unlucky, the kernel of T may "eat up" some of those lines and $\mathcal{T}(g)$ will not lie in $\mathcal{P}_0(\mathbb{R}^k)$. This is the reason why we chose T to be "generic". Thus if we prove the theorem for some "model" polyhedra P, we can extend it (with some care) to polyhedra obtained from P by linear transformations.

Now, we show that the result holds for a *simplex*, which we define as a compact polyhedron $\Delta \subset \mathbb{R}^d$ that is the non-empty intersection of $d+1$ sufficiently generic halfspaces $H_1, \ldots, H_{d+1}$. We notice that

$$[H_1 \cup \ldots \cup H_{d+1}] = [\mathbb{R}^d]$$

and expanding $[H_1 \cup \ldots \cup H_{d+1}]$ by the inclusion-exclusion formula we represent $[\mathbb{R}^d]$ as the alternating sum of the indicators $[H_{i_1} \cap \ldots \cap H_{i_k}]$ of intersections of halfspaces. All such intersections contain lines except for the simplex $\Delta = [H_1 \cap \ldots \cap H_{d+1}]$ itself (the intersection of all $d+1$ halfspaces) and the tangent cones $[H_1 \cap \ldots \cap H_{i-1} \cap H_{i+1} \cap \ldots \cap H_{d+1}]$ (the intersections of all but one halfspace) at the vertices of Δ, see Figure 9.

It follows now that the result holds for all projections of simplices, that is for polytopes (bounded polyhedra). To obtain the formula for a general polyhedron, one needs some structural results about unbounded polyhedra, namely that every unbounded polyhedron is the Minkowski sum of its *recession cone* and a polytope, see Review Problem 11 and Supplementary Problem 3. □

Defintion 3. Let $A \subset \mathbb{R}^d$ be a non-empty set. The set

$$A^\circ = \left\{ y \in \mathbb{R}^d : \ \langle x, y \rangle \leq 1 \quad \text{for all} \quad x \in A \right\}$$

is called the *polar* of A.

It is easy to see that A° is a non-empty closed convex set containing the origin. The *Bipolar Theorem* asserts that $(A^\circ)^\circ = A$ provided A is a closed convex set

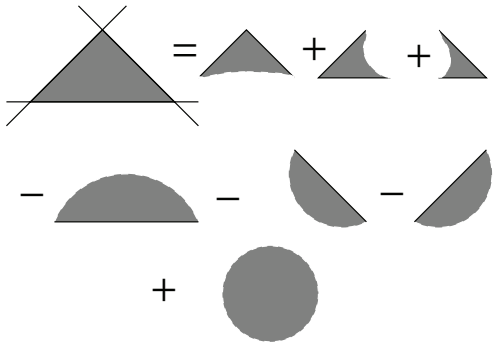


Figure 9. A triangle is the sum of the angles at its vertices minus the half-planes based on its sides plus the whole plane.

containing the origin. One can show that if P is a (rational) polyhedron then P° is a (rational) polyhedron, see Figure 10.

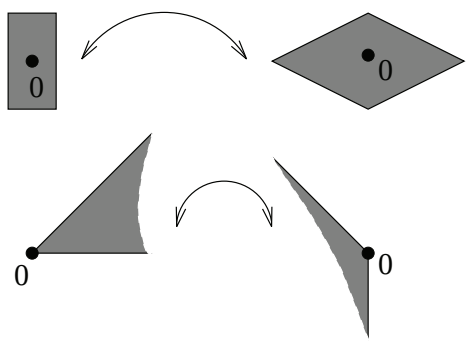


Figure 10. Some (bounded and unbounded) polyhedra and their polars.

It is somewhat surprising that the polarity correspondence preserves linear relations among the indicator functions of polyhedra.

Theorem 4. *There exists linear transformations $\mathcal{D} : \mathcal{P}(\mathbb{R}^d) \longrightarrow \mathcal{P}(\mathbb{R}^d)$, $\mathcal{D} : \mathcal{P}(\mathbb{Q}^d) \longrightarrow \mathcal{P}(\mathbb{Q}^d)$, such that $\mathcal{D}[P] = [P^\circ]$ for every non-empty (rational) polyhedron P.*

The idea of the proof. We define $\mathcal{D}$ as a limit of certain operators $\mathcal{D}_\epsilon$. For $\epsilon > 0$, let us define the kernel $G_\epsilon : \mathbb{R}^d \times \mathbb{R}^d \longrightarrow \mathbb{R}$ by

$$G_\epsilon(x, y) = \begin{cases} 1 & \text{if } \langle x, y\rangle < 1 + \epsilon \\ 0 & \text{otherwise.} \end{cases}$$

For $f \in \mathcal{P}(\mathbb{R}^d)$, $\mathcal{P}(\mathbb{Q}^d)$ and $y \in \mathbb{R}^d$, let $g_{y,\epsilon}(x) = f(x)G_\epsilon(x, y)$. One can check that $g_{y,\epsilon} \in \mathcal{P}(\mathbb{R}^d), \mathcal{P}(\mathbb{Q}^d)$, so we can apply the Euler characteristic χ to $g_{y,\epsilon}$. Let us define $h_\epsilon = \mathcal{D}_\epsilon(f)$ by $h_\epsilon(y) = \chi(g_{y,\epsilon})$. Finally, we define $h = \mathcal{D}(f)$ by $h(y) = \lim_{\epsilon \longrightarrow 0+} h_\epsilon(y)$. One can check then that $\mathcal{D}$ satisfies the desired properties. $\square$

It follows from Theorem 4 that whenever we have a linear identity

$$\sum_{i=1}^{m} \alpha_i [P_i] = 0$$

among the indicator functions of polyhedra, we have the same identity

$$\sum_{i=1}^{m} \alpha_i [P_i^\circ] = 0$$

for the indicator functions of their polars, see Figure 11.

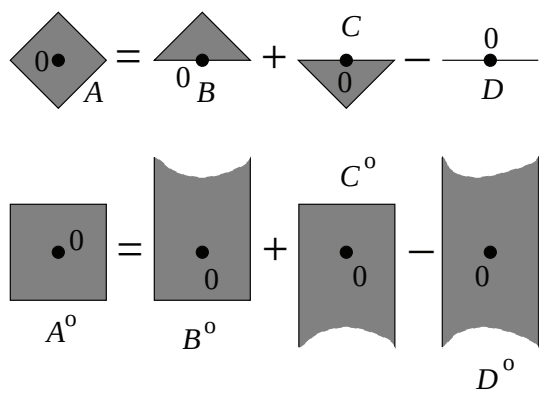


Figure 11. We have $[A] = [B] + [C] - [D]$ and $[A^\circ] = [B^\circ] + [C^\circ] - [D^\circ]$.

An important feature of the polarity transform is that P° contains a line if and only if P lies in an affine hyperplane, that is, not full dimensional. Continuing our analogy with analysis, we can say that in the Euler characteristic based polyhedral combinatorics, the polarity transform $\mathcal{D}$ plays the role akin to that of the Fourier transform in the Lebesgue measure based analysis. An observation in support of this statement can be found in Preview Problem 3.

Problems

Review problems.

1. Complete the proof of Theorem 1.

2. In Theorem 1, suppose that $P \subset \mathbb{R}^d$ is defined by m linear inequalities. Estimate the number of inequalities needed to define $T(P)$.

3. Check the proof of Theorem 2.

4. Let $P \subset \mathbb{R}^d$ be a polyhedron defined by m linear inequalities

$$\sum_{j=1}^{d} a_{ij} x_j \leq b_i \quad \text{for} \quad i = 1, \ldots, m.$$

Let $x \in P$ be a point. We say that the inequality is *active* on x if equality holds at x. Let $a_i = (a_{i1}, \ldots, a_{id})$ be the vector of coefficients of the i-th inequality.

Prove that $v \in P$ is a vertex of P if and only if there are at least d inequalities active on v such that their vectors form a basis of $\mathbb{R}^d$.

5. Prove that a polyhedron has finitely many vertices, if any.

6. Let P be a rational polyhedron and let $v \in P$ be a vertex. Prove that v has rational coordinates.

7. Let P be a polyhedron and let $v \in P$ be a point. Prove that $\mathrm{co}(P, v)$ is the polyhedron defined by the inequalities of P that are active on v.

8. Prove that a non-empty polyhedron has a vertex if and only if it does not contain lines.

9. Prove that $v \in P$ is a vertex of P if and only if $\mathrm{co}(P, v)$ does not contain lines.

10. Let $P \subset \mathbb{R}^d$ is a polyhedron, let $v \in P$ be a point, let $T : \mathbb{R}^d \longrightarrow \mathbb{R}^k$ be a linear transformation, let $Q = T(P)$, and let $u = T(v)$. Prove that $\mathrm{co}(Q, u) = T(\mathrm{co}(P, v))$.

11. Let $P \subset \mathbb{R}^d$ be a non-empty (rational) polyhedron. Let us define the *recession cone* K_P by
$$K_P = \Big\{x \in \mathbb{R}^d : \quad y + tx \in P \quad \text{for all} \quad y \in P \quad \text{and all} \quad t \geq 0\Big\}.$$
Show that K_P is a (rational) polyhedron.

12. Prove that a non-empty polyhedron $P \subset \mathbb{R}^d$ lies in an affine hyperplane if and only if P° contains a line.

13. Let us define a "simpler" version of the kernel G in Theorem 4 by
$$G(x, y) = \begin{cases} 1 & \text{if } \langle x, y\rangle \leq 1 \\ 0 & \text{otherwise} \end{cases}$$
and let $\mathcal{D}'$ be the corresponding operator. Check that if P is a non-empty polyhedron then $\mathcal{D}'([P]) = [Q]$, where
$$Q = \Big\{y \in \mathbb{R}^d : \ \langle x, y\rangle \leq 1 \quad \text{for some} \quad x \in P\Big\}.$$
Compare Q with P°, cf. Definition 3.

Supplementary problems.

1. For sets $A, B \subset \mathbb{R}^d$, we define their *Minkowski sum* $A + B = \big\{x + y : \ x \in A, y \in B\big\}$. Prove that the Minkowski sum of polyhedra is a polyhedron and that the Minkowski sum of rational polyhedra is a rational polyhedron.

2. Prove that there exists a bilinear operation, called convolution, $\star : \mathcal{P}(\mathbb{R}^d) \times \mathcal{P}(\mathbb{R}^d) \longrightarrow \mathcal{P}(\mathbb{R}^d)$ such that $[P]\star[Q] = [P+Q]$ for any two polyhedra $P, Q \subset \mathbb{R}^d$. This gives $\mathcal{P}(\mathbb{R}^d)$ another (more interesting) commutative algebra structure. Note that $[0]$ plays the role of the identity, so $f \star [0] = [0] \star f = f$ for all $f \in \mathcal{P}(\mathbb{R}^d)$.

3. Let $P \subset \mathbb{R}^d$ be a non-empty polyhedron not containing lines and let Q be the convex hull of the set of vertices of P. Prove that P can be represented as the Minkowski sum $P = Q + K_P$, where K_P is the recession cone of P, cf. Review Problem 11.

4. Using Problem 3 above, complete the proof of Theorem 3.

5. Suppose that $P \subset \mathbb{R}^d$ is a bounded polyhedron. Prove that $[P]$ is invertible with respect to the convolution operation $\star$ of Problem 2 above : there exists an $f \in \mathcal{P}(\mathbb{R}^d)$ such that $f \star [P] = [0]$. More precisely, if P is a bounded polyhedron with a non-empty interior $\operatorname{int} P$, we can choose $f = (-1)^d[-\operatorname{int} P]$ (that is, we take the interior of P, reflect it about the origin, and take the indicator of the set we got with the appropriate sign).

6. Let $P \subset \mathbb{R}^d$ be a polyhedron. We say that two points $x, y \in P$ are *equivalent*, if $\operatorname{co}(P, x) = \operatorname{co}(P, y)$. An equivalence class of points in P is just an open face $F \subset P$. For an $x \in F$, we denote $\operatorname{co}(P, x)$ by $\operatorname{co}(P, F)$. Prove the following *Gram-Brianchon Theorem*
$$[P] = \sum_F (-1)^{\dim F} \left[\operatorname{co}(P, F)\right],$$
where the sum is taken over all non-empty faces F of P, including $F = P$.

7. Let $P \subset \mathbb{R}^d$ be a bounded polyhedron (polytope) containing the origin in its interior. For a face F of P, let $P_F = \operatorname{conv}(F, 0)$ be the convex hull of the face F and the origin. Prove that
$$(-1)^{d-1}[P] = \sum_F (-1)^{\dim F}[P_F],$$
where the sum is taken over all faces $F \neq P$ of P, including the empty face (cf. Supplementary Problem 1 to Lecture 1).

8. Prove that the polar of a (rational) polyhedron is a (rational polyhedron) and that $(A^\circ)^\circ = A$ if A is closed, convex, and contains 0.

9. Complete the proof of Theorem 4.

10. Show that if we apply the polarity transform $\mathcal{D}$ to both sides of the identity in Problem 7 above, we get the Gram-Brianchon identity of Problem 6.

Preview problems.

1. A polyhedron $K \subset \mathbb{R}^d$ is called a (polyhedral) *cone* if $0 \in K$ and $\lambda x \in K$ for all $x \in K$ and all $\lambda \geq 0$ (note that the tangent cone of Definition 1 is not necessarily a cone in the sense of this definition, since the vertex of the tangent cone is not necessarily the origin). Prove that if K is a cone then K° is a cone and that $(K^\circ)^\circ = K$.

2. Let $K_1, K_2 \subset \mathbb{R}^d$ be polyhedral cones. Prove that $[K_1 \cap K_2]^\circ = [K_1 + K_2]$, where "+" is the Minkowski sum, see Supplementary Problem 1.

3. Let $\mathcal{D}$ be the transform of Theorem 4 and let $f_1, f_2 \in \mathcal{P}(\mathbb{R}^d)$ be linear combinations of indicator functions of polyhedral cones. Prove that $\mathcal{D}(f_1 f_2) =$

$\mathcal{D}(f_1) \star \mathcal{D}(f_2)$, where $\star$ is the convolution operation from Supplementary Problem 2.

Remarks: For the Fourier-Motzkin elimination (Theorem 1), see Sections I.9 of [**Ba02**] and Sections 1.2-1.3 of [**Zi95**]. A nice exposition of the Euler characteristic and the theory of valuations is given in [**KR97**]. Much of the material of this lecture can be found in [**Ba02**]: Section II.4-5 (vertices of polyhedra), Section IV.1 (polarity), Section VIII.4 (tangent cones). Analogies between integral operators in the classical analysis and valuations are drawn, for example, in [**KP93**].

LECTURE 3

Generating Functions and Cones. Continued Fractions

Generating Functions and Cones

Now we turn to integer points. For an integer point $m = (m_1, \dots, m_d)$, we introduce the monomial $\mathbf{x}^m = x_1^{m_1} \cdots x_d^{m_d}$ in d complex variables $\mathbf{x} = (x_1, \dots, x_d)$. Given a set $S \subset \mathbb{R}^d$, we consider the sum

$$f(S, \mathbf{x}) = \sum_{m \in S \cap \mathbb{Z}^d} \mathbf{x}^m.$$

Our goal is to find a reasonably short expression for this sum as a rational function in $\mathbf{x}$. Our inspiration is the formula

$$\sum_{m=0}^{+\infty} x^m = \frac{1}{1-x} \quad \text{for} \quad |x| < 1.$$

Here is an obvious multivariate generalization of the formula.

Example 1. Let $\mathbb{R}_+^d$ be the non-negative orthant, that is the set of points with all coordinates non-negative. We have

$$\begin{aligned} \sum_{m \in \mathbb{R}_+^d \cap \mathbb{Z}^d} \mathbf{x}^m &= \left(\sum_{m_1=0}^{+\infty} x_1^{m_1} \right) \cdots \left(\sum_{m_d=0}^{+\infty} x_d^{m_d} \right) \\ &= \prod_{i=1}^{d} \frac{1}{1-x_i} \quad \text{provided} \quad |x_i| < 1 \quad \text{for} \quad i = 1, \dots, d. \end{aligned}$$

In general, we say that $f(S, \mathbf{x})$ is defined by a particular rational function if there is a non-empty open set $U \subset \mathbb{C}^d$ such that for all $\mathbf{x} \in U$ the defining series for $f(S, \mathbf{x})$ converges absolutely to that rational function and the convergence is uniform on compact subsets of U. In all cases we encounter, only existence of such an U, but not its precise shape, will be of importance.

Our next step is less obvious. What if the orthant gets somewhat "skewed"?

Defintion 1. Let $u_1, \dots, u_d \in \mathbb{Z}^d$ be linearly independent integer vectors. The *simple rational cone* generated by $u_1, \dots, u_d$ is the set

$$K = K(u_1, \dots, u_d) = \Big\{ \sum_{i=1}^{d} \alpha_i u_i : \quad \alpha_i \geq 0 \quad \text{for} \quad i = 1, \dots, d \Big\}.$$

The *fundamental parallelepiped* of $u_1, \ldots, u_d$ is the set

$$\Pi = \Pi(u_1, \ldots, u_d) = \Big\{ \sum_{i=1}^{d} \alpha_i u_i : \quad 0 \le \alpha_i < 1 \quad \text{for} \quad i = 1, \ldots, d \Big\}.$$

Note that the parallelepiped is "semi-open", see Figure 12.

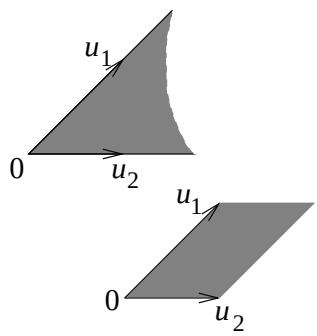


Figure 12. A simple rational cone and its fundamental parallelepiped.

If we replace vectors u_i by their positive integer multiples $k_i u_i$, $i = 1, \ldots, d$, the cone K will not change (although the parallelepiped Π will). This is all the freedom we have in choosing $u_1, \ldots, u_d$ for a given K.

Let us define the dual set of vectors $u_i^*, i = 1, \ldots, d$, by $\langle u_i, u_j^* \rangle = \delta_{ij}$. Then

$$K = \Big\{ x : \quad \langle x, u_i^* \rangle \ge 0 \quad \text{for} \quad i = 1, \ldots, d \Big\},$$

from which it follows that simple rational cones are rational polyhedra.

Theorem 1. *For a simple rational cone $K = K(u_1, \ldots, u_d)$, we have*

$$f(K, \mathbf{x}) = \left(\sum_{m \in \Pi \cap \mathbb{Z}^d} \mathbf{x}^m \right) \prod_{i=1}^{d} \frac{1}{1 - \mathbf{x}^{u_i}}.$$

Proof. The proof consists of the observation that *every* point $m \in K \cap \mathbb{Z}^d$ can be *uniquely* written as $m = m_1 + m_2$, where $m_1 \in \Pi \cap \mathbb{Z}^d$ and m_2 is a non-negative integer combination of $u_1, \ldots, u_d$. Indeed, since u lies in the cone K, it can be written in the form

$$m = \sum_{i=1}^{d} \alpha_i u_i \quad \text{for some real numbers} \quad \alpha_i \ge 0.$$

Let $\lfloor \alpha \rfloor$ denote the largest integer not exceeding α (a.k.a the *integer part* of α) and let $\{\alpha\} = \alpha - \lfloor \alpha \rfloor$ (the *fractional part* of α). Then

$$m_1 = \sum_{i=1}^{d} \{\alpha_i\} u_i \quad \text{and} \quad m_2 = \sum_{i=1}^{d} \lfloor \alpha_i \rfloor u_i.$$

To prove uniqueness, suppose that we have two decompositions $m = m_1 + m_2$ and $m = m_1' + m_2'$, where m_1 and m_2 are integer points from the parallelepiped Π and m_2 and m_2' are non-negative integer combinations of $u_1, \ldots, u_d$. Then we can write

$m_1 - m_1' = m_2' - m_2$, from which $m_1 - m_1'$ is an integer combination of $u_1, \ldots, u_d$. However, since $m_1, m_1' \in \Pi$, we should be able to write

$$m_1 - m_1' = \sum_{i=1}^{d} \beta_i u_i \quad \text{where} \quad -1 < \beta_i < 1 \quad \text{for} \quad i = 1, \ldots, d.$$

Thus we must have $\beta_i = 0$ and $m_1 = m_1'$, $m_2 = m_2'$.

It remains to show that there is some non-empty open set $U \subset \mathbb{C}^d$ of $\mathbf{x}$ for which the series

$$f(K, \mathbf{x}) = \sum_{m \in K \cap \mathbb{Z}^d} \mathbf{x}^m$$

converges absolutely and uniformly on compact subsets of U. Since $u_1, \ldots, u_d$ are linearly independent, we can find a vector $c = (c_1, \ldots, c_d)$, such that $\langle c, u_i \rangle < 0$ for $i = 1, \ldots, d$, where $\langle x, y \rangle = x_1 y_1 + \ldots + x_d y_d$ is the standard scalar product in $\mathbb{R}^d$. We can take, for example, $c = -u_1^* - \ldots - u_d^*$. Let $\mathbf{x}_0 = (e^{c_1}, \ldots, e^{c_d})$. Then for all $\mathbf{x}$ in a sufficiently small neighborhood U of $\mathbf{x}_0$, the series converges as desired. Since the product $\prod_{i=1}^{d} (1 - \mathbf{x}^{u_i})^{-1}$ encodes the sum of $\mathbf{x}^m$ over all m that are non-negative integer combinations of $u_1, \ldots, u_d$ (cf. Example 1), the proof follows. $\square$

Theorem 1 provides us with a finite formula for an infinite series, but there is still something unsatisfactory about it. Namely, the sum over integer points in the fundamental parallelepiped is not very explicit and, although finite, can be quite large. Although the set of integer points lying in the parallelepiped can be complicated, we can tell the number of such points exactly.

Theorem 2. *The number of integer points in the fundamental parallelepiped is equal to the volume of the parallelepiped.*

Sketch of proof. Let Λ be the set of all integer combinations of $u_1, \ldots, u_d$:

$$\Lambda = \Big\{ \sum_{i=1}^{d} \alpha_i u_i : \quad \alpha_i \in \mathbb{Z} \quad \text{for} \quad i = 1, \ldots, d \Big\}.$$

Let us consider all translates $\Pi + u$ with $u \in \Lambda$. We claim that the translations $\Pi + u : \ u \in \Lambda$ cover $\mathbb{R}^d$ without overlapping. The proof can be extracted from the proof of Theorem 1. Let us take a sufficiently large, regular looking region $X \subset \mathbb{R}^d$ (say, a ball of a large radius), and let us count integer points in X. On one hand, we can approximate the number of integer points in X by the volume $\operatorname{vol} X$ of X. One the other hand, the set is covered by roughly $(\operatorname{vol} X)/(\operatorname{vol} \Pi)$ translations of the parallelepiped Π and each translation contains the same number of integer points. Hence we must have $|\Pi \cap \mathbb{Z}^d| = \operatorname{vol} \Pi$. $\square$

We note that the volume of the fundamental parallelepiped of $u_1, \ldots, u_d$ is equal to the absolute value of the determinant of the matrix with the columns $u_1, \ldots, u_d$.

For a simple illustration of Theorem 2, see Figure 13.

This leads us to the following crucial definition.

Defintion 2. Let $u_1, \ldots, u_d \in \mathbb{Z}^d$ be linearly independent vectors and let K be the simple cone generated by $u_1, \ldots, u_d$. We say that K is *unimodular* if the volume of

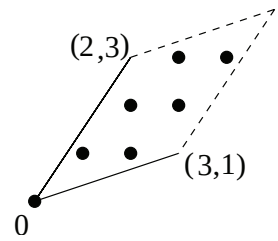


Figure 13. The number of integer points in a fundamental parallelogram is equal to the area of the parallelogram.

the fundamental parallelepiped Π is 1. Equivalently, K is unimodular if the origin is the unique integer point in Π. Equivalently,

$$f(K, \mathbf{x}) = \prod_{i=1}^{d} \frac{1}{1 - \mathbf{x}^{u_i}}.$$

One of our goals is to devise an efficient procedure of *decomposing* a given simple cone into a certain *combination* of unimodular cones. The first non-trivial case is $d = 2$ (every 1-dimensional cone is unimodular) and there such a procedure has long been known.

Continued Fractions

Let us choose a number $a \in \mathbb{R}$. The following procedure produces what is called the *continued fraction* expansion $[a_0; a_1, \ldots, a_n, \ldots]$ of a. First, we write

$$a = \lfloor a \rfloor + \{a\} \quad \text{and let} \quad a_0 = \lfloor a \rfloor.$$

Now, if $\{a\} = 0$, we stop. Otherwise, $0 < \{a\} < 1$, we let $b = 1/\{a\}$, so $b > 1$. We write

$$b = \lfloor b \rfloor + \{b\} \quad \text{and let} \quad a_1 = \lfloor b \rfloor.$$

If $\{b\} = 0$ we stop. Otherwise, we let new $b := 1/\{\text{old } \ b\}$, and continue. In the end, we get the expansion

$$a = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ldots}}}.$$

The expansion can be finite (if a is rational) or infinite (if a is irrational). We define the k-th *convergent* $[a_0; a_1, \ldots, a_k]$ by cutting the expansion at a_k. For example, the 4-th convergent $[a_0; a_2, a_3, a_4]$ is

$$a = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{a_4}}}}.$$

As an example, let us compute the continued fractions expansion of $a = 164/31$:

$$\frac{164}{31} = 5 + \frac{9}{31} = 5 + \cfrac{1}{3 + \cfrac{4}{9}} = 5 + \cfrac{1}{3 + \cfrac{1}{2 + \cfrac{1}{4}}}.$$

Hence $164/31 = [5; 3, 2, 4]$. Now we compute the convergents:

$$[5; 3, 2] = 5 + \cfrac{1}{3 + \cfrac{1}{2}} = \frac{37}{7}, \quad [5; 3] = 5 + \frac{1}{3} = \frac{16}{3}, \quad [5] = \frac{5}{1}.$$

Computing $f(K, \mathbf{x})$ for 2-dimensional Cones

Continued fractions can be applied to obtain short formulas for $f(K, \mathbf{x})$, where $K \subset \mathbb{R}^2$ is a simple cone. Instead of developing a comprehensive theory, we give one computational example.

Let $K \subset \mathbb{R}^2$ be the cone generated by the vectors $(1, 0)$ and $(31, 164)$. The volume of the fundamental parallelepiped is 164, so the formula for $f(K, \mathbf{x})$ provided by Theorem 1 would contain a sum of 164 monomials. We will find a shorter formula, and, moreover, will compute it by hand.

First, we compute the continued fraction expansion of $164/31 = [5; 3, 2, 4]$ and the convergents $[5] = 5/1$, $[5; 3] = 16/3$, $[5; 3, 2] = 37/7$, cf. the above example.

Now, we do some "surgery" on cones. Let us consider the following cones, given by their generators

K_0	generated by	$(1, 0)$	and	$(0, 1)$;	
K_1	generated by	$(1, 0)$	and	$(1, 5)$;	
K_2	generated by	$(1, 0)$	and	$(3, 16)$;	
K_3	generated by	$(1, 0)$	and	$(7, 37)$;	and, finally,
K_4	generated by	$(1, 0)$	and	$(31, 164)$.	

We observe that K_0 is a unimodular cone with

$$f(K_0, \mathbf{x}) = \frac{1}{(1 - x_1)(1 - x_2)},$$

while we are trying to compute $f(K_4, \mathbf{x})$.

The crucial observation is that to pass from K_i to K_{i+1} we have either to "cut" from K_i a unimodular cone (if i is even) or to "paste" to K_i a unimodular cone (i is odd), see Figure 14.

Hence, starting with K_0, we

cut the unimodular cone generated by	$(0, 1)$	and	$(1, 5)$;
paste the unimodular cone generated by	$(1, 5)$	and	$(3, 16)$;
cut the unimodular cone generated by	$(3, 16)$	and	$(7, 37)$;
paste the unimodular cone generated by	$(7, 37)$	and	$(31, 164)$

to finally get K_4. Taking into account "boundary effects" (when we cut and paste, some points on the boundary get double-counted), which, luckily, cancel each other,

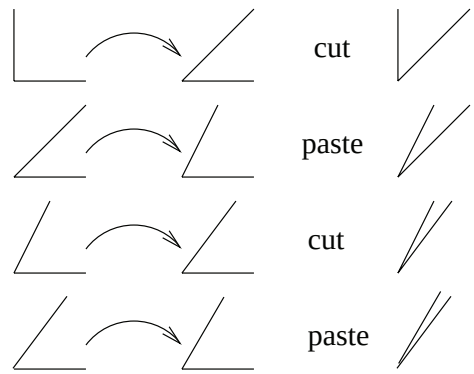


Figure 14. Cutting and pasting unimodular cones.

we get:

$$\begin{aligned} f(K,\mathbf{x}) =& \frac{1}{(1-x_1)(1-x_2)} - \frac{1}{(1-x_2)(1-x_1x_2^5)} + \frac{1}{1-x_1x_2^5} \\ &+ \frac{1}{(1-x_1x_2^5)(1-x_1^3x_2^{16})} - \frac{1}{1-x_1x_2^5} \\ &- \frac{1}{(1-x_1^3x_2^{16})(1-x_1^7x_2^{37})} + \frac{1}{1-x_1^7x_2^{37}} \\ &+ \frac{1}{(1-x_1^7x_2^{37})(1-x_1^{31}x_2^{164})} - \frac{1}{1-x_1^7x_2^{37}}, \end{aligned}$$

so finally,

$$\begin{aligned} f(K,\mathbf{x}) =& \frac{1}{(1-x_1)(1-x_2)} - \frac{1}{(1-x_2)(1-x_1x_2^5)} + \frac{1}{(1-x_1x_2^5)(1-x_1^3x_2^{16})} \\ &- \frac{1}{(1-x_1^3x_2^{16})(1-x_1^7x_2^{37})} + \frac{1}{(1-x_1^7x_2^{37})(1-x_1^{31}x_2^{164})}. \end{aligned}$$

The formula is reasonably short.

Given an arbitrary 2-dimensional rational cone generated by $u_1, u_2 \in \mathbb{Z}^2$, we can always change the coordinates by applying a linear transformation which preserves $\mathbb{Z}^2$ so that u_2 becomes equal to $(1,0)$. Suppose that $u_1 = (q,p)$ for integers p and $q > 0$. To compute the generating function $f(K,\mathbf{x})$, we compute the continued fraction expansion of p/q and obtain K by cutting and pasting the unimodular cones computed from the convergents of p/q. If the k-th convergent is p_k/q_k, we cut or paste, depending on the parity of $k > 1$, the cone generated by (q_k, p_k) and (q_{k-1}, p_{k-1}), which is always unimodular, see Review Problems 4 and 5.

The Computational Complexity

Let $K \subset \mathbb{R}^2$ be the cone generated by $u_1 = (1,0)$ and $u_2 = (q,p)$ as above. The fundamental parallelepiped of K contains $|p|$ points, so if we compute $f(K,\mathbf{x})$ by the formula of Theorem 1, the resulting rational function will contain $|p|$ terms. If, instead, we use the continued fractions method, we get an expression for $f(K,\mathbf{x})$ containing about $\log(\min(|p|,|q|) + O(1)$ terms. For large $|p|$, the difference is quite significant. Looking more closely, we observe that to *define* the cone K, that is,

to write the coordinates of its generators, we need about $\log|p| + \log|q| + O(1)$ bits (or digits) since to write an integer a we need about $\log|a| + O(1)$ bits (or digits). Thus we say that the *input size* of the problem of computing $f(K, \mathbf{x})$ is about $\log|p| + \log|q| + O(1)$. The number of operations required to compute $f(K, \mathbf{x})$ via continued fractions is about $O(\log^2|p| + \log^2|q| + 1)$, that is, bounded by a *polynomial* in the input size. In contrast, the number operations required to compute $f(K, \mathbf{x})$ via Theorem 1 (and even to write down the answer) is *exponential* in the input size of K. In Lecture 5, for any dimension d (fixed in advance), we present a polynomial time algorithm, which, given a rational cone $K \subset \mathbb{R}^d$ as an input, computes $f(K, \mathbf{x})$ as a rational function.

Problems

Review problems.

1. Check the proof of Theorem 1.
2. Make the proof of Theorem 2 rigorous.
3. Let K be the 2-dimensional simple cone generated by $u_1 = (1, 0)$ and $u_2 = (1, n)$ for some positive integer n. Compute $f(K, \mathbf{x})$.
4. Let $[a_0; a_1, \ldots, a_n \ldots]$ be the continued fraction expansion of a real number a and let $p_k/q_k = [a_0; a_1, \ldots, a_k]$ be the k-th convergent (we assume that p_k and q_k are coprime). Prove that for $k \geq 2$
$$p_k = a_k p_{k-1} + p_{k-2} \quad \text{and} \quad q_k = a_k q_{k-1} + q_{k-2}.$$
Deduce that
$$p_{k-1}q_k - p_k q_{k-1} = (-1)^{k-1} \quad \text{for} \quad k \geq 0.$$
5. Justify the procedure of computing $f(K, \mathbf{x})$ for the cone K generated by $(1, 0)$ and (q, p) via continued fractions.
6. Let $K \subset \mathbb{R}^d$ be the set defined by
$$K = \Big\{x \in \mathbb{R}^d : \quad \langle u_i, x\rangle \leq 0 \quad \text{for} \quad i = 1, \ldots, d\Big\}$$
for some linearly independent vectors $u_1, \ldots, u_d \in \mathbb{Z}^d$. Prove that K is a simple rational cone.
7. Let $u_1, \ldots, u_d \in \mathbb{Z}^d$ be linearly independent vectors and let $u_1^*, \ldots, u_d^*$ be defined by $\langle u_i, u_j^*\rangle = \delta_{ij}$. Prove that $u_1^*, \ldots, u_d^* \in \mathbb{Z}^d$ if and only if the cone K generated by $u_1, \ldots, u_d$ is unimodular (we assume that $u_1, \ldots, u_d$ are the minimal generators of K).

Supplementary problems.

1. Let $u_1, \ldots, u_d$ be linearly independent vectors in $\mathbb{Z}^d$. Let K be the cone generated by $u_1, \ldots, u_d$ and let
$$\operatorname{int} K = \Big\{\sum_{i=1}^{d} \alpha_i u_i : \quad \alpha_i > 0 \quad \text{for} \quad i = 1, \ldots, d\Big\}$$

be the interior of K. Let

$$\overline{\Pi} = \Big\{ \sum_{i=1}^{d} \alpha_i u_i : \quad 0 < \alpha_i \leq 1 \quad \text{for} \quad i = 1, \ldots, d \Big\}.$$

Prove that

$$f(\operatorname{int} K, \mathbf{x}) = \left(\sum_{m \in \overline{\Pi} \cap \mathbb{Z}^d} \mathbf{x}^m \right) \prod_{i=1}^{d} \frac{1}{1 - \mathbf{x}^{u_i}}.$$

Deduce the *reciprocity relation* $f(\operatorname{int} K, \mathbf{x}^{-1}) = (-1)^d f(K, \mathbf{x})$.

2. Deduce from Theorem 2 the following *Pick's* Theorem: if $P \subset \mathbb{R}^2$ is a convex polygon with integer vertices and non-empty interior, then the number of integer points in P is equal to the area of P plus half of the number of integer points on the boundary of P plus 1, see Figure 15.

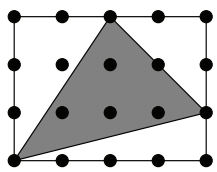


Figure 15. The number of integer points in the triangle (8) is equal to the area of the triangle (5) plus half of the number of integer points on the boundary (2) plus 1.

Preview problems.

1. Let $K \subset \mathbb{R}^d$ be a unimodular cone generated by integer vectors $u_1, \ldots, u_d$ and let $K + v$ be the translation of K by a rational vector $v \in \mathbb{Q}^d$. Prove that

$$f(K + v, \mathbf{x}) = \mathbf{x}^w \prod_{i=1}^{d} \frac{1}{1 - \mathbf{x}^{u_i}} \quad \text{with} \quad w = \sum_{i=1}^{d} \lceil \langle v, u_i^* \rangle \rceil u_i,$$

where $u_1^*, \ldots, u_d^*$ are defined by $\langle u_i^*, u_j \rangle = \delta_{ij}$.

2. Construct an efficient (polynomial time) algorithm to *sample* a random integer point in a given fundamental parallelepiped Π from the uniform distribution on $\Pi \cap \mathbb{Z}^d$ (the dimension d needs not to be fixed in advance).

Remarks: For generating functions and rational cones, see Section 4.6 of [**St97**] and Section VIII.1 of [**Ba02**]. A classical reference for continued fractions is [**Kh97**]. For the theory of computational complexity, see [**Pa94**].

LECTURE 4
Rational Polyhedra and Rational Functions

Let $P \subset \mathbb{R}^d$ be a rational polyhedron. Our goal is to understand the generating function

$$f(P, \mathbf{x}) = \sum_{m \in P \cap \mathbb{Z}^d} \mathbf{x}^m.$$

Previously, we discussed what happens if $P = K$ is a simple rational cone. Step by step, we go to larger and larger classes of polyhedra.

Defintion 1. A rational polyhedron $K \subset \mathbb{R}^d$ is called a *rational cone* provided $0 \in K$ and $\lambda x \in K$ for every $x \in K$ and every $\lambda \geq 0$. Equivalently, K is a rational cone if K can be defined by a system of finitely many homogeneous linear inequalities with integer coefficients:

$$K = \Big\{ x : \quad \sum_{j=1}^{d} a_{ij} x_j \leq 0 \quad \text{for} \quad i = 1, \ldots, m \Big\}.$$

If 0 is a vertex of K, the cone is called *pointed*.

The first real difference between the concept of a rational cone and that of a simple rational cone transpires at $d = 3$. While simple rational cones in $\mathbb{R}^d$ are defined by exactly d inequalities, non-simple rational cones may require more or fewer inequalities.

Theorem 1. *Let $K \subset \mathbb{R}^d$ be a pointed rational cone. Then $f(K, \mathbf{x})$ is a rational function in $\mathbf{x}$ of the type*

$$f(K, \mathbf{x}) = \sum_{i=1}^{n} \frac{p_i(\mathbf{x})}{(1 - \mathbf{x}^{u_{i1}}) \cdots (1 - \mathbf{x}^{u_{id}})},$$

where $p_i(\mathbf{x})$ are Laurent polynomials in $\mathbf{x}$ and $u_{ij} \in \mathbb{Z}^d$ are non-zero vectors.

A plausible argument. Since 0 is a vertex of K, there is a vector $c \in \mathbb{R}^d$, $c = (c_1, \ldots, c_d)$ such that $\langle c, x \rangle < 0$ for all $x \in K \setminus \{0\}$. Now, for any $\mathbf{x}$ from a sufficiently small neighborhood U of $x_0 = (e^{c_1}, \ldots, e^{c_d})$ the series $\sum_{m \in K \cap \mathbb{Z}^d} \mathbf{x}^m$ converges absolutely and uniformly on compact subsets of U. It seems intuitively obvious and indeed correct that K can be cut into simple rational cones, so we can deduce the formula for $f(K, \mathbf{x})$ from Theorem 1, Lecture 3, and the inclusion-exclusion formula. It takes some time though to make the proof rigorous, cf. Figure 16. $\square$

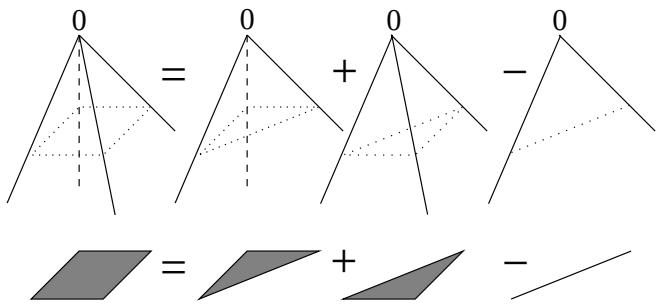


Figure 16. The indicator of a cone with a square base can be written as the sum of the indicators of cones with triangular bases minus the indicator of a flat cone based on the interval.

Next, we consider an arbitrary rational polyhedron with a vertex.

Theorem 2. *Let $P \subset \mathbb{R}^d$ be a rational polyhedron with a vertex (equivalently, a non-empty rational polyhedron without lines). Then $f(P, \mathbf{x})$ is a rational function*

$$f(P, \mathbf{x}) = \sum_{i=1}^{n} \frac{p_i(\mathbf{x})}{(1 - \mathbf{x}^{u_{i1}}) \cdots (1 - \mathbf{x}^{u_{id}})},$$

where $p_i(\mathbf{x})$ are Laurent polynomials in $\mathbf{x}$ and $u_{ij} \in \mathbb{Z}^d$ are non-zero vectors.

Sketch of proof. The idea is to consider P as a section of a pointed rational cone $K \subset \mathbb{R}^{d+1}$. We think of $\mathbb{R}^d$ as the affine hyperplane $x_{d+1} = 1$ in $\mathbb{R}^{d+1}$. Given the inequalities defining P,

$$\sum_{j=1}^{d} a_{ij} x_j \leq b_i \quad \text{for} \quad i = 1, \ldots, m,$$

we define K by the inequalities

$$\sum_{j=1}^{d} a_{ij} x_j - b_i x_{d+1} \leq 0, \quad x_{d+1} \geq 0.$$

Clearly, K is a rational cone and P is the section of K by the affine hyperplane $x_{d+1} = 1$, cf. Figure 17. One can also prove that K is pointed via the following chain of implications:

$$P \text{ contains no lines} \quad \Longrightarrow \quad K \text{ contains no lines} \quad \Longrightarrow \quad K \text{ is pointed.}$$

Finally, we obtain $f(P, \mathbf{x})$ by differentiating with respect to x_{d+1}:

$$f(P, \mathbf{x}) = \frac{\partial}{\partial x_{d+1}} f(K, \mathbf{x}) \quad \text{evaluated at} \quad x_{d+1} = 0.$$

□

Suppose now that P is a rational polyhedron with lines. In this case, the series $\sum_{m \in P \cap \mathbb{Z}^d} \mathbf{x}^m$ does not converge anywhere. As we hinted in the introductory examples of Lecture 1, we want to define $f(P, \mathbf{x}) \equiv 0$ in this case. Quite surprisingly, this naive solution works just fine. The following remarkable result was proved by J. Lawrence, and, independently, by A. Khovanski and A. Pukhlikov in early 1990's.

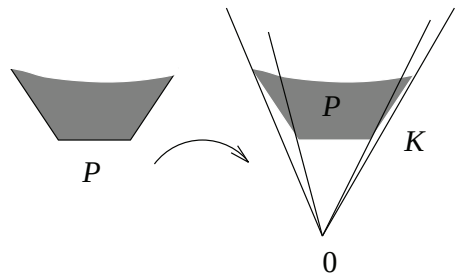


Figure 17. Representing a d-dimensional polyhedron P as a hyperplane section of a $(d+1)$-dimensional cone K.

Theorem 3. *There exists a map*

$$\mathcal{F}: \quad \mathcal{P}(\mathbb{Q}^d) \longrightarrow \mathbb{C}(\mathbf{x})$$

from the algebra of rational polyhedra to the ring of rational functions in d variables $\mathbf{x} = (x_1, \ldots, x_d)$ such that

1. *The map $\mathcal{F}$ is a valuation, that is, a linear transformation,*
2. *If $P \subset \mathbb{R}^d$ is a rational polyhedron without lines then $\mathcal{F}[P] = f(P, \mathbf{x})$ is the rational function such that*
$$f(P, \mathbf{x}) = \sum_{m \in P \cap \mathbb{Z}^d} \mathbf{x}^m$$
for all $\mathbf{x}$ such that the series converges absolutely;
3. *If $P \subset \mathbb{R}^d$ is a polyhedron containing a line then $\mathcal{F}[P] = 0$.*

Sketch of proof. We know how to define $\mathcal{F}$ on the indicators $[P]$ of rational polyhedra P without lines, as in Part (2) of the theorem. Our proof consists of two steps:

the first step is to show that $\mathcal{F}$ can be extended to a valuation on $\mathcal{P}(\mathbb{Q}^d)$;

the second step is to show that once we extended $\mathcal{F}$ to a valuation, we necessarily have $\mathcal{F}[P] = 0$ for rational polyhedra P with lines.

It is clear how we *should* extend $\mathcal{F}$ onto polyhedra with lines (it is not yet clear that we *can*). Any rational polyhedron P can be cut into rational polyhedral pieces P_i without lines, so we *should* compute $\mathcal{F}[P]$ from $\mathcal{F}[P_i] = f(P_i, \mathbf{x})$ via the inclusion-exclusion formula. The problem is, of course, to show that this extension is not self-contradictory. This, in turn, reduces to proving that whenever we have a finite linear dependence

$$\sum_{i \in I} \alpha_i [P_i] = 0 \tag{1}$$

of indicators of rational polyhedra P_i without lines, we must have the same linear dependence

$$\sum_{i \in I} \alpha_i f(P_i, \mathbf{x}) = 0 \tag{2}$$

of their generating functions. Suppose for a moment that in (1) there exists a non-empty open set $U \subset \mathbb{C}^d$ such that for $\mathbf{x} \in U$, each of the series $\sum_{m \in P_i \cap \mathbb{Z}^d} \mathbf{x}^m$ converges absolutely to $f(P_i, \mathbf{x})$. Then (2) follows by a standard argument from analysis. The problem is that there may not be a single set U which works for

all polyhedra P_i in (1). To handle this difficulty, we break the global identity (1) into small "local" pieces, prove (2) for every such piece and then "glue" the global identity (2) from the local pieces.

Let us choose a representation

$$[\mathbb{R}^d] = \sum_{j \in J} \beta_j [Q_j],$$

where $\{Q_j\}$ is a finite family of rational polyhedra without lines and β_j are numbers. One way to obtain the representation is to cut $\mathbb{R}^d$ by the coordinate hyperplanes and express $[\mathbb{R}^d]$ as a linear combination of indicators of coordinate orthants and their intersections using the inclusion-exclusion formula. Multiplying the above formula by $[P_i]$, we get

$$[P_i] = \sum_{j \in J} \beta_j [P_i \cap Q_j] \quad \text{for all} \quad i \in I.$$

Let us fix some $i \in I$. Then P_i is a rational polyhedron without lines and $P_i \cap Q_j$ are rational polyhedral pieces of P_i. Therefore, there is a non-empty open set $U_i \subset \mathbb{C}^d$ such that for all $\mathbf{x} \in U_i$ all the series defining $f(P_i, x)$ and $f(P_i \cap Q_j)$ converge and so we have the identity

$$f(P_i, \mathbf{x}) = \sum_{j \in J} \beta_j f(P_i \cap Q_j, \mathbf{x}) \quad \text{for all} \quad i \in I. \tag{3}$$

Now, let us fix some $j \in J$. Multiplying (1) by Q_j, we get

$$\sum_{i \in I} \alpha_i [P_i \cap Q_j] = 0 \quad \text{for all} \quad j \in J.$$

Again, all $P_i \cap Q_j$ are rational polyhedral pieces of a rational polyhedron Q_j without lines, and, since we can find a single domain $U_j \subset \mathbb{C}^d$ for which all the relevant series converge, we get

$$\sum_{i \in I} \alpha_i f(P_i \cap Q_j, \mathbf{x}) = 0 \quad \text{for all} \quad j \in J. \tag{4}$$

From (3) and (4) we get (2). This completes the first step of the proof.

Thus we are able to extend $\mathcal{F}$ to a valuation on $\mathcal{P}(\mathbb{Q}^d)$. It remains to prove Part (3) of the Theorem. One can show that if P is a rational polyhedron with lines, then there exists a non-zero $m \in \mathbb{Z}^d$ such that $P + m = P$ (there is a non-zero integer translation of P which maps P onto itself). On the other hand, from elementary analysis we deduce that we must have $f(P + m, \mathbf{x}) = \mathbf{x}^m f(P, \mathbf{x})$ for any rational polyhedron P without lines. By linearity, $\mathcal{F}[P + m] = \mathbf{x}^m \mathcal{F}[P]$ for *any* rational polyhedron P. Hence, if $P + m = P$, we must have $\mathcal{F}[P] = \mathbf{x}^m \mathcal{F}[P]$, from which $\mathcal{F}[P] = 0$. $\square$

Suppose that $P \subset \mathbb{R}^d$ is a rational polyhedron without lines (maybe even bounded) and that we want to compute a short formula for the rational generating function $f(P, \mathbf{x})$. Theorem 3 allows us to employ various identities in the algebra $\mathcal{P}(\mathbb{Q}^d)$ of rational polyhedra, including those that involve polyhedra with lines. In particular, we get the following result, first obtained by M. Brion in 1988.

Theorem 4. *Let $P \subset \mathbb{R}^d$ be a rational polyhedron with vertices. Then*

$$f(P, \mathbf{x}) = \sum_v f\big(\mathrm{co}(P, v), \mathbf{x}\big),$$

where the sum is taken over all vertices v of P and $\mathrm{co}(P, v)$ *is the tangent cone of P at v.*

Proof. The proof follows by Theorem 3 of this lecture and Theorem 3 of Lecture 2. □

Note that the tangent cone $\mathrm{co}(P, v)$ is not a rational cone per se, but a rational translation of a rational cone.

Example 1. Let $d = 1$ and let P be the interval $[0, n] \subset \mathbb{R}^1$ for some positive integer n. Then P is a rational polyhedron with the vertices at 0 and n, see Figure 18. The tangent cone $\mathrm{co}(P, 0)$ at 0 is the ray $[0, +\infty)$ and the corresponding generating function is

$$\sum_{m=0}^{+\infty} x^m = \frac{1}{1-x}.$$

The tangent cone $\mathrm{co}(P, n)$ at n is the ray $(-\infty, n]$ and the corresponding generating function is

$$\sum_{m=-\infty}^{n} x^m = \frac{x^n}{1-x^{-1}} = \frac{-x^{n+1}}{1-x}.$$

Note that there is not a single value of x for which both series converge. Nevertheless, Theorem 4 predicts that the sum of the two functions gives the generating function for P:

$$\sum_{m=0}^{n} x^m = \frac{1}{1-x} - \frac{x^{n+1}}{1-x},$$

which is indeed the case.

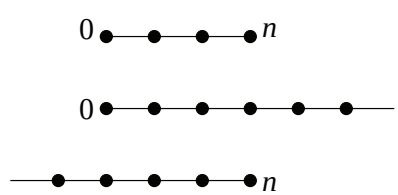


Figure 18. An interval and its tangent cones.

Problems

Review problems

1. Complete the proof of Theorem 2.
2. Let $P \subset \mathbb{R}^d$ be a rational polyhedron with a line. Prove that there exists a non-zero vector $m \in \mathbb{Z}^d$ such that $P + m = P$.
3. Complete the proof of Theorem 3.
4. Check Theorem 4 for the triangle in the plane with the vertices $(0, 0)$, $(0, 1)$, and $(1, 0)$.

A supplementary problem

1. Let $K \subset \mathbb{R}^d$ be a pointed rational cone with non-empty interior $\operatorname{int} K$. Prove the *reciprocity relation* $f(\operatorname{int} K, \mathbf{x}^{-1}) = (-1)^d f(K, \mathbf{x})$.

non-empty interior $\operatorname{int} K$.

Preview problems

1. Prove that the polar of a unimodular cone is a unimodular cone.
2. Let $K \subset \mathbb{R}^2$ be the cone generated by $u_1 = (1, 0)$ and $u_2 = (q, p)$ for some positive integers p and q. Compare the following two ways of computing $f(K, \mathbf{x})$. The first way is the continued fractions method of Lecture 3. The second way is as follows: consider the polar K° (check that K° is the cone generated by $(-p, q)$ and $(0, -1)$). Represent $[K^\circ]$ as a linear combination of the indicators of unimodular cones using the continued fractions method. Apply Theorem 4 of Lecture 1 to obtain a unimodular decomposition of $K = (K^\circ)^\circ$. Compute $f(K, \mathbf{x})$ from that decomposition. What kind of identities do we get for $f(K, \mathbf{x})$?

   This question was asked by one of the attendees.

Remarks: Theorem 3 is proved in [**La91**] and, independently, in [**KP92**]. The first proof of Theorem 4 [**Br88**] uses algebraic geometry. For the material of this lecture, see Sections VIII.3–4 of [**Ba02**] (Theorems 3 and 4), Section 4.6 of [**St97**], and [**B+05**] (generating functions for rational cones and the reciprocity relation).

LECTURE 5
Computing Generating Functions Fast

We discuss *how* to compute the generating function $f(P, \mathbf{x})$ fast, but before we do that, we discuss *why* we want to compute it and what *fast* means.

Why Do We Need Generating Functions?

Let $P \subset \mathbb{R}^d$ be a bounded rational polyhedron (rational polytope). For a variety of reasons, we need to compute the number $|P \cap \mathbb{Z}^d|$ of integer points in P (the counting problem). If we are able to compute the generating function

$$f(P, \mathbf{x}) = \sum_{m \in \mathbb{Z}^d} \mathbf{x}^m,$$

which is just a Laurent polynomial in $\mathbf{x}$, we can get the number of integer points $|P \cap \mathbb{Z}^d|$ by substituting $\mathbf{x} = (1, \ldots, 1)$. Our technique allows us to compute $f(P, \mathbf{x})$ as a reasonably short rational function of the type

$$f(P, \mathbf{x}) = \sum_i \frac{p_i(\mathbf{x})}{(1 - \mathbf{x}^{u_{i1}}) \cdots (1 - \mathbf{x}^{u_{id}})},$$

where $p_i(\mathbf{x})$ are Laurent polynomials in $\mathbf{x}$. This seems to pose a little problem since $\mathbf{x} = (1, \ldots, 1)$ is a pole of every fraction. Nevertheless, the poles cancel each other, as in the model example

$$\sum_{m=0}^{n} x^m = \frac{1}{1-x} - \frac{x^{n+1}}{1-x}.$$

We deal with singularities by approaching the point $(1, \ldots, 1)$ via some curve and computing the appropriate limit. One of the standard choices is the curve $\mathbf{x}(t) = (e^{tc_1}, \ldots, e^{tc_d})$, where $c = (c_1, \ldots, c_d)$ is a sufficiently generic vector: we need $\langle c, u_{ij} \rangle \neq 0$ for all i, j. Then $\mathbf{x}(0) = (1, \ldots, 1)$ and the limit $f(P, \mathbf{x}(t))$ as $t \longrightarrow 0$ can be computed by using standard analysis techniques.

Generating functions help to solve *integer programming problems*, that is the problems of optimizing a given linear function on the set $P \cap \mathbb{Z}^d$ of integer points in a given rational polytope. In short, generating functions $f(P, \mathbf{x})$ encode all the information about the set of integer points in P in a compact form. One remarkable fact is that to find a short formula for $f(P, \mathbf{x})$ for a *bounded* polyhedron, we employ the full power of the algebra $\mathcal{P}(\mathbb{Q}^d)$ and identities in the algebra involving *unbounded* polyhedra and even polyhedra with lines (Theorems 3 and 4 of Lecture 4).

What "Fast" and "Short" Means

We mentioned more than once that we want to compute the generating function $f(P, \mathbf{x})$ "fast" and that we want it in a "reasonably short" form. The exact meaning of these words is understood through the theory of computational complexity.

The polyhedron P is defined by a system of linear inequalities

$$\sum_{j=1}^{d} a_{ij} x_j \leq b_i, \quad i = 1, \ldots, n.$$

The *input size* of P is the number of bits needed to write down the inequalities, assuming that a_{ij} and b_i are integers written in the binary system. For example, to write an integer a, we need about $\log |a| + O(1)$ bits. Thus we say that the algorithm for computing $f(P, \mathbf{x})$ is reasonably fast and the resulting formula is reasonably short if the time we need to compute $f(P, \mathbf{x})$ and the space we need to write it down grows only modestly when the input size of P grows. More precisely, we say that we have a *polynomial time* algorithm for a particular class of rational polyhedra if there is a polynomial *poly* such that the running time of the algorithm on every polyhedron P from the class does not exceed *poly*(input size of P). One example of a polynomial time algorithm is provided by the continued fraction method for computing $f(K, \mathbf{x})$ where K is a 2-dimensional rational cone, see Lecture 3.

It is probably hopeless to search for a polynomial time algorithm in the class of *all* rational polyhedra. However, once the dimension d is fixed such algorithms exist.

Theorem 1. *Let us fix d. Then there exists a polynomial time algorithm, which, given a rational polyhedron $P \subset \mathbb{R}^d$, computes the generating function $f(P, \mathbf{x})$ in the form*

$$f(P, \mathbf{x}) = \sum_i \alpha_i \frac{\mathbf{x}^{v_i}}{(1 - \mathbf{x}^{u_{i1}}) \cdots (1 - \mathbf{x}^{u_{id}})},$$

where $\alpha_i \in \{-1, 1\}$, $v_i \in \mathbb{Z}^d$, and $u_{ij} \in \mathbb{Z}^d \setminus \{0\}$ for all i, j.

Since the running time of the algorithm includes the time needed to write down the output, the space needed to write down $f(P, \mathbf{x})$ is also bounded by a polynomial in the input size.

There exist several versions of the main algorithm behind Theorem 1. Different versions have different advantages under different circumstances. Moreover, the algorithm of Theorem 1 appears to be *practical*. It has been implemented (in fact, by at least two independent groups). The main procedure behind Theorem 1 is a certain *unimodular cone decomposition*. We sketch it below.

Preliminaries

The main result we need is Minkowski's *Convex Body Theorem*. Let $A \subset \mathbb{R}^d$ be a set, such that

(1) A is convex, that is, for every two points $x, y \in A$, the interval $[x, y] = \{\alpha x + (1 - \alpha) y : \ 0 \leq \alpha \leq 1\}$ also lies in A;

(2) A is symmetric about the origin, that is, for every $x \in A$, the point $-x$ also lies in A;

(3) A has a sufficiently large volume: $\operatorname{vol} A > 2^d$.

(4) Moreover, if A is compact, we may assume that $\operatorname{vol} A \geq 2^d$.

Minkowski's Convex Body Theorem asserts that A necessarily contains a non–zero integer point. Here is the idea of the proof: consider $X = \{x/2 : \ x \in A\}$, so that $\operatorname{vol} X > 1$. Consider the set of all integer translates $X + u : \ u \in \mathbb{Z}^d$. Argue that some two different translates must overlap: $(X + u_1) \cap (X + u_2) \neq \emptyset$. Deduce that there is a non-zero lattice point in A.

If A is a rational polyhedron, then such a non-zero integer point in A can be found efficiently, though we don't discuss how.

The unimodular decomposition of a cone

Let $K \subset \mathbb{R}^d$ be a simple rational cone generated by linearly independent vectors $u_1, \ldots, u_d \in \mathbb{Z}^d$. Our goal is to construct unimodular cones K_i such that

$$[K] = \sum_i \alpha_i [K_i] \quad + \quad \text{indicators of lower-dimensional cones}$$

and $\alpha_i \in \{-1, 1\}$. The algorithm runs in polynomial time if the dimension d fixed.

Let Π be the fundamental parallelepiped of K. Then $\operatorname{vol} \Pi$ is a positive integer and $\operatorname{vol} \Pi = 1$ if and only if K is unimodular. Let us call $\operatorname{vol} \Pi$ the *index* of K and denote it $\operatorname{ind} K$. Thus $\operatorname{ind} K$ measures how far K is from being unimodular. We will iterate a certain procedure which gradually reduces the index of K.

Let

$$A = \Big\{ \sum_{i=1}^d \alpha_i u_i : \quad |\alpha_i| \leq (\operatorname{ind} K)^{-1/d} \Big\}.$$

Then $\operatorname{vol} A = 2^d$ and by Minkowski's Convex Body Theorem there is a non-zero point $v \in A \cap \mathbb{Z}^d$, cf. Figure 19.

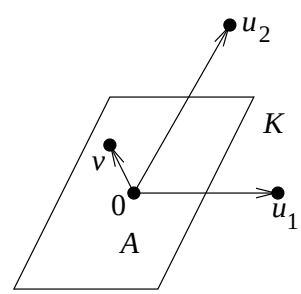


Figure 19. Finding the point.

As we mentioned, we can find this point efficiently if the dimension d is fixed.

For $i = 1, \ldots, d$, let K_i be the cone generated by $u_1, \ldots, u_{i-1}, v, u_{i+1}, \ldots, u_d$. Then one can show that

$$\operatorname{ind} K_i \leq (\operatorname{ind} K)^{\frac{d-1}{d}}.$$

Let $\alpha_i = 1$ or $\alpha_i = -1$ depending on whether the orientations of the bases $u_1, \ldots, u_{i-1}, v, u_{i+1}, \ldots, u_d$ and $u_1, \ldots, u_d$ are the same or the opposite. Then

$$[K] = \sum_{i=1}^d \alpha_i [K_i] \quad + \quad \text{indicators of lower-dimensional cones,}$$

see Figures 20 and 21.

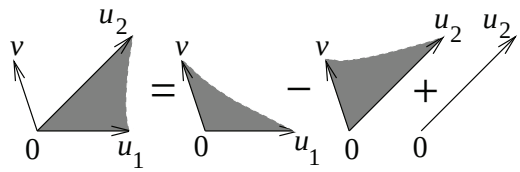


Figure 20. Writing the cone as a linear combination of cones with smaller indices for $d = 2$.

Now we iterate the procedure. After k iterations, we get d^k cones K_i with

$$\operatorname{ind} K_i \leq (\operatorname{ind} K)^{\left(\frac{d-1}{d}\right)^k}.$$

In other words, the number of cones grows exponentially in the number k of iterations while the indices of the cones decrease doubly exponentially in k. It follows that when d is fixed, to obtain a unimodular decomposition, we need $k = O(\log\log(\operatorname{ind} K))$ iterations, which results in a polynomial time algorithm.

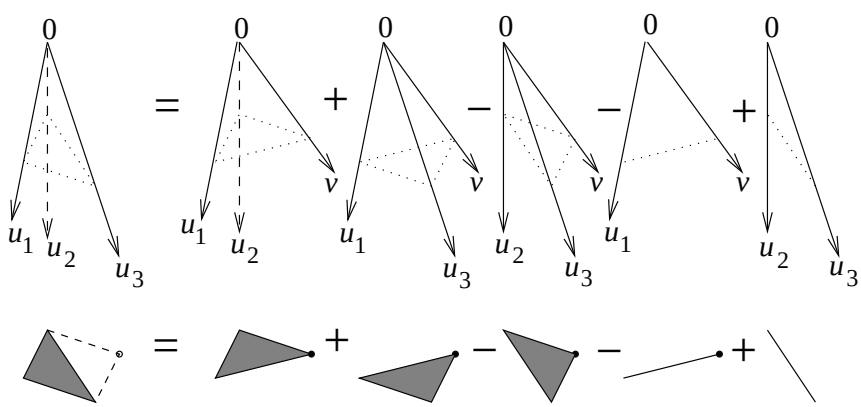


Figure 21. Writing the cone as a linear combination of cones with smaller indices for $d = 3$ (the sections of the cones by a plane are shown below).

There are certain similarities between the described procedure and the unimodular decomposition obtained from the continued fractions method in dimension 2. There are differences, too. In the method just described, there is a certain flexibility in choosing vector v, while the continued fractions method is quite rigid. This is, of course, due to the fact that we know much more about integer points in dimension 2 than in higher dimensions. On the other hand, there is a version of our algorithm that reduces to the continued fractions method in dimension 2.

Polarity and discarding lower-dimensional cones

When we apply the procedure described above, there is an apparent nuisance of dealing with lower-dimensional cones. However, there is a certain trick which allows us simply to forget about them.

Let $K \subset \mathbb{R}^d$ be a simple rational cone. Let

$$K^\circ = \left\{ x \in \mathbb{R}^d : \quad \langle x, y \rangle \leq 0 \quad \text{for all} \quad y \in K \right\}$$

be the polar of K, see Lecture 2. It is not hard to prove that K° is a simple rational cone, that K° is unimodular if and only if K is unimodular, and that $(K^\circ)^\circ = K$. Thus we modify the above procedure as follows.

Given a simple rational cone K, we compute the polar K°. Then we apply the unimodular decomposition and get

$$[K^\circ] = \sum_i \alpha_i [K_i] \quad + \quad \text{indicators of lower-dimensional cones,}$$

where K_i are unimodular cones. Next, we compute K_i° and observe that

$$[K] = \sum_i \alpha_i [K_i^\circ] \quad + \quad \text{indicators of cones with lines,}$$

see Theorem 4 and Review Problem 14 of Lecture 1.

By Theorem 3 of Lecture 4,

$$f(K, \mathbf{x}) = \sum_i \alpha_i f(K_i^\circ, \mathbf{x}),$$

since we can ignore polyhedra with lines.

Problems

Review problems

1. Check that the procedure of computing $f(K, \mathbf{x})$ for a simple rational cone $K \subset \mathbb{R}^2$ via continued fractions (see Lecture 3) indeed runs in polynomial time.
2. Prove Minkowski's Convex Body Theorem.
3. Check that the algorithm for the unimodular decomposition of a cone indeed works. 4. Let $K \subset \mathbb{R}^d$ be a unimodular cone. Prove that K° is a unimodular cone.

Supplementary problems

1. Let a_1 and a_2 be positive coprime integers and let $S \subset \mathbb{Z}$ be the set of all non-negative integer combinations of a_1 and a_2. Prove that
$$\sum_{m \in S} x^m = \frac{1 - x^{a_1 a_2}}{(1 - x^{a_1})(1 - x^{a_2})}.$$
2. Let a_1, a_2 and a_3 be positive coprime integers and let $S \subset \mathbb{Z}$ be the set of all non-negative integer combinations of a_1, a_2 and a_3. Prove that
$$\sum_{m \in S} x^m = \frac{1 - x^{b_1} - x^{b_2} - x^{b_3} + x^{b_4} + x^{b_5}}{(1 - x^{a_1})(1 - x^{a_2})(1 - x^{a_3})},$$
for some, not necessarily distinct, integers $b_i = b_i(a_1, a_2, a_3)$, $i = 1, 2, 3, 4, 5$.
3. Let $a_1, \ldots, a_n \in \mathbb{Z}_+^d$ be some integer vectors with non-negative coordinates. Let $S \subset \mathbb{Z}_+^d$ be the set of all non-negative integer combinations of $a_1, \ldots, a_n$. Prove that the series
$$\sum_{m \in S} \mathbf{x}^m \quad \text{where} \quad \mathbf{x} = (x_1, \ldots, x_d)$$

converges for $|x_i| < 1$, $i = 1, \ldots, d$, to a rational function of $\mathbf{x}$.

Concluding Remarks

The algorithmic theory of counting lattice points in polyhedra is discussed in [**BP99**]; some of the algorithms suggested there are implemented, see [**L+04**] and [**V+04**]. For other algorithmic questions concerning lattice points, see [**G+93**]. For Minkowski's Theorems and other topics in the geometry of numbers, see [**GL87**]. We conclude these lectures by discussing various related topics and open questions.

Something curvilinear?

Is it possible to extend the developed theory onto something non-polyhedral, such as Euclidean balls? Probably not, as it appears to be in the realm of totally different forces, more akin to theta functions than to rational functions. For example, let

$$B = \left\{(x_1, \ldots, x_4) : \quad \sum_{i=1}^{4} x_i^2 \leq n\right\}$$

be the standard Euclidean ball of radius $\sqrt{n}$ in dimension 4. Suppose for a moment that we can efficiently enumerate integer points in B. Then we can count integer points on the sphere $x_1^2 + x_2^2 + x_3^2 + x_4^2 = n$. However, the number of such points, that is, the number of ways to represent n as a sum of four squares of integers, by Jacobi's formula is equal to

$$8 \sum_{4 \not| \, p \, | \, n} p$$

(in words: eight times the sum of the divisors of n that are not divisible by 4). Thus we gain some insight into divisors of n, and, pushing it a bit further, we can come up with an efficient algorithm for factoring integers, see [**B+86**] and [**Dy91**]. The existence of such an algorithm is not entirely impossible, but somewhat doubtful.

Irrational polyhedra?

How can we enumerate integer points in irrational polyhedra? There are some obvious difficulties with generating functions. Consider, for example, a cone $K \subset \mathbb{R}^2$ defined by the inequalities $x_1 \geq 0$ and $x_2 \leq \sqrt{2}x_1$. Just as before, we can write the generating function

$$f(K, \mathbf{x}) = \sum_{m \in K \cap \mathbb{Z}^2} \mathbf{x}^m.$$

The problem is that $f(K, \mathbf{x})$ is no longer a rational function in $\mathbf{x}$. To build an interesting theory, we would like to extend $f(K, \mathbf{x})$ analytically far beyond the region of convergence of the defining series, and it is not clear how to do that.

There is a little trick, however, which allows us to incorporate irrational polyhedra P to some extent. Let us first change the coordinates and consider the *exponential sum*:

$$F(P; c) = \sum_{m \in P \cap \mathbb{Z}^d} e^{\langle c, m \rangle},$$

where $c = (c_1, \ldots, c_d) \in \mathbb{C}^d$. We obtain $F(P; c)$ from $f(P, \mathbf{x})$ by substituting $\mathbf{x} = (e^{c_1}, \ldots, e^{c_d})$. Let $\rho : \mathbb{C}^d \longrightarrow \mathbb{C}$ be a polynomial and let us consider the weighted version

$$F(P, \rho; c) = \sum_{m \in P \cap \mathbb{Z}^d} \rho(m) e^{\langle c, m \rangle}$$

of the exponential sum. One can think of $F(P, \rho; c)$ as the result of applying the differential operator

$$D = \rho\left(\frac{\partial}{\partial c_1}, \ldots, \frac{\partial}{\partial c_d}\right)$$

to $F(P; c)$. If P is an irrational polyhedron, all "bad things" happen along the boundary ∂P of P, so let us cut them out by choosing ρ such that $\rho(x) = 0$ for all $x \in \partial P$ (such a ρ can be obtained by multiplying the equations that define the facets of P). One can show that in this case $F(P, \rho; c)$ indeed extends to a meromorphic function on $\mathbb{C}^d$ and there is a way to extend our theory, see [**Ba93**] for details. This extension, however, is not particularly interesting (it lacks interesting examples so far).

Let's add projections!

There are interesting sets $S \subset \mathbb{Z}^d$ of integer points, which are intimately related to sets of integer points in rational polyhedra but have a more complicated logical structure. Such sets can be quite complicated even in dimension $d = 1$. For example, let us fix positive coprime integers $a_1, \ldots, a_d$ and let $S \subset \mathbb{Z}$ be the set of all integers that are non-negative integer combinations of $a_1, \ldots, a_d$. In other words, S is the semigroup generated by $a_1, \ldots, a_d$. We can think of S as a *projection* of the set of integer points in a rational polyhedron. Let $P = \mathbb{R}^d_+$ be the non-negative orthant and let $T : \mathbb{R}^d \longrightarrow \mathbb{R}$ be the projection

$$(x_1, \ldots, x_d) \longmapsto a_1 x_1 + \ldots + a_d x_d.$$

Then $S = T(P \cap \mathbb{Z}^d)$, the image of the set of integer points in P under the linear transformation T. In [**BW03**], it is proved that for such sets S (obtained from the set of integer points $P \cap \mathbb{Z}^d$ in a rational polyhedron $P \subset \mathbb{R}^d$ by a projection) the generating function

$$f(S; \mathbf{x}) = \sum_{m \in S} \mathbf{x}^m$$

admits a short representation as a rational function in $\mathbf{x}$, which can be computed in polynomial time when the dimension d is fixed.

There have been some advances towards the general theory of sets of integer points encoded by short rational generating functions. For example, in [**BW03**] it is proved that if two sets $S_1, S_2 \subset \mathbb{Z}^d$ are *defined* by their short rational generating functions $f(S_1, \mathbf{x})$ and $f(S_2, \mathbf{x})$, then the generating function $f(S, \mathbf{x})$ of $S = S_1 \cap S_2$ can be computed in polynomial time as a short rational function.

However, we are still quite far from having a full-fledged theory for sets S with short rational generating functions. Suppose, for example, that S is the projection of the difference $X \setminus Y$, where X and Y are the projections of the sets of integer points $P \cap \mathbb{Z}^{k_1}$ and $Q \cap \mathbb{Z}^{k_2}$ in some rational polyhedra P and Q. We don't know how to handle such a set S (our lack of understanding is mitigated by the lack of interesting examples of such complicated constructions). Also, algorithms of [**BW03**] seem to be outrageously impractical.

Polytopes of large dimension?

The theory we described in these lectures provides efficient algorithms if the dimension d of the given rational polytope is fixed in advance. There are certain classes of polytopes for which the algorithms still remain efficient (polynomial time), even when the dimension is allowed to grow, see [**Ba93**] and [**BP99**]. If the dimension d is allowed to grow, the **P** vs. **NP** issue leaves us with little hope to find efficient algorithms for testing whether a given rational polyhedron contains an integer point, let alone for computing the number of such points. The problem, however, remains practically important and it seems that various probabilistic approaches of approximate counting look the most promising here, cf. [**Dy03**].

There seems to be a possibility of a "hybrid" algebraic/probabilistic approach based on the following simple observation. Let $K = K(u_1, \ldots, u_d)$ be a simple rational cone generated by integer vectors $u_1, \ldots, u_d$ and let Π be its fundamental parallelepiped. Theorem 1 of Lecture 3 allows us to compute $f(K, \mathbf{x})$ in terms of the sum $\sum_{m \in \Pi \cap \mathbb{Z}^d} \mathbf{x}^m$. This sum is potentially very big, but it is very easy to *sample* a random integer point $m \in \Pi \cap \mathbb{Z}^d$, cf. Preview Problem 2 in Lecture 3. Indeed, let $\Lambda \subset \mathbb{Z}^d$ be the lattice generated by $u_1, \ldots, u_d$. Then the points $\Pi \cap \mathbb{Z}^d$ are in one-to-one correspondence with the elements of $\mathbb{Z}^d/\Lambda$: if $n \in \mathbb{Z}^d$ is an integer point, then the point $m \in \Pi \cap \mathbb{Z}^d$ such that $m - n \in \Lambda$ is computed as follows: we write $n = \sum_{i=1}^d \alpha_i u_i$ and let $m = \sum_{i=1}^d \{\alpha_i\} u_i$. Hence the problem of sampling $m \in \Pi \cap \mathbb{Z}^d$ reduces to that of sampling coset representatives $n \in \mathbb{Z}^d/\Lambda$, which can be done efficiently. One can ask what happens if we try to count integer points in a given polytope P by using Brion's Theorem (Theorem 4 of Lecture 4) and computing the generating functions of the supporting cones of P approximately via random sampling.

BIBLIOGRAPHY

[Ba93] A. Barvinok, *Computing the volume, counting integral points, and exponential sums*, Discrete Comput. Geom., **10**, (1993), pp. 123-141.

[Ba02] A. Barvinok, *A Course in Convexity*, Graduate Studies in Mathematics, vol 54, Amer. Math. Soc., Providence, RI, 2002.

[Br88] M. Brion, *Points entiers dans les polyédres convexes (French)* , Ann. Sci. École Norm. Sup. (4) **21** (1988), pp. 653–663.

[BP99] A. Barvinok and J. Pommersheim, *An algorithmic theory of lattice points in polyhedra*, New Perspectives in Algebraic Combinatorics (Berkeley, CA, 1996–97), Math. Sci. Res. Inst. Publ., vol 38, Cambridge Univ. Press, Cambridge, 1999, pp. 91–147.

[BW03] A. Barvinok and K. Woods, *Short rational generating functions for lattice point problems*, J. Amer. Math. Soc. **16** (2003), pp. 957–979

[B+86] E. Bach, G. Miller, and J. Shallit, *Sums of divisors, perfect numbers and factoring*, SIAM J. Comput. **15** (1986), pp. 1143–1154.

[B+05] M. Beck and F. Sottile *Irrational proofs of three theorems of Stanley* (preprint) `arXiv math.CO/0501359`, European Journal of Combinatorics, to appear.

[Dy91] M. Dyer, *On counting lattice points in polyhedra*, SIAM J. Comput. **20** (1991), pp. 695–707.

[Dy03] M. Dyer, *Approximate counting by dynamic programming*, Proceedings of the 35th Annual ACM Symposium on the Theory of Computing (STOC 2003), 2003, pp. 693–699.

[GL87] P.M. Gruber and C.G. Lekkerkerker, *Geometry of Numbers. Second edition,* North-Holland Mathematical Library, vol. 37, North-Holland, Amsterdam, 1987.

[G+93] M. Grötschel, L. Lovász, and A. Schrijver, *Geometric Algorithms and Combinatorial Optimization. Second edition,* Algorithms and Combinatorics, vol. 2, Springer-Verlag, Berlin, 1993.

[Kh97] A. Ya. Khinchin, *Continued Fractions*, Translated from the third (1961) Russian edition. Reprint of the 1964 translation, Dover Publications, Inc., Mineola, NY, 1997.

[KR97] D. Klain and G.-C. Rota, *Introduction to Geometric Probability*, Lezioni Lincee, Cambridge Univ. Press, Cambridge, 1997.

[KP92] A.G. Khovanskii and A.V. Pukhlikov, *The Riemann-Roch theorem for integrals and sums of quasipolynomials on virtual polytopes. (Russian),* translation in St. Petersburg Math. J. 4 (1993), no. 4, pp. 789–812, Algebra i Analiz **4**, no. 4 (1992), pp. 188–216.

[KP93] A.G. Khovanskii and A.V. Pukhlikov, *Integral transforms based on Euler characteristic and their applications*, Integral Transform. Spec. Funct. **1** (1993), pp. 19–26.

[La91] J. Lawrence, *Rational-function-valued valuations on polyhedra,* Discrete and Comput. Geometry (New Brunswick, NJ, 1989/1990), DIMACS Ser. Discrete Math. Theoret. Comput. Sci., vol. 6, Amer. Math. Soc., Providence, RI, 1991, pp. 199–208.

[L+04] J.A. De Loera, R. Hemmecke, J. Tauzer, and R. Yoshida, *Effective lattice point counting in rational convex polytopes*, Journal of Symbolic Computation **38** (2004), pp. 1273–1302.
see also `http://www.math.ucdavis.edu/~latte/`.

[Pa94] C.H. Papadimitriou, *Computational Complexity*, Addison-Wesley, addr Reading, MA, 1994.

[St97] R.P. Stanley, *Enumerative Combinatorics. Vol. 1,* Corrected reprint of the 1986 original. Cambridge Studies in Advanced Mathematics, vol. 49, Cambridge Univ. Press, Cambridge, 1997.

[V+04] S. Verdoolaege, R. Seghir, K. Beyls, V. Loechner, and M. Bruynooghe, *Analytical computation of Ehrhart polynomials: enabling more compiler analyses and optimizations*, Proceedings of the 2004 International Conference on Compilers, Architecture, and Synthesis for Embedded Systems (CASES 2004), 2004, pp. 248–258.
see also `http://www.kotnet.org/~skimo/barvinok/`.

[Zi95] G. Ziegler, *Lectures on Polytopes*, Graduate Texts in Mathematics, vol. 152, Springer-Verlag, New York, 1995.

Root Systems and Generalized Associahedra

Sergey Fomin and Nathan Reading

IAS/Park City Mathematics Series
Volume 14, 2004

Root Systems and Generalized Associahedra

Sergey Fomin and Nathan Reading

These lecture notes provide an overview of root systems, generalized associahedra, and the combinatorics of clusters. Lectures 1-2 cover classical material: root systems, finite reflection groups, and the Cartan-Killing classification. Lectures 3–4 provide an introduction to cluster algebras from a combinatorial perspective. Lecture 5 is devoted to related topics in enumerative combinatorics.

There are essentially no proofs but an abundance of examples. We label unproven assertions as either "lemma" or "theorem" depending on whether they are easy or difficult to prove. We encourage the reader to try proving the lemmas, or at least get an idea of why they are true.

For additional information on root systems, reflection groups and Coxeter groups, the reader is referred to [**9, 25, 34**]. For basic definitions related to convex polytopes and lattice theory, see [**58**] and [**31**], respectively. Primary sources on generalized associahedra and cluster combinatorics are [**13, 19, 21**]. Introductory surveys on cluster algebras were given in [**22, 56, 57**].

Note added in press (February 2007): Since these lecture notes were written, there has been much progress in the general area of cluster algebras and Catalan combinatorics of Coxeter groups and root systems. We have not attempted to update the text to reflect these most recent advances. Instead, we refer the reader to the online Cluster Algebras Portal, maintained by the first author.

[1]Department of Mathematics, University of Michigan, Ann Arbor, MI 48109-1109, USA.
E-mail address: `fomin@umich.edu, nreading@umich.edu`.
This work was partially supported by NSF grants DMS-0245385 (S.F.) and DMS-0202430 (N.R.).

©2007 Sergey Fomin and Nathan Reading

Acknowledgments

We thank Christos Athanasiadis, Jim Stasheff and Andrei Zelevinsky for careful readings of earlier versions of these notes and for a number of editorial suggestions, which led to the improvement of the paper.

S.F.: I am grateful to the organizers of the 2004 Graduate Summer School at Park City (Ezra Miller, Vic Reiner, and Bernd Sturmfels) for the invitation to deliver these lectures, and for their support, understanding, and technical help.

Sections 3.3-3.4 and Lecture 4 present results of an ongoing joint project with Andrei Zelevinsky centered around cluster algebras.

N.R.: I would like to thank Vic Reiner for teaching the course which sparked my interest in Coxeter groups; Anders Björner and Francesco Brenti for making a preliminary version of their forthcoming book available to the students in Reiner's course; and John Stembridge, whose course and lecture notes have deepened my knowledge of Coxeter groups and root systems.

Some of the figures in these notes are inspired by figures produced by Satyan Devadoss, Vic Reiner and Rodica Simion. Several figures were borrowed from [**13, 19, 20, 21, 23**].

LECTURE 1
Reflections and Roots

1.1. The Pentagon Recurrence

Consider a sequence $f_1, f_2, f_3, \ldots$ defined recursively by $f_1 = x$, $f_2 = y$, and

$$f_{n+1} = \frac{f_n + 1}{f_{n-1}}. \tag{1}$$

Thus, the first five entries are

$$x,\ y,\ \frac{y+1}{x},\ \frac{x+y+1}{xy},\ \frac{x+1}{y}. \tag{2}$$

Unexpectedly, the sixth and seventh entries are x and y, respectively, so the sequence is periodic with period five! We will call (1) the *pentagon recurrence.*[1]

This sequence has another important property. *A priori*, we can only expect its terms to be rational functions of x and y. In fact, each f_i is a *Laurent polynomial* (actually, with nonnegative integer coefficients). This is an instance of what is called the *Laurent phenomenon.*

It will be helpful to represent this recurrence as the evolution of a "moving window" consisting of two consecutive terms f_i and f_{i+1}:

$$\begin{bmatrix} f_1 \\ f_2 \end{bmatrix} \xrightarrow{\tau_1} \begin{bmatrix} f_3 \\ f_2 \end{bmatrix} \xrightarrow{\tau_2} \begin{bmatrix} f_3 \\ f_4 \end{bmatrix} \xrightarrow{\tau_1} \begin{bmatrix} f_5 \\ f_4 \end{bmatrix} \xrightarrow{\tau_2} \begin{bmatrix} f_5 \\ f_6 \end{bmatrix} \longrightarrow \cdots,$$

where the maps τ_1 and τ_2 are defined by

$$\tau_1 : \begin{bmatrix} f \\ g \end{bmatrix} \longmapsto \begin{bmatrix} \frac{g+1}{f} \\ g \end{bmatrix} \quad \text{and} \quad \tau_2 : \begin{bmatrix} f \\ g \end{bmatrix} \longmapsto \begin{bmatrix} f \\ \frac{f+1}{g} \end{bmatrix}. \tag{3}$$

Both τ_1 and τ_2 are involutions: $\tau_1^2 = \tau_2^2 = 1$, where 1 denotes the identity map. The 5-periodicity of the recurrence (1) translates into the identity $(\tau_2\tau_1)^5 = 1$. That is, the group generated by τ_1 and τ_2 is a dihedral group with 10 elements.

Let us now consider a similar but simpler pair of maps. Throw away the $+1$'s that occur in the definitions of τ_1 and τ_2, and take logarithms. We then obtain a pair of linear maps

$$s_1 : \begin{bmatrix} x \\ y \end{bmatrix} \longmapsto \begin{bmatrix} y - x \\ y \end{bmatrix} \quad \text{and} \quad s_2 : \begin{bmatrix} x \\ y \end{bmatrix} \longmapsto \begin{bmatrix} x \\ x - y \end{bmatrix}.$$

A (linear) *hyperplane* in a vector space V is a linear subspace of codimension 1. A (linear) *reflection* is a map that fixes all the points in some linear hyperplane, and has an eigenvalue of -1. The maps s_1 and s_2 are linear reflections satisfying $(s_2 s_1)^3 = 1$. Thus, the group $\langle s_1, s_2 \rangle$ is a dihedral group with 6 elements.

We are led to wonder if the dihedral behavior of $\langle \tau_1, \tau_2 \rangle$ is related to, or even explained by the dihedral behavior of $\langle s_1, s_2 \rangle$. To test this unlikely-sounding hypothesis, let us try to find similar examples. What other pairs (s, s') of linear

[1]The discovery of this recurrence and its 5-periodicity are sometimes attributed to R. C. Lyness (1942); see, e.g., [**15**]. It was probably already known to N. H. Abel. This recurrence is closely related to (and easily deduced from) the famous "pentagonal identity" for the dilogarithm function, first obtained by W. Spence (1809) and rediscovered by Abel (1830) and C. H. Hill (1830). See, e.g., [**37**].

reflections generate finite dihedral groups? To keep things simple, we set $s = s_1$ and confine the choice of s' to maps of the form

$$s' : \begin{bmatrix} x \\ y \end{bmatrix} \longmapsto \begin{bmatrix} x \\ L(x,y) \end{bmatrix},$$

where L is a linear function. Keeping in mind that s_1 and s_2 arose as logarithms, we require that L have integer coefficients.

After some work, one determines that besides $x - y$, the functions $2x - y$ and $3x - y$ are the only good choices for L. More specifically, define

$$s_3 : \begin{bmatrix} x \\ y \end{bmatrix} \longmapsto \begin{bmatrix} x \\ 2x - y \end{bmatrix} \quad \text{and} \quad s_4 : \begin{bmatrix} x \\ y \end{bmatrix} \longmapsto \begin{bmatrix} x \\ 3x - y \end{bmatrix}.$$

Then $(s_3 s_1)^4 = 1$ and $(s_4 s_1)^6 = 1$. Thus, $\langle s_1, s_3 \rangle$ and $\langle s_1, s_4 \rangle$ are dihedral groups with 8 and 12 elements, respectively.

By analogy with (3), we next define

$$\tau_3 : \begin{bmatrix} f \\ g \end{bmatrix} \longmapsto \begin{bmatrix} f \\ \frac{f^2+1}{g} \end{bmatrix} \quad \text{and} \quad \tau_4 : \begin{bmatrix} f \\ g \end{bmatrix} \longmapsto \begin{bmatrix} f \\ \frac{f^3+1}{g} \end{bmatrix}.$$

Calculations show that $(\tau_3 \tau_1)^6 = 1$, and the group $\langle \tau_1, \tau_3 \rangle$ is dihedral with 12 elements. We can think of τ_1 and τ_3 as defining a "moving window" for the sequence

$$x,\ y,\ \frac{y+1}{x},\ \frac{x^2+(y+1)^2}{x^2 y},\ \frac{x^2+y+1}{xy},\ \frac{x^2+1}{y},\ x,\ y,\ \ldots \tag{4}$$

Notice that the Laurent phenomenon holds: these rational functions are Laurent polynomials—again, with nonnegative integer coefficients.

Likewise, $(\tau_4 \tau_1)^8 = 1$, the group $\langle \tau_1, \tau_4 \rangle$ is dihedral with 16 elements, and τ_1 and τ_4 define an 8-periodic sequence of Laurent polynomials.

In the first two lectures, we will develop the basic theory of finite reflection groups that will include their complete classification. This theory will later help explain the periodicity and Laurentness of the sequences discussed above, and provide appropriate algebraic and combinatorial tools for the study of other similar recurrences.

1.2. Reflection Groups

Our first goal will be to understand the finite groups generated by linear reflections in a vector space V. It turns out that for such a group, it is always possible to define a Euclidean structure on V so that all of the reflections in the group are ordinary *orthogonal reflections*. The study of groups generated by orthogonal reflections is a classical subject, which goes back to the classification of Platonic solids by the ancient Greeks.

Let V be a Euclidean space. In what follows, all reflecting hyperplanes pass through the origin, and all reflections are orthogonal. A finite *reflection group* is a finite group generated by some reflections in V. In other words, we choose a collection of hyperplanes such that the group of orthogonal transformations generated by the corresponding reflections is finite. Infinite reflection groups are also interesting, but in these lectures, "reflection group" will always mean a finite one.

The set of reflections in a reflection group W is typically larger than a minimal set of reflections generating W. This is illustrated in Figure 1.1, where W is the group of symmetries of a regular pentagon. This 10-element group is generated by

two reflections s and t whose reflecting lines make an angle of $\pi/5$. It consists of 5 reflections, 4 rotations, and the identity element. In Figure 1.1, each of the 5 lines is labeled by the corresponding reflection.

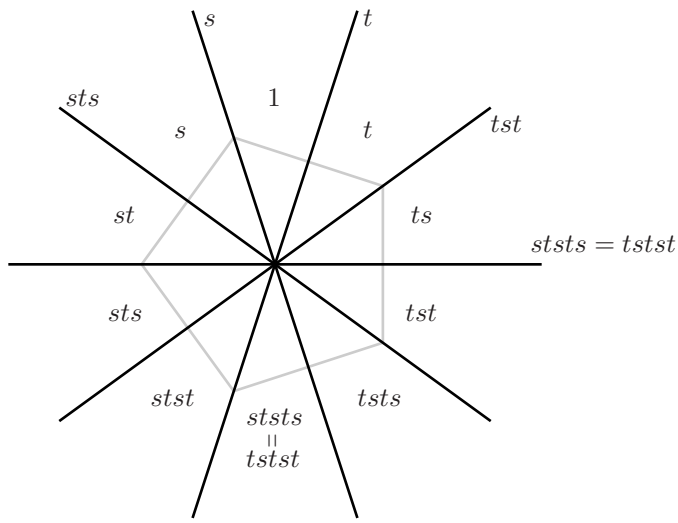


Figure 1.1. The reflection group $I_2(5)$.

Lemma 1.1. *If t is the reflection fixing a hyperplane H and w an orthogonal transformation, then wtw^{-1} is the reflection fixing the hyperplane wH.*

Lemma 1.2. *Let W be a finite group generated by a finite set T of reflections. Then the set of all reflections in W is $\{wtw^{-1} : w \in W, t \in T\}$.*

The set $\mathcal{H}$ of all reflecting hyperplanes of a reflection group W is called a *Coxeter arrangement*. In light of Lemmas 1.1 and 1.2, one can give an alternate definition of a Coxeter arrangement: A Coxeter arrangement is a collection $\mathcal{H}$ of hyperplanes which is closed under reflections in the hyperplanes. Like any hyperplane arrangement in V, a Coxeter arrangement cuts V into connected components called *regions*. That is, the regions are the connected components of the complement to the union of all hyperplanes in $\mathcal{H}$.

The regions are in one-to-one correspondence with the elements of W, as follows. Once and for all, fix an arbitrary region R_1 to represent the identity element.

Lemma 1.3. *The map $w \mapsto R_w \stackrel{\text{def}}{=} w(R_1)$ is a bijection between a reflection group W and the set of regions of the corresponding Coxeter arrangement.*

To illustrate, each of the 10 regions in Figure 1.1 is labeled by the corresponding element of the group.

The choice of a region representing the identity element leads to a distinguished choice of a minimal set of generating reflections. The *facet hyperplanes* of R_1 are the hyperplanes in $\mathcal{H}$ whose intersection with the closure of R_1 has dimension $n-1$.

Lemma 1.4. *The reflections in the facet hyperplanes of R_1 generate W. This generating set is minimal by inclusion.*

1.3. Symmetries of Regular Polytopes

A *regular polytope* in a Euclidean space is a convex polytope whose symmetry group (i.e., the group of isometries of the space that leave the polytope invariant) acts transitively on *complete flags of faces*, i.e., on nested collections of the form

$$\text{vertex} \subset \text{edge} \subset \text{2-dim. face} \subset \cdots$$

Theorem 1.5. *The symmetry group of any regular polytope is a reflection group.*

The converse is false—see Remark 1.12.

We illustrate Theorem 1.5 with several concrete examples.

Example 1.6. Consider a regular m-gon on a Euclidean plane, centered at the origin. The symmetry group of the m-gon is denoted by $I_2(m)$. This group contains (and is generated by) m reflections, which correspond to the m lines of reflective symmetry of the m-gon.

The group $I_2(m)$ is a dihedral group with $2m$ elements. It is generated by two reflections s and t satisfying $(st)^m = 1$. To define s and t, we use the construction of Lemma 1.4. Pick a side of the polygon, and consider two reflecting lines: one perpendicular to the side and another passing through one of its endpoints. The case $m = 5$ is shown in Figure 1.1.

Example 1.7. Take a regular tetrahedron in 3-space, with the vertices labeled 1, 2, 3, and 4. Its symmetry group is obviously isomorphic to the *symmetric group* $\mathcal{S}_4$, which consists of the permutations of the set $\{1, 2, 3, 4\}$. For each edge of the tetrahedron, choose a plane which is perpendicular to the edge and contains the other two vertices. Reflections in these six hyperplanes generate the symmetry group.

In general, the symmetry group of a regular simplex can be described as follows. Let $(e_1, \ldots, e_{n+1})$ be the standard basis in $\mathbb{R}^{n+1}$. The *standard n-dimensional simplex* (or n-simplex) is the convex hull of the endpoints of the vectors $e_1, \ldots, e_{n+1}$. Thus the standard 1-simplex is a line segment in $\mathbb{R}^2$, the standard 2-simplex is an equilateral triangle in $\mathbb{R}^3$, and the standard 3-simplex is the regular tetrahedron described above, sitting in $\mathbb{R}^4$. The symmetry group A_n of the standard n-simplex is canonically isomorphic to $\mathcal{S}_{n+1}$, the symmetric group of permutations of the set $[n+1] \stackrel{\text{def}}{=} \{1, 2, \ldots, n+1\}$.

For each edge $[e_i, e_j]$ of the standard simplex, there is a hyperplane $x_i - x_j = 0$ perpendicular to the edge and containing all the other vertices. Reflection through this hyperplane interchanges the endpoints of the edge and fixes the rest of the vertices. These $\binom{n+1}{2}$ reflections generate A_n.

To construct a minimal generating set of reflections, we again use Lemma 1.4. Let R_1 be the connected component of the complement to the $\binom{n+1}{2}$ reflecting hyperplanes defined by

$$R_1 = \{x_1 < x_2 < \cdots < x_{n+1}\}. \tag{5}$$

The facet hyperplanes of R_1 are given by the equations

$$x_i - x_{i+1} = 0, \quad \text{for } i = 1, \ldots, n.$$

Then Lemma 1.4 reduces to the well-known fact that the symmetric group $\mathcal{S}_{n+1}$ is generated by the adjacent transpositions $s_1, \ldots, s_n$. (Here each s_i exchanges i and $i+1$, keeping everything else in its place.)

Figure 1.2 illustrates the special case $n = 2$, the symmetry group of the standard 2-simplex (shaded). The plane of the page represents the plane $x + y + z = 1$ in $\mathbb{R}^3$.

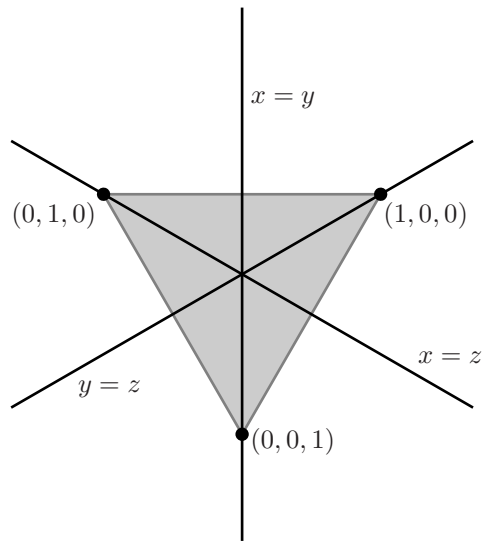


Figure 1.2. The reflection group A_2.

Example 1.8. The *n-crosspolytope* is the convex hull of (the endpoints of) the vectors $\pm e_1, \pm e_2, \ldots, \pm e_n$ in $\mathbb{R}^n$. For example, the 3-crosspolytope is the regular octahedron. The symmetry group of this polytope is the *hyperoctahedral group* B_n. As in the previous examples, it is generated by the reflections it contains.

The special case $n = 3$ (the symmetry group B_3 of a regular octahedron) is shown in Figure 1.3. The dotted lines show the intersections of reflecting hyperplanes with the front surface of the octahedron. Each edge of the octahedron is also contained in a reflecting plane.

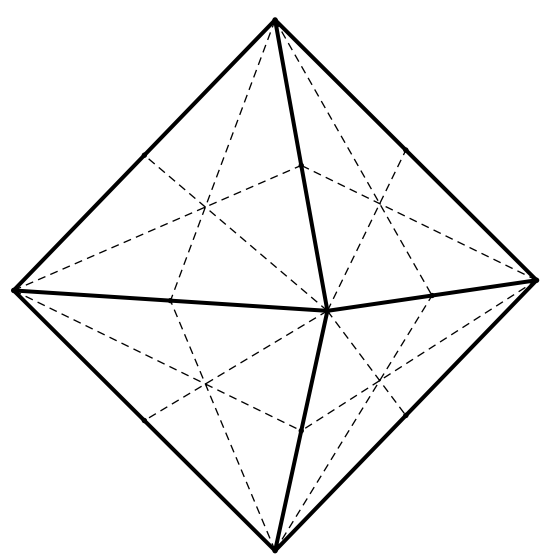


Figure 1.3. The reflection group B_3

There are two types of reflections in the symmetry group of the crosspolytope. One type of reflection transposes a vertex with its negative and fixes all other vertices. Also, for each pair $i \neq j$, there is a reflection which transposes e_i and e_j, transposes $-e_i$ and $-e_j$, and fixes all other vertices.

To construct a minimal set of reflections generating B_n, take the minimal generating set for A_{n-1} given in Example 1.7 and adjoin the reflection that interchanges e_1 and $-e_1$.

The group B_n is also the symmetry group of the n-dimensional cube.

Example 1.9. The symmetry group of a regular dodecahedron (or a regular icosahedron) is the reflection group H_3. Figure 1.4 shows the dodecahedron and a minimal set of three reflections generating its symmetry group. The dotted lines show the intersections of the corresponding three hyperplanes with the front surface of the dodecahedron.

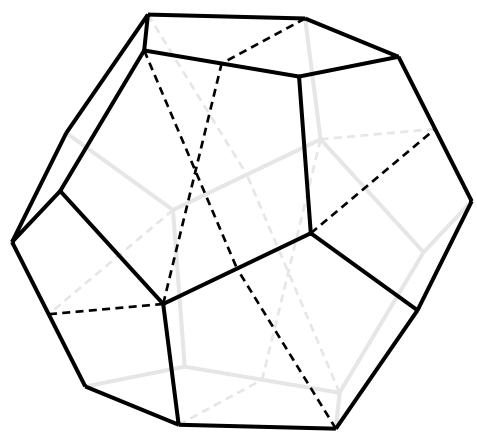


Figure 1.4. The reflection group H_3

Example 1.10. In 4-space, there are six types of regular polytopes. The obvious three are the 4-simplex, the 4-cube, and the 4-crosspolytope. There are two regular polytopes whose symmetry group is the reflection group called H_4. One of these, the 120-*cell*, has 600 vertices and 120 dodecahedral faces; the other, the 600-*cell*, has 120 vertices and 600 tetrahedral faces. The remaining regular 4-dimensional polytope is the 24-*cell*, with 24 vertices and 24 octahedral faces. Its symmetry group is a reflection group denoted by F_4.

Not every reflection group is the symmetry group of a regular polytope. A counterexample is constructed as follows.

Example 1.11. Let $n \geq 3$. Returning to the crosspolytope, ignore the reflections which transpose an opposite pair of vertices. The remaining reflections generate a reflection group called D_n, which is a proper subgroup of B_n. The reflections of D_3 are represented by the dotted lines in Figure 1.3. We note that the Coxeter arrangements of types A_3 and D_3 are related by an orthogonal transformation, so the reflection groups A_3 and D_3 are isomorphic to each other.

Remark 1.12. It can be shown that, for $n \geq 4$, the group D_n is not a symmetry group of a regular polytope. See Section 2.3 for further details.

1.4. Root Systems

Root systems are configurations of vectors obtained by replacing each reflecting hyperplane of a reflection group by a pair of opposite normal vectors; the resulting configuration should be invariant under the action of the group. Here is a formal definition. A finite *root system* is a finite non-empty collection Φ of nonzero vectors in V called *roots* with the following properties:

(i) Each one-dimensional subspace of V either contains no roots, or contains two roots $\pm\alpha$.

(ii) For each $\alpha \in \Phi$, the reflection σ_α permutes Φ.

The following lemma shows that the study of root systems is essentially equivalent to the study of reflection groups.

Lemma 1.13. *For a finite root system Φ, the group generated by the reflections $\{\sigma_\alpha : \alpha \in \Phi\}$ is finite. The corresponding reflecting hyperplanes form a Coxeter arrangement. Conversely, for any reflection group W, there is a root system Φ such that the orthogonal reflections $\{\sigma_\alpha\}_{\alpha\in\Phi}$ are precisely the reflections in W.*

In Section 1.2, we fixed a region R_1 of the associated Coxeter arrangement $\mathcal{H}$. The *simple roots* in Φ are the roots normal to the facet hyperplanes of R_1 and pointing into the half-space containing R_1. The *rank* of Φ is the cardinality n of the set of simple roots Π. Since W acts transitively on the regions of $\mathcal{H}$, the rank of Φ does not depend on the choice of Π, and is equal to the dimension of the linear span of Φ. It will be convenient to fix an indexing set I so that $\Pi = \{\alpha_i : i \in I\}$. The standard choice is $I = [n] = \{1, \ldots, n\}$.

For any $\alpha \in \Phi$, the coefficients c_i in the expansion $\alpha = \sum_{i\in I} c_i\alpha_i$ are called the *simple root coordinates* of α. The set Φ_+ of *positive roots* consists of all roots whose simple root coordinates are all non-negative. The *negative roots* Φ_- are those with non-positive simple root coordinates.

Lemma 1.14. *Φ is the disjoint union of Φ_+ and Φ_-.*

In these lectures, we focus on the study of the important class of finite *crystallographic* root systems. These are the finite non-empty collections of vectors that, in addition to the axioms (i)–(ii) above, satisfy the "crystallographic condition"

(iii) For any $\alpha, \beta \in \Phi$, we have $\sigma_\alpha(\beta) = \beta - a_{\alpha\beta}\alpha$ with $a_{\alpha\beta} \in \mathbb{Z}$. (See Figure 1.5.)

Equivalently, the simple root coordinates of any root are integers.

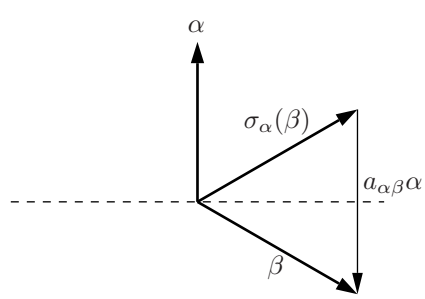


Figure 1.5. Reflecting β in the hyperplane perpendicular to α.

For the rest of these lectures, a "root system" will always be presumed finite and crystallographic.

Example 1.15. A root system of rank 1 is called A_1; it consists of a pair of vectors $\pm\alpha$. There are four non-isomorphic (finite crystallographic) root systems of rank 2, called $A_1 \times A_1$, A_2, B_2 and G_2; see Figure 1.6.

$A_1 \times A_1$

$$\sigma_{\alpha_1} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \sigma_{\alpha_2} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

A_2

$$\sigma_{\alpha_1} = \begin{bmatrix} -1 & 1 \\ 0 & 1 \end{bmatrix} \qquad \sigma_{\alpha_2} = \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix}$$

B_2

$$\sigma_{\alpha_1} = \begin{bmatrix} -1 & 2 \\ 0 & 1 \end{bmatrix} \qquad \sigma_{\alpha_2} = \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix}$$

G_2

$$\sigma_{\alpha_1} = \begin{bmatrix} -1 & 3 \\ 0 & 1 \end{bmatrix} \qquad \sigma_{\alpha_2} = \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix}$$

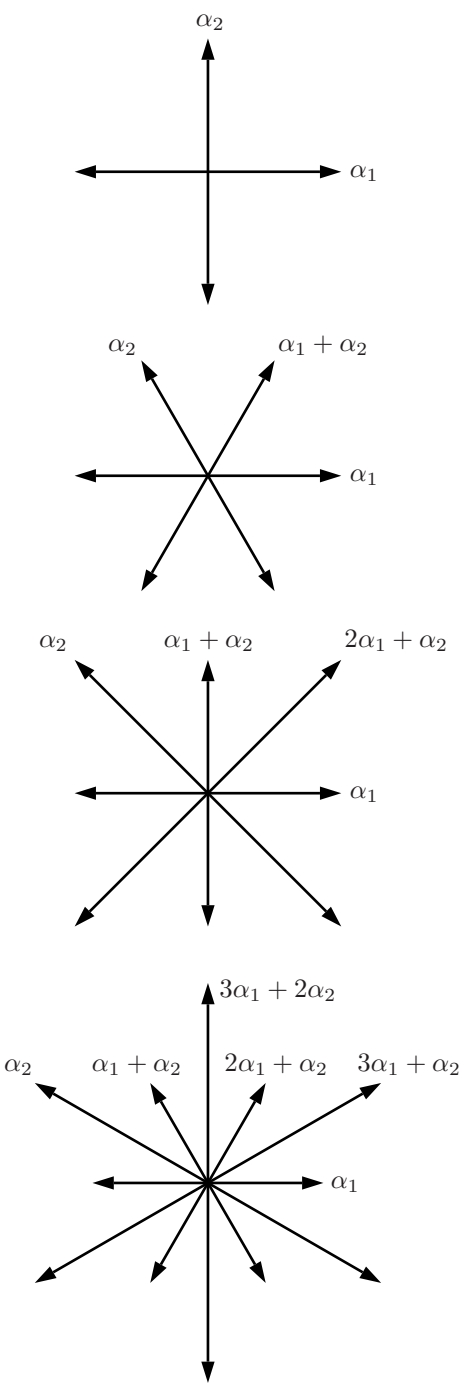


Figure 1.6. The finite crystallographic root systems of rank 2

For the root systems A_2, B_2 and G_2, the reflections σ_{α_1} and σ_{α_2} have appeared earlier in Section 1.1. (The matrices of these reflections in the basis (α_1, α_2) of simple roots are shown in Figure 1.6.) In these three cases, the pair $(\sigma_{\alpha_1}, \sigma_{\alpha_2})$ coincides with (s_2, s_1), (s_3, s_1), and (s_4, s_1), respectively, in the notation of Section 1.1.

1.5. Root Systems of Types A, B, C, and D

Here we present four classical families of root systems, traditionally denoted by A_n, B_n, C_n and D_n. The corresponding reflection groups have types A_n, B_n, B_n and D_n (cf. Examples 1.7, 1.8, and 1.11). In each case, n is the rank of a root system.

We realize each root system inside a Euclidean space with a fixed orthonormal basis $(e_1, e_2, \dots)$, and describe particular choices of the sets of simple and positive roots. There is no "canonical" way to make these choices. Our realizations of root systems coincide with those in [**9, 34**], but our choices of simple/positive roots (which are motivated by notational convenience alone) are different.

The root system A_n

The root system A_n can be realized as the set of vectors $e_i - e_j$ in $\mathbb{R}^{n+1}$ with $i \neq j$. Let R_1 be given by (5). Then the n simple roots are $\alpha_i \stackrel{\text{def}}{=} e_{i+1} - e_i$, for $i = 1, \dots, n$, and the positive roots are $e_i - e_j$, for $1 \leq j < i \leq n+1$.

Figure 1.7 shows a planar projection of the root system A_3. The positive roots are labeled by their simple root coordinates. The solid lines are in the plane of the page. Thick dotted lines are above the plane, while thin dotted lines are below it.

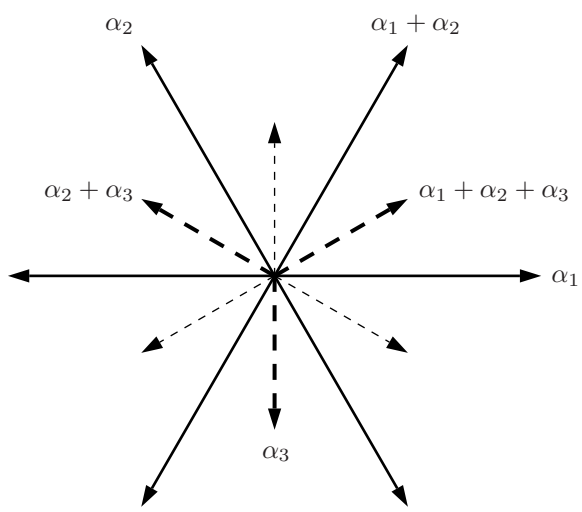


Figure 1.7. The root system A_3

The root systems B_n and C_n

The root system B_n can be realized as the set of vectors in $\mathbb{R}^n$ of the form $\pm e_i$ or $\pm e_i \pm e_j$ with $i \neq j$. Choose $R_1 = \{0 < x_1 < x_2 < \cdots < x_n\}$. Then the vectors $\alpha_0 = e_1$ and $\alpha_i = e_{i+1} - e_i$ for $i \in [n-1]$ form a set of simple roots. The positive roots are e_i for $i \in [n]$ and $e_i \pm e_j$ for $1 \leq j < i \leq n$. See Figure 1.8.

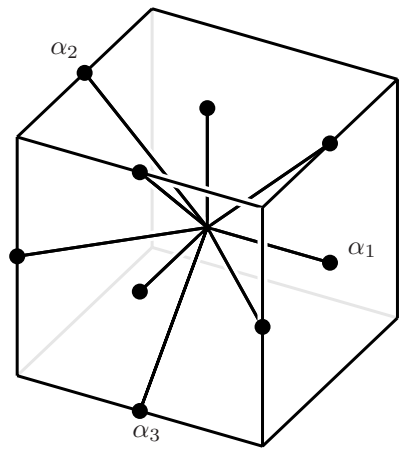


Figure 1.8. The root system B_3. The endpoints of the 9 positive roots are shown as black circles on the cube's front. The negative roots are not shown.

The root system C_n can be realized as the set of vectors in $\mathbb{R}^n$ of the form $\pm 2e_i$ or $\pm e_i \pm e_j$. The vectors $\alpha_0 = 2e_1$ and $\alpha_i = e_{i+1} - e_i$ form a set of simple roots. The positive roots are $2e_i$ and $e_i \pm e_j$. See Figure 1.9.

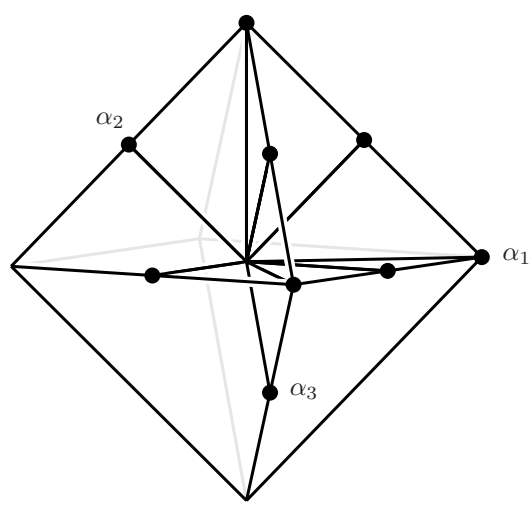


Figure 1.9. The root system C_3. The endpoints of the 9 positive roots are shown on the front of the octahedron. The negative roots are not shown.

The root system C_n is a rescaling of B_n, so the corresponding reflection groups W coincide. In contrast to the type A_n, the action of W on the roots of B_n or C_n is not transitive: there are two orbits, corresponding to two different lengths of roots.

The root system D_n

The root system D_n can be realized as the vectors $\pm e_i \pm e_j$ with $i \neq j$. One choice of simple roots is $\alpha_0 = e_2 + e_1$ and $\alpha_i = e_{i+1} - e_i$, giving the positive roots $e_i \pm e_j$ for $1 \leq j < i \leq n$. This comes from setting $R_1 = \{-x_2 < x_1 < x_2 < \cdots < x_n\}$.

LECTURE 2
Dynkin Diagrams and Coxeter Groups

2.1. Finite Type Classification

The most fundamental result in the theory of (finite crystallographic) root systems is their complete classification, obtained by W. Killing and E. Cartan in late nineteenth – early twentieth century. (See the historical notes in [**9**].) To present this classification, we will need a few preliminaries.

First, we will need the notion of isomorphism. The *ambient space* $Q_{\mathbb{R}} = Q_{\mathbb{R}}(\Phi)$ of a root system Φ is the real span of Φ. It inherits a Euclidean structure from V. Root systems Φ and Φ' are *isomorphic* if there is an isometry map $Q_{\mathbb{R}}(\Phi) \to Q_{\mathbb{R}}(\Phi')$ of their ambient spaces that sends Φ to some dilation $c\Phi'$ of Φ'.

The *Cartan matrix* of a root system Φ is the integer matrix $[a_{ij}]_{i,j\in I}$, where a_{ij} is such that $\sigma_{\alpha_i}(\alpha_j) = \alpha_j - a_{ij}\alpha_i$, as in part (iii) of the definition of a root system. (This convention agrees with [**21, 35**] but is "transposed" to the one in [**9, 34**].)

Lemma 2.1. *Root systems Φ and Φ' are isomorphic if and only if they have the same Cartan matrix, up to simultaneous rearrangement of rows and columns.*

Example 2.2. The Cartan matrices for the root systems of rank two are:

$$A_1 \times A_1 : \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \qquad A_2 : \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

$$B_2 : \begin{bmatrix} 2 & -2 \\ -1 & 2 \end{bmatrix} \qquad G_2 : \begin{bmatrix} 2 & -3 \\ -1 & 2 \end{bmatrix}$$

Example 2.3. *The Cartan matrices for the root systems of type A_4, B_4, C_4, and D_4 are, respectively:*

$$A_4 : \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix} \qquad B_4 : \begin{bmatrix} 2 & -2 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix}$$

$$C_4 : \begin{bmatrix} 2 & -1 & 0 & 0 \\ -2 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix} \qquad D_4 : \begin{bmatrix} 2 & 0 & -1 & 0 \\ 0 & 2 & -1 & 0 \\ -1 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix}$$

The Cartan matrices of (finite crystallographic) root systems are sometimes called *Cartan matrices of finite type*. This class of matrices is completely characterized by several elementary properties.

Theorem 2.4. *An integer $n \times n$ matrix $[a_{ij}]$ is a Cartan matrix of a root system if and only if*

(i) $a_{ii} = 2$ *for every* i*;*
(ii) $a_{ij} \leq 0$ *for any* $i \neq j$*, with* $a_{ij} = 0$ *if and only if* $a_{ji} = 0$*;*
(iii) *there exists a diagonal matrix* D *with positive diagonal entries such that* DAD^{-1} *is symmetric and positive definite.*

Remark 2.5. Condition (iii) can be replaced by

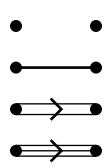
(iii′) *there exists a diagonal matrix* D' *with positive integer diagonal entries such that* $D'A$ *is symmetric and positive definite.*

Example 2.6. For the root systems $A_1 \times A_1$ and A_2, the 2×2 identity matrix serves as D. For B_2 and G_2, take $D = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{2} \end{bmatrix}$ and $D = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{3} \end{bmatrix}$, respectively.

The characterization in Theorem 2.4 can be used to completely classify the Cartan matrices of finite type, or the corresponding root systems. It turns out that each of those is built from blocks taken from a certain relatively short list. Let us be more precise.

A root system Φ is called *reducible* if Φ is a disjoint union of root systems Φ_1 and Φ_2 such that every $\beta_1 \in \Phi_1$ is normal to every $\beta_2 \in \Phi_2$. If such a decomposition does not exist, Φ is called *irreducible*. The parallel definition for Cartan matrices is that a Cartan matrix of finite type is *indecomposable* if its rows and columns cannot be simultaneously rearranged to bring the matrix into block-diagonal form with more than one block.

The Cartan matrices of finite type can be encoded by their *Dynkin diagrams*. The vertices of a Dynkin diagram are labeled by the elements of the indexing set I; thus they are in bijection with the simple roots. Each pair of vertices i and j is then connected as shown below (with the vertex i on the left):



if $a_{ij} = a_{ji} = 0$
if $a_{ij} = a_{ji} = -1$
if $a_{ij} = -1$ and $a_{ji} = -2$
if $a_{ij} = -1$ and $a_{ji} = -3$

(It follows from Theorem 2.4 that these are the only possible pairs of values for a_{ij} and a_{ji}. Cf. Example 2.2.)

Lemma 2.7. *A Cartan matrix of finite type (resp., a root system) is indecomposable (resp., irreducible) if and only if its Dynkin diagram is connected.*

Theorem 2.8 (Cartan-Killing classification of irreducible root systems and Cartan matrices of finite type)**.** *The complete list of Dynkin diagrams of irreducible root systems is presented in Figure 2.1.*

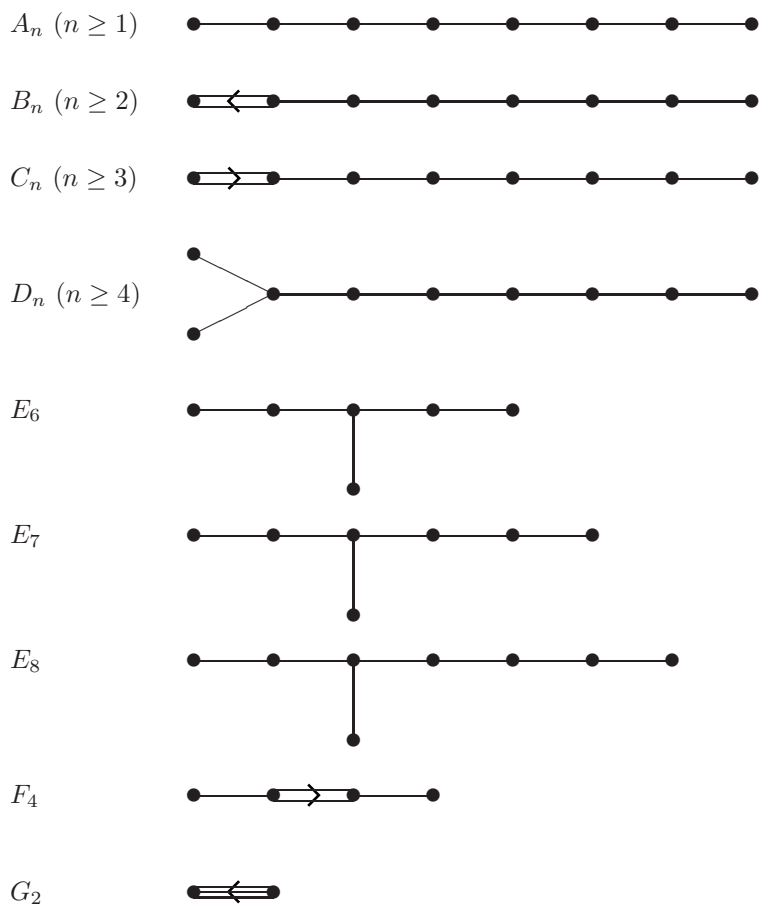


Figure 2.1. Dynkin diagrams of finite irreducible root systems.

Root systems are just one example among a large number of mathematical objects of "finite type" which are classified by (some class of) Dynkin diagrams. The appearance of the ubiquitous Dynkin diagrams in a variety of seemingly unrelated classification problems has fascinated several generations of mathematicians, and helped establish nontrivial connections between different areas of mathematics. See Section 2.3 and references therein.

2.2. Coxeter Groups

Let Φ be a (finite crystallographic) root system and $\alpha \neq \beta$ a pair of roots in Φ. The angle between the corresponding reflecting hyperplanes is a rational multiple of π with denominator 2, 3, 4 or 6. Thus the rotation $\sigma_\alpha \sigma_\beta$ has order 2, 3, 4, or 6 as an element of the associated reflection group W. The insight that the order of a product of reflections is directly related to the angle between the corresponding hyperplanes leads to the definition of a Coxeter group.

Definition 2.9. A *Coxeter system* (W, S) is a pair consisting of a group W together with a finite subset $S \subset W$ satisfying the following conditions:

(i) each $s \in S$ is an involution: $s^2 = 1$;
(ii) some pairs $\{s, t\} \subset S$ satisfy relations of the form $(st)^{m_{st}} = 1$ with $m_{st} \geq 2$;
(iii) the relations in (i)–(ii) form a presentation of the group W.

In other words, S generates W, and any identity in W is a formal consequence of (i)–(ii) and the axioms of a group.

A group W is called a *Coxeter group* if it has a presentation of the above form.

The following theorem demonstrates that the notion of a Coxeter group indeed captures the geometric essence of reflection groups.

Theorem 2.10. *Any finite Coxeter group is isomorphic to a reflection group.*

Conversely, a reflection group associated with a (finite crystallographic) root system Φ is a Coxeter group, in the following sense. Let Π be the set of simple roots in Φ. For each simple root $\alpha_i \in \Pi$, the associated *simple reflection* is $s_i \stackrel{\text{def}}{=} \sigma_{\alpha_i}$.

Theorem 2.11. *Let W be the group generated by the reflections $\{\sigma_\beta : \beta \in \Phi\}$. Let*

$$S = \{s_i\}_{i \in I} = \{\sigma_\alpha : \alpha \in \Pi\}$$

be the set of simple reflections. Then (W, S) is a Coxeter system.

Furthermore, W is a *crystallographic Coxeter group*, where the adjective "crystallographic" refers to restricting the integers m_{st} to the set $\{2, 3, 4, 6\}$.

2.3. Other "Finite Type" Classifications

The classification of root systems is similar or identical to several other classifications of objects of "finite type," briefly reviewed below.

Non-crystallographic root systems

Lifting the crystallographic restriction does not allow very many additional root systems. The only non-crystallographic irreducible finite root systems are those of types H_3, H_4 and $I_2(m)$ for $m = 5$ or $m \geq 7$. See [**34**].

Coxeter groups and reflection groups

By Theorems 2.10 and 2.11, the classification of finite Coxeter groups is parallel to the classification of reflection groups and is essentially the same as the classification of root systems. The difference is that the root systems B_n and C_n correspond to the same Coxeter group B_n. A Coxeter group is encoded by its *Coxeter diagram*, a graph whose vertex set is S, with an edge s—t whenever $m_{st} > 2$. If $m_{st} > 3$, the edge is labeled by m_{st}. Figure 2.2 shows the Coxeter diagrams of the finite irreducible Coxeter systems, including the non-crystallographic Coxeter groups H_3, H_4 and $I_2(m)$. The group G_2 appears as $I_2(6)$. See [**34**] for more details.

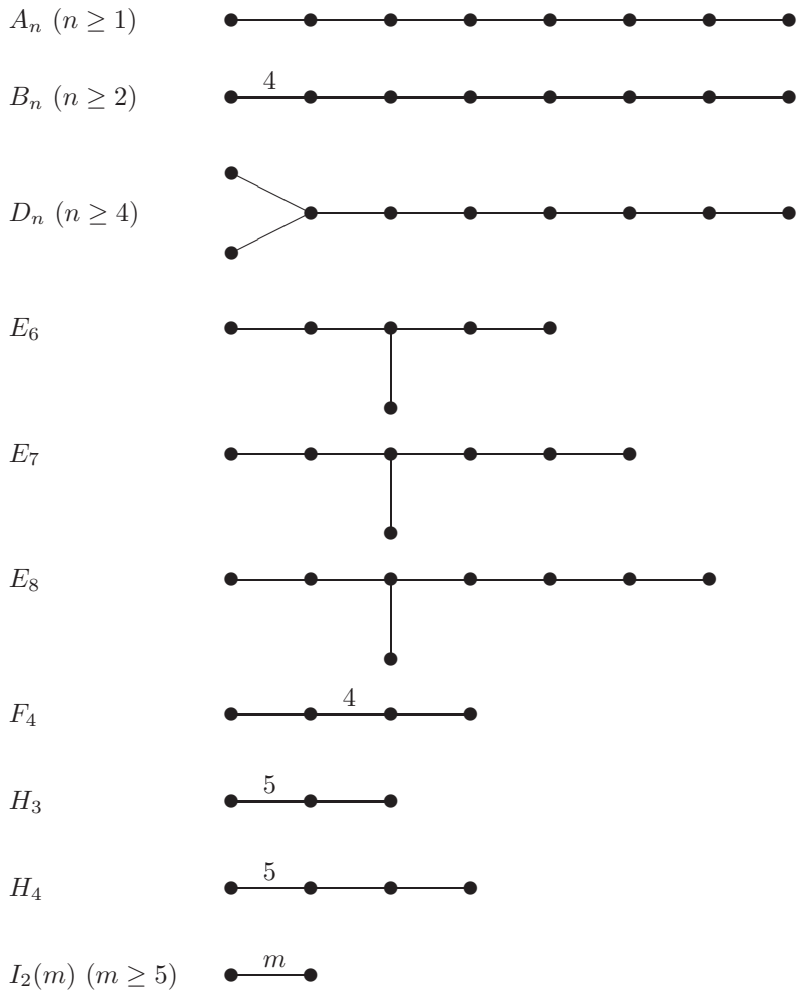


Figure 2.2. Coxeter diagrams of finite irreducible Coxeter systems

Regular polytopes

By Theorem 1.5, the symmetry group of a regular polytope is a reflection group. In fact, it is a Coxeter group whose Coxeter diagram is *linear*: the underlying graph is a path with no branching points. This narrows down the possibilities, leading to the conclusion that there are no other regular polytopes besides the ones described in Section 1.2. In particular, there are no "exceptional" regular polytopes beyond dimension 4: only simplices, cubes, and crosspolytopes. See [**14**].

Lie algebras

The original motivation for the Cartan-Killing classification of root systems came from Lie theory. Complex finite-dimensional simple Lie algebras correspond naturally, and one-to-one, to finite irreducible crystallographic root systems. There exist innumerable expositions of this classical subject; see, e.g., [**25**].

Quivers of finite type

A *quiver* is a directed graph; its *representation* assigns a complex vector space to each vertex, and a linear map to each directed edge. A quiver is *of finite type* if it has only a finite number of indecomposable representations (up to isomorphism); a representation is *indecomposable* if it cannot be obtained as a nontrivial direct sum. By Gabriel's Theorem, a quiver is of finite type if and only if its underlying graph is a Dynkin diagram of type A, D or E. See [**45**] and references therein.

Et cetera

And the list goes on: simple singularities, finite subgroups of $SU(2)$, symmetric matrices with nonnegative integer entries and eigenvalues between -2 and 2, *etc.* For more, see [**28, 33, 59**]. In Section 4.2, we will present yet another classification that is parallel to Cartan-Killing: the classification of the cluster algebras of finite type.

2.4. Reduced Words and Permutohedra

Each element $w \in W$ can be written as a product of elements of S:

$$w = s_{i_1} \cdots s_{i_\ell} .$$

A shortest factorization of this form (or the corresponding sequence of subscripts $(i_1, \ldots, i_\ell)$) is called a *reduced word* for w; the number of factors ℓ is called the *length* of w.

Any finite Coxeter group has a unique element $w_\circ$ of maximal length. In the symmetric group $\mathcal{S}_{n+1} = A_n$, this is the permutation $w_\circ$ that reverses the order of the elements of the set $\{1, \ldots, n+1\}$.

Example 2.12. Let $W = \mathcal{S}_4$ be the Coxeter group of type A_3. The standard choice of simple reflections yields $S = \{s_1, s_2, s_3\}$, where s_1, s_2 and s_3 are the transpositions which interchange 1 with 2, 2 with 3, and 3 with 4, respectively. (Cf. Example 1.7.)

The word $s_1 s_2 s_1 s_3 s_2 s_3$ is a non-reduced word for the permutation that interchanges 1 with 3 and 2 with 4. This permutation has two reduced words $s_2 s_1 s_3 s_2$ and $s_2 s_3 s_1 s_2$.

An example of a reduced word for $w_\circ$ is $s_1 s_2 s_1 s_3 s_2 s_1$. There are 16 such reduced words altogether. (Cf. Example 2.14 and Theorem 2.15.)

Recall from Section 1.2 that we label the regions R_w of the Coxeter arrangement by the elements of the reflection group W, so that R_w is the image of R_1 under the action of w. More generally, $R_{uv} = u(R_v)$.

Lemma 2.13. *In the Coxeter arrangement associated with a reflection group W, regions R_u and R_v are adjacent (that is, share a codimension* 1 *face) if and only if $u^{-1}v$ is a simple reflection.*

Thus, moving to an adjacent region is encoded by multiplying *on the right* by a simple reflection; cf. Figure 1.1. (Warning: this simple reflection is generally *not* the same as the reflection through the hyperplane separating the two adjacent regions.) Consequently, reduced words for an element $w \in W$ correspond to equivalence classes of paths from R_1 to R_w in the ambient space of the Coxeter arrangement. More precisely, we consider the paths that cross hyperplanes of the arrangement

one at a time, and cross each hyperplane at most once; two paths are equivalent if they cross the same hyperplanes in the same order.

In order to make the correspondence between paths and reduced words more explicit, one can restrict the paths to the edges of the W-*permutohedron*, a convex polytope that we will now define. Fix a point x in the interior of R_1. The W-permutohedron is the convex hull of the orbit of x under the action of W. The name "permutohedron" comes from the fact that the vertices of an A_n-permutohedron are obtained by permuting the coordinates of a generic point in $\mathbb{R}^{n+1}$.

Example 2.14. The A_2, B_2 and G_2 permutohedra are respectively a hexagon, an octagon and a dodecagon; under the right choices of x, these polygons are regular. Figures 2.3 and 2.4 show the permutohedra of types A_3 and B_3. Each of these realizations derives from a choice of $x \in R_1$ which makes the permutohedron an Archimedean solid, so that in particular its facets are all regular polygons. The non-crystallographic H_3-permutohedron is also an Archimedean solid[1].

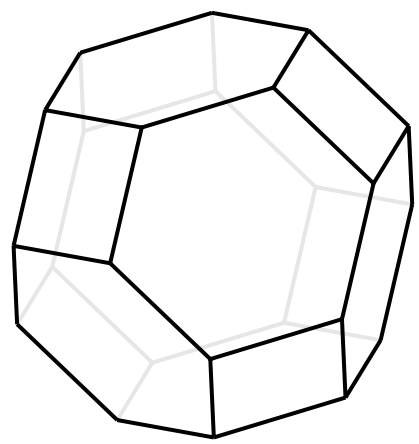


Figure 2.3. The permutohedron of type A_3

In both pictures, the bottom vertex can be associated with the identity element $1 \in W$, so that the top vertex is $w_\circ$. A reduced word for w corresponds to a path along edges from 1 to w which moves up in a monotone fashion. There are 16 such paths from 1 to $w_\circ$ in the A_3-permutohedron; cf. Example 2.12.

The following beautiful formula is due to R. Stanley [**49**].

Theorem 2.15. *The number of reduced words for $w_\circ$ in the reflection group A_n is*

$$\frac{\binom{n+1}{2}!}{1^n 3^{n-1} 5^{n-2} \cdots (2n-1)^1}.$$

[1]An *Archimedean solid* is a non-regular polytope whose all facets are regular polygons, and whose symmetry group acts transitively on vertices. In dimension 3, there are 13 Archimedean solids. The permutohedra of types A_3, B_3, and H_3 are also known as the *truncated octahedron*, *great rhombicuboctahedron*, and *great rhombicosidodecahedron*, respectively. See, e.g., [**55**].

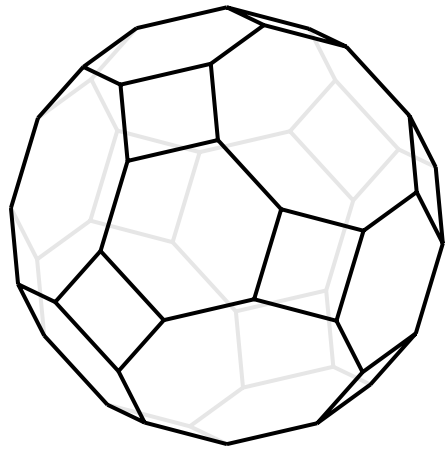


Figure 2.4. The permutohedron of type B_3

2.5. Coxeter Element and Coxeter Number

The underlying graph of the Coxeter diagram for a finite Coxeter group has no cycles. Hence it is bipartite, i.e., we can write a disjoint union $I = I_+ \cup I_-$ such that each of the sets I_+ and I_- is totally disconnected in the Coxeter diagram. An example is shown in Figure 2.5, where the elements of I_+ and I_- are marked by $+$ and $-$, respectively.

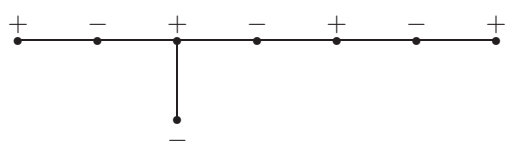


Figure 2.5. Bi-partition of the nodes of the Coxeter diagram of type E_8

The simple reflections associated with I_+ (resp., I_-) commute pairwise. Consequently, the following is well-defined:

$$c = \left(\prod_{i \in I_+} s_i\right)\left(\prod_{i \in I_-} s_i\right).$$

The element $c \in W$ is called the *Coxeter element*[2].

Example 2.16. In type A_n, let I_- (resp., I_+) consist of the odd (resp., even) numbers in $I = [n]$. Then for example in $A_5 = \mathcal{S}_6$, we have $c = s_2 s_4 s_1 s_3 s_5$.

Thinking of W as a reflection group, the Coxeter element c is an interesting orthogonal transformation. One important feature of c is that it fixes a certain two-dimensional plane L (as a set, not pointwise). The action of c on L can be analyzed to determine the order of c as an element of W. This order is called the *Coxeter number* of W, and is denoted by h.

[2] More broadly, one often calls the product of the elements in S (in any order) a Coxeter element, but for our present purposes the definition above will do.

Example 2.17. Figure 2.6 shows the Coxeter arrangement of type A_3 and the plane L fixed by the Coxeter element $c = s_2 s_1 s_3$ (dotted). The great circles represent the intersections of the six reflecting hyperplanes with a unit hemisphere. The sphere is opaque, so only half of each circle is visible, and appears either as a *half* of an ellipse or as a straight line segment. (The "equator" does not represent a hyperplane in the arrangement.) The restriction of c onto L has order 4, so the Coxeter number for A_3 is $h = 4$.

Example 2.18. Figure 2.7 is a similar picture for B_3, illustrating that the Coxeter number for B_3 is $h = 6$. In this picture, the equator does represent a hyperplane in the arrangement.

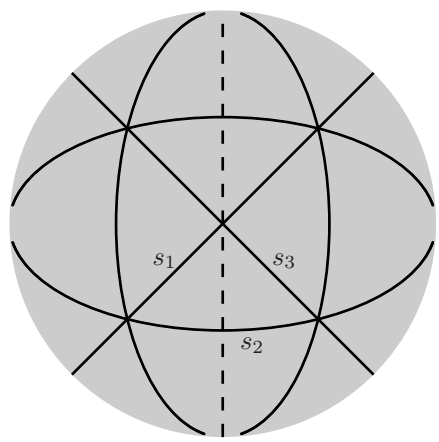


Figure 2.6. The Coxeter arrangement A_3 and the plane fixed by the Coxeter element

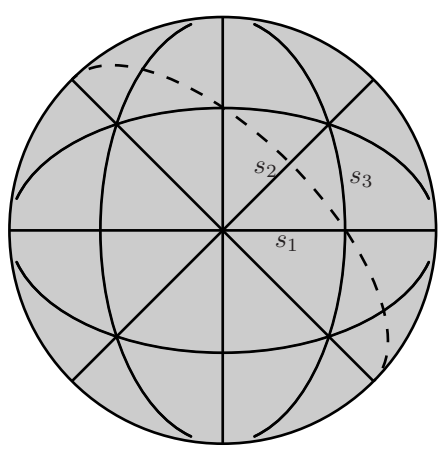


Figure 2.7. The Coxeter arrangement B_3 and the plane fixed by the Coxeter element

The action of c on L also leads to a determination of its eigenvalues, which all have the form $e^{2mi\pi/h}$, where m is a positive integer less than h. The n values of m which arise in this way are called the *exponents* of W. We denote the exponents by $e_1, \ldots, e_n$. They pop up everywhere in the combinatorics of root systems and Coxeter groups. For instance, the order (i.e., cardinality) of W is expressed in terms of the exponents by

$$|W| = \prod_{i=1}^{n} (e_i + 1)\,.$$

See Section 5.1 for more examples.

For a finite irreducible Coxeter group W Figure 2.8 tabulates some classical numerical invariants associated to W and the corresponding (not necessarily crystallographic) root system Φ.

type of Φ	$\lvert\Phi_+\rvert$	h	$e_1, \ldots, e_n$	$\lvert W\rvert$
A_n	$n(n+1)/2$	$n+1$	$1, 2, \ldots, n$	$(n+1)!$
B_n, C_n	n^2	$2n$	$1, 3, 5, \ldots, 2n-1$	$2^n n!$
D_n	$n(n-1)$	$2(n-1)$	$1, 3, 5, \ldots, 2n-3, n-1$	$2^{n-1} n!$
E_6	36	12	$1, 4, 5, 7, 8, 11$	$2^7 3^4 5$
E_7	63	18	$1, 5, 7, 9, 11, 13, 17$	$2^{10} 3^4 5 \cdot 7$
E_8	120	30	$1, 7, 11, 13, 17, 19, 23, 29$	$2^{14} 3^5 5^2 7$
F_4	24	12	$1, 5, 7, 11$	$2^7 3^2$
G_2	6	6	$1, 5$	$2^2 3$
H_3	15	10	$1, 5, 9$	$2^3 3 \cdot 5$
H_4	60	30	$1, 11, 19, 29$	$2^6 3^2 5^2$
$I_2(m)$	m	m	$1, m-1$	$2m$

Figure 2.8. Number of positive roots, Coxeter number, exponents, and the order of W.

LECTURE 3
Associahedra and Mutations

3.1. Associahedron

We start by discussing two classical problems of combinatorial enumeration.

(i) Count the number of *bracketings* (*parenthesizations*) of a non-associative product of $n+2$ factors. Note that we need n pairs of brackets in order to make the product unambiguous.

(ii) Count the number of *triangulations* of a convex $(n+3)$-gon by diagonals. Note that each triangulation involves exactly n diagonals.

Example 3.1. In the special cases $n = 1, 2, 3$, there are, respectively:

- 2 bracketings $(ab)c$ and $a(bc)$ of a product of 3 factors;
- 5 bracketings $((ab)c)d$, $(a(bc))d$, $a((bc)d)$, $(ab)(cd)$, and $a(b(cd))$ of a product of 4 factors;
- 14 bracketings of a product of 5 factors (check!).

As to triangulations, there are:

- 2 triangulations of a convex quadrilateral ($n = 1$);
- 5 triangulations of a pentagon ($n = 2$, Figure 3.3);
- 14 triangulations of a hexagon ($n = 3$, Figure 3.4).

Theorem 3.2. *Both bracketings and triangulations described above are enumerated by the Catalan numbers* $\frac{1}{n+2}\binom{2n+2}{n+1}$.

There are a great many families of combinatorial objects enumerated by the Catalan numbers; more than a hundred of those are listed in [**50**]. This list includes: ballot sequences; Young diagrams and tableaux satisfying certain restrictions; non-crossing partitions; trees of various kinds; Dyck paths; permutations avoiding patterns of length 3; and much more. In Lecture 5, we will discuss several additional members of the "Catalan family," together with their analogues for arbitrary root systems. (We will see that the ordinary Catalan numerology should be considered as "type A.")

A bijection between bracketings and triangulations is described in Figure 3.1.

For a fixed n, the bracketings naturally form the set of vertices of a graph whose edges correspond to applications of the associativity axiom. Figure 3.2 shows this graph for $n = 2$.

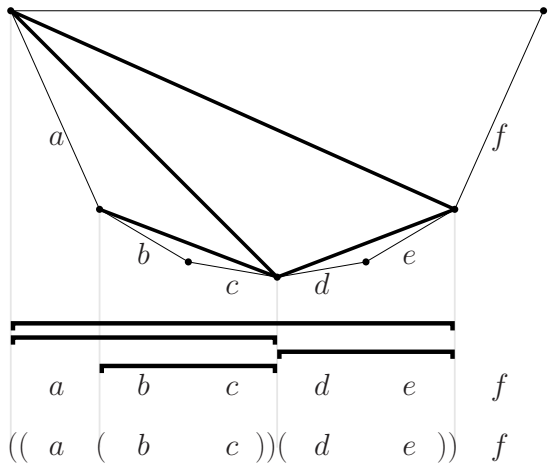


Figure 3.1. The bijection between triangulations and bracketings.

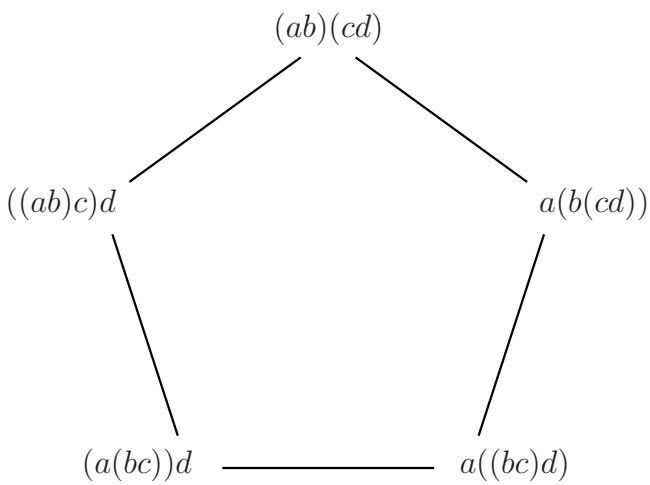


Figure 3.2. Applying associativity to the bracketings of *abcd*.

The bijection illustrated in Figure 3.1 translates an application of the associativity axiom into a diagonal *flip* on the corresponding triangulation. That is, one removes a diagonal to create a quadrilateral, then replaces the removed diagonal with the other diagonal of the quadrilateral.

We call the graph defined by diagonal flips the *exchange graph*. The exchange graphs for $n = 2$ and $n = 3$ are shown in Figures 3.3 and 3.4.

The drawing of the exchange graph in Figure 3.4 fails to convey its crucial property: this exchange graph is the 1-skeleton of a convex polytope, the 3-dimensional *associahedron*. (Sometimes it is also called the *Stasheff polytope*, after J. Stasheff, who first defined it in [**52**].) Figure 3.5 shows a polytopal realization of this associahedron.

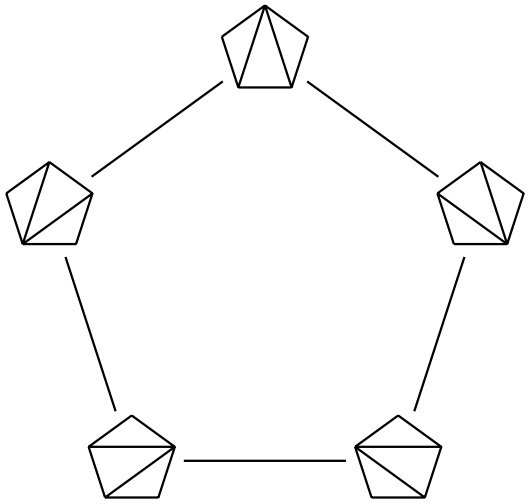


Figure 3.3. The exchange graph for triangulations of a pentagon.

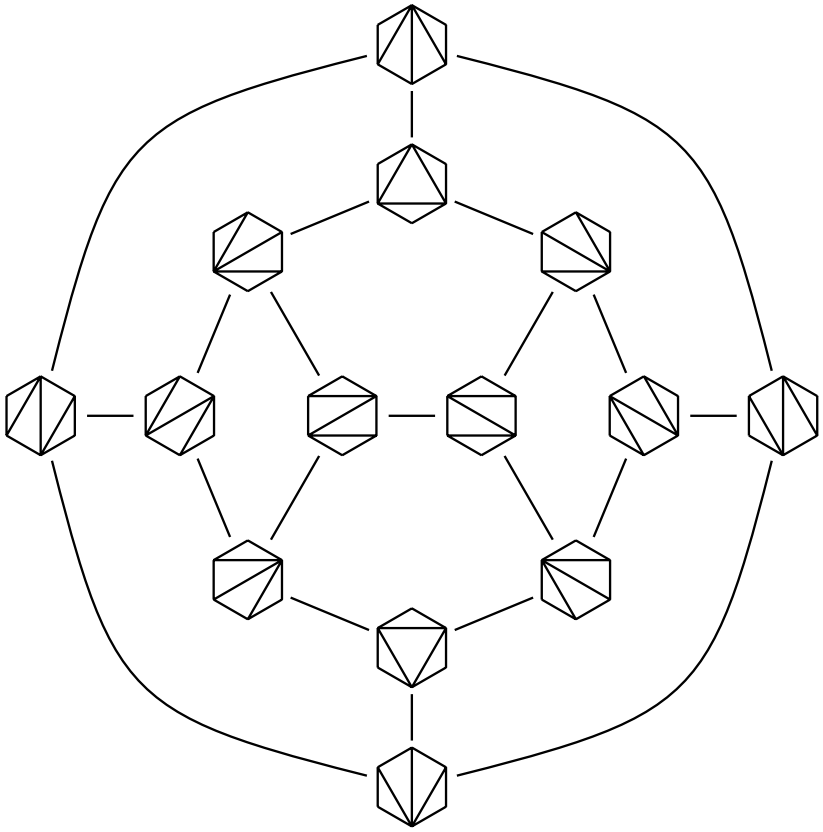


Figure 3.4. The exchange graph for triangulations of a hexagon.

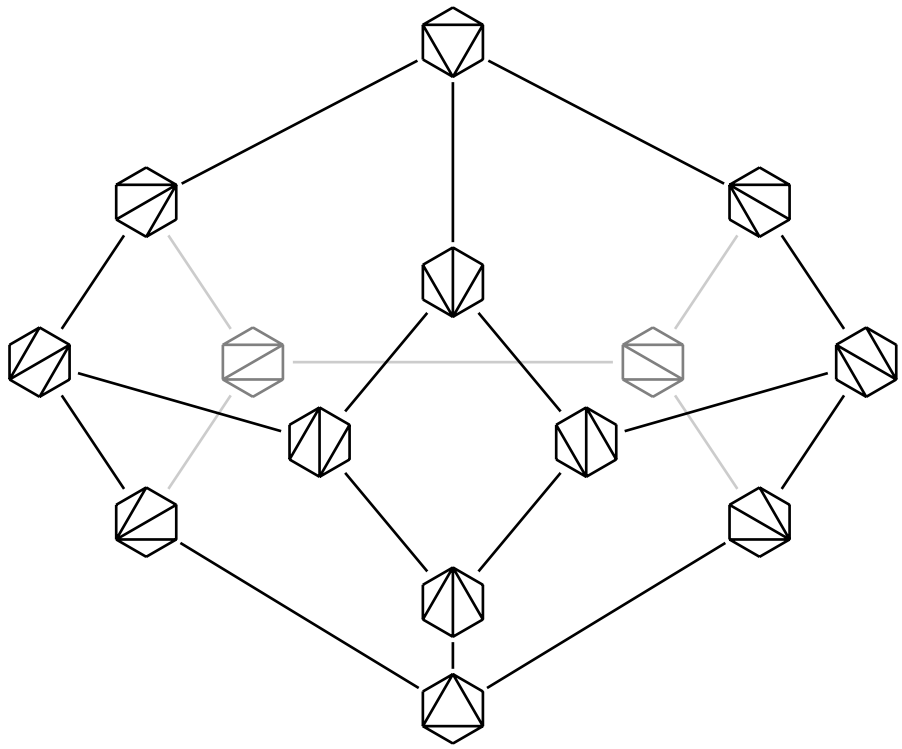


Figure 3.5. The 3-dimensional associahedron.

In order to formally define the n-dimensional associahedron, we start by describing the object which is dual to it, in the same sense in which the octahedron is dual to the cube, and the dodecahedron is dual to the icosahedron.

Definition 3.3 (The dual complex of an associahedron)**.** Consider the following simplicial complex:

vertices:	diagonals of a convex $(n+3)$-gon
simplices:	partial triangulations of the $(n+3)$-gon (viewed as collections of non-crossing diagonals)
maximal simplices:	triangulations of the $(n+3)$-gon (collections of n non-crossing diagonals).

Figure 3.6 shows this simplicial complex for $n = 3$, superimposed on a faint copy of the exchange graph. Note that the facial structures of the 3-dimensional associahedron and its dual complex are indeed "dual" to each other: two vertices of one polyhedron are adjacent if and only if the corresponding faces of the other polyhedron share an edge.

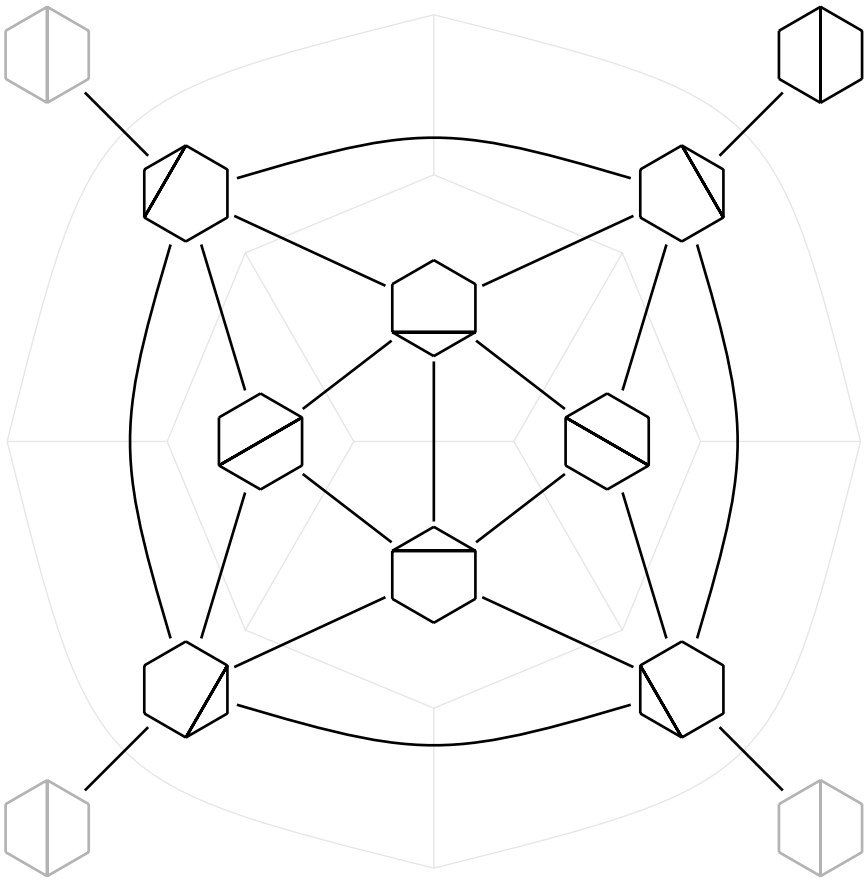


Figure 3.6. The simplicial complex dual to the 3-dimensional associahedron.

It is not clear *a priori* that these complexes are topological spheres. But, as already mentioned, more is true.

Theorem 3.4. *The simplicial complex described in Definition 3.3 can be realized as the boundary of an n-dimensional convex polytope.*

Theorem 3.4 (or its equivalent reformulations) were proved independently by J. Milnor, M. Haiman, and C. W. Lee (first published proof [**36**]). This theorem also follows as a special case of the very general theory of secondary polytopes developed by I. M. Gelfand, M. Kapranov and A. Zelevinsky [**30**].

Definition 3.5 (The associahedron)**.** The n-dimensional *associahedron* is the convex polytope (defined up to combinatorial equivalence) that is dual (or polar, see [**58**, Sec. 2.3]) to the polytope of Theorem 3.4.

The facial structure of an associahedron as a cell complex is dual to that of its polar:

(6)

vertices:	triangulations
faces:	partial triangulations
facets:	diagonals
edges:	diagonal flips

The labeling of the facets of an n-dimensional associahedron by the diagonals of an $(n+3)$-gon is illustrated in Figure 3.7 for the special case $n = 3$ (compare to Figure 3.4).

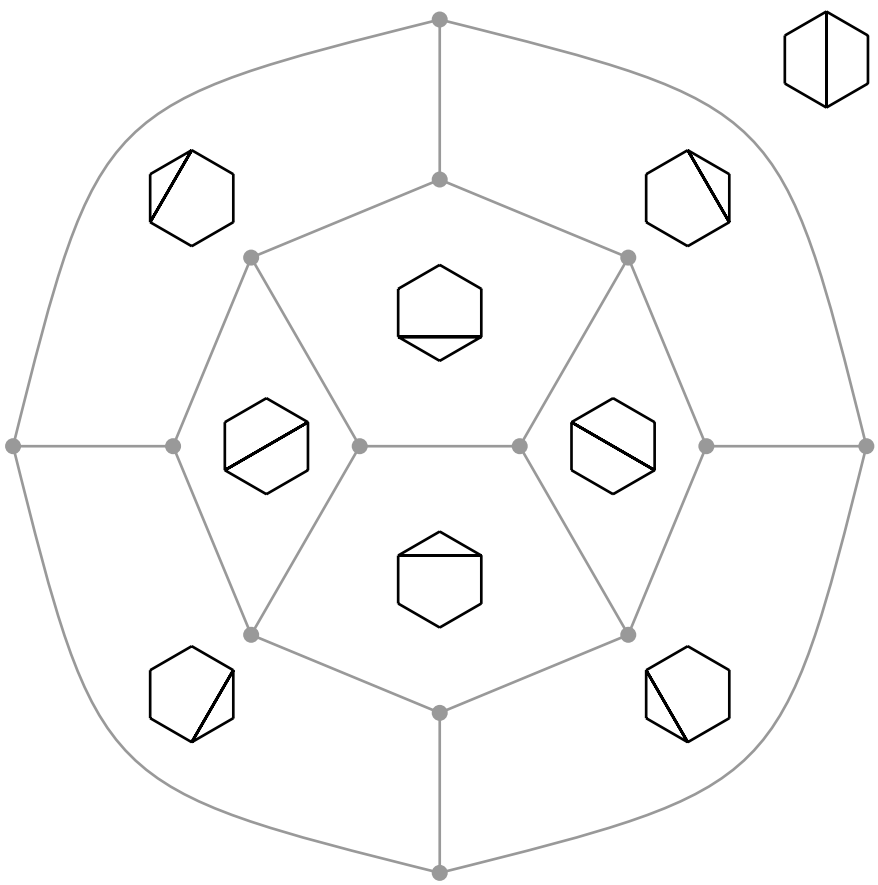


Figure 3.7. Labeling the facets of the associahedron by diagonals

We note that we could have defined the associahedron directly, as a cell complex whose cell structure is described by (6). (This would require resolving some technical issues that we would rather avoid here.) The fact that these cell complexes are polytopal—i.e., the fact that a combinatorially defined associahedron can be realized as a convex polytope—is essentially equivalent to Theorem 3.4.

Associahedra play an important role in homotopy theory and the study of operads [**53**], in the analysis of real moduli/configuration spaces [**16**], and other branches of mathematics. In these notes, we restrict our attention to the combinatorial aspects of the associahedra.

An n-dimensional polytope is called *simple* if every vertex is incident to exactly n edges. This is the case for the associahedron, as every triangulation of an $(n+3)$-gon is adjacent to precisely n others in the exchange graph.

Much is known about the facial structure and enumerative invariants of the associahedron. For example, each face is the direct product of smaller associahedra. The entries of the *h-vector* of the associahedron are the Narayana numbers (see Section 5.2). This allows one to calculate the number of faces of each dimension.

3.2. Cyclohedron

The n-dimensional *cyclohedron* (also known as the *Bott-Taubes polytope* [**8**]) is constructed similarly to the associahedron using centrally-symmetric triangulations of a regular $(2n+2)$-gon. Each edge of the cyclohedron represents either a diagonal flip involving two diameters of the polygon, or a pair of two centrally-symmetric diagonal flips. Figures 3.8 and 3.9 show the 2- and 3-dimensional cyclohedra respectively. As these figures suggest, the cyclohedron is a convex polytope for any n. Explicit polytopal realizations of cyclohedra were constructed by M. Markl [**38**] and R. Simion [**47**]. Each face of a cyclohedron is a product of smaller cyclohedra and associahedra.

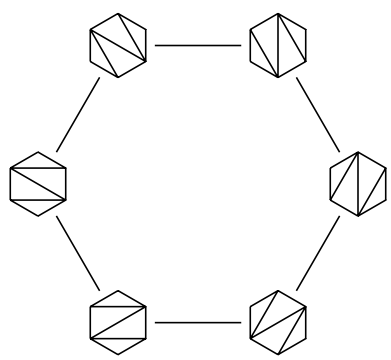


Figure 3.8. The 2-dimensional cyclohedron

Further details about the combinatorics of cyclohedra, and about their appearance in the study of configuration spaces can be found in [**17**].

The geometry of associahedra and cyclohedra is related to the geometry of permutohedra, as the following theorem (due to Tonks [**54**]) shows.

Theorem 3.6. *The* 1*-skeleton of the* n*-dimensional associahedron (resp., cyclohedron) can be obtained from the* 1*-skeleton of the permutohedron of type* A_n *(resp., type* B_n*) by contraction of edges.*

Theorem 3.6 is further discussed in Section 5.4 in connection with Theorem 5.11. For $n = 3$, the theorem is illustrated in Figure 3.10. (Cf. Figures 2.3 and 2.4.)

In light of Theorem 3.6, the cyclohedron can be viewed as a "type B counterpart" of the associahedron (which is a "type A" object).

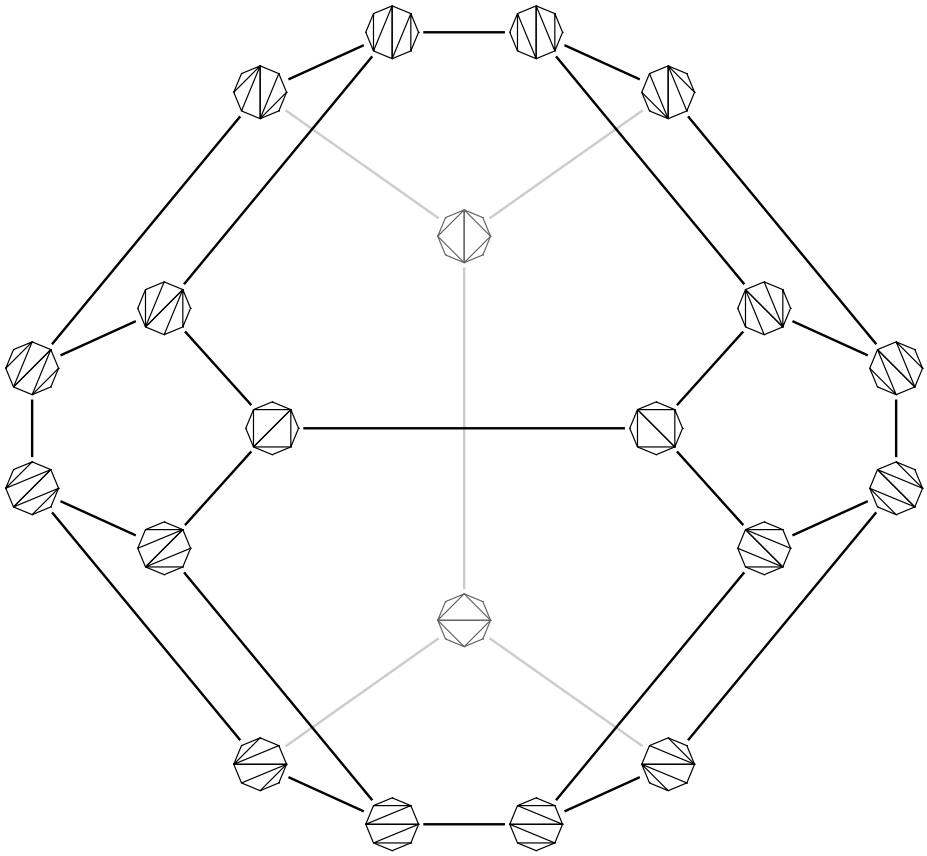


Figure 3.9. The 3-dimensional cyclohedron

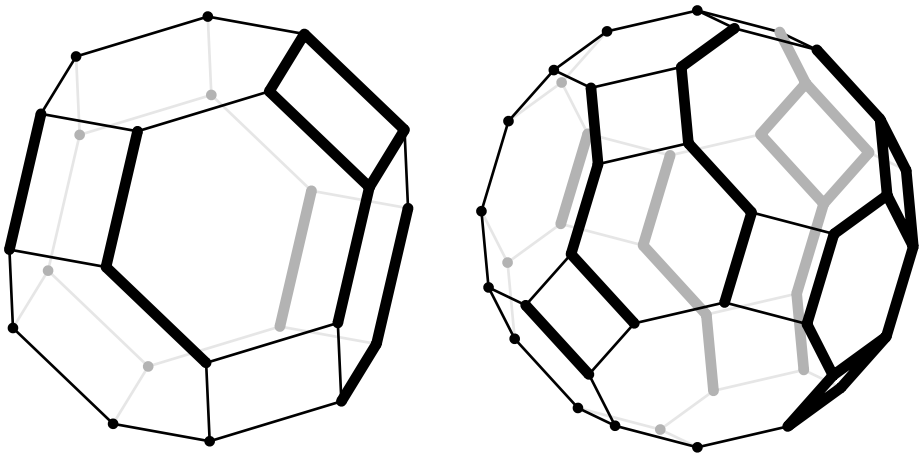


Figure 3.10. Contracting edges of permutohedra of types A_3 and B_3 yields an associahedron and a cyclohedron

3.3. Matrix Mutations

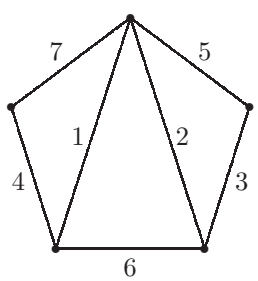
Having looked closely at the associahedron and the cyclohedron, one is naturally led to wonder: are these two just a pair of isolated constructions, or is there a general framework that includes them as special cases? Given that the associahedra and the cyclohedra are related to the root systems of types A and B, respectively, is there a classification of polytopes arising within this framework that is similar to the Cartan-Killing classification of root systems?

As a first step towards answering these questions, we will develop the machinery of *matrix mutations*, which encode the combinatorics of various models similar to the associahedron and the cyclohedron. We begin our discussion of matrix mutations by continuing the example of the associahedron.

Fix a triangulation T of the $(n+3)$-gon. Label the n diagonals of T arbitrarily by the numbers $1, \ldots, n$, and label the $n+3$ sides of T by the numbers $n+1, \ldots, 2n+3$. The combinatorics of T can be encoded by the (signed) *edge-adjacency matrix* $\tilde{B} = (b_{ij})$. This is the $(2n+3) \times n$ matrix whose entries are given by

$$b_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ label two sides in some triangle of } T \text{ so that } j \text{ follows } i \\ & \quad \text{in the clockwise traversal of the triangle's boundary;} \\ -1 & \text{if the same holds, with the counter-clockwise order;} \\ 0 & \text{otherwise.} \end{cases}$$

Note that the first index i is a label for a side or a diagonal of the $(n+3)$-gon, while the second index j must label a diagonal. The *principal part* of $\tilde{B}$ is an $n \times n$ submatrix $B = (b_{ij})_{i,j \in [n]}$ that encodes the signed adjacencies between the diagonals of T. An example is shown in Figure 3.11.



$$\tilde{B} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \\ 0 & 1 \\ -1 & 0 \\ 0 & -1 \\ 1 & -1 \\ 1 & 0 \end{bmatrix} \qquad B = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

Figure 3.11. Matrices B and $\tilde{B}$ for a triangulation

In the language of matrices $\tilde{B}$ and B, diagonal flips can be described as certain transformations called matrix mutations.

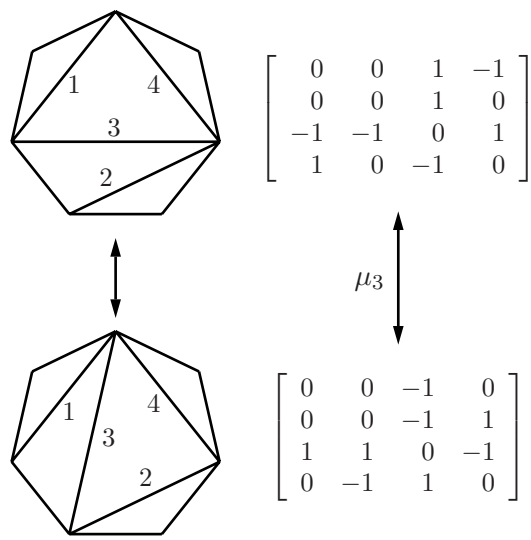
Definition 3.7. Let $B = (b_{ij})$ and $B' = (b'_{ij})$ be integer matrices. We say that B' is obtained from B by a *matrix mutation* in direction k, and write $B' = \mu_k(B)$, if

$$b'_{ij} = \begin{cases} -b_{ij} & \text{if } k \in \{i, j\}; \\ b_{ij} + |b_{ik}| b_{kj} & \text{if } k \notin \{i, j\} \text{ and } b_{ik} b_{kj} > 0; \\ b_{ij} & \text{otherwise.} \end{cases} \tag{7}$$

It is easy to check that a matrix mutation is an involution, i.e., $\mu_k(\mu_k(B)) = B$.

Lemma 3.8. *Assume that $\tilde{B}$ and $\tilde{B}'$ (resp., B and B') are the edge-adjacency matrices (resp., their principal parts) for two triangulations T and T' obtained from each other by flipping the diagonal numbered k; the rest of the labels are the same in T and T'. Then $\tilde{B}' = \mu_k(\tilde{B})$ (resp., $B' = \mu_k(B)$).*

Lemma 3.8 is illustrated in Figures 3.12 and 3.13. Note that the numbering of diagonals used in defining the matrices $\tilde{B}$ and B can change as we move along the exchange graph. For instance, the sequence of 5 flips shown in Figure 3.13 results in switching the labels of the two diagonals.



$$\begin{bmatrix} 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 0 \\ -1 & -1 & 0 & 1 \\ 1 & 0 & -1 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & -1 & 1 \\ 1 & 1 & 0 & -1 \\ 0 & -1 & 1 & 0 \end{bmatrix}$$

Figure 3.12. A diagonal flip and the corresponding matrix mutation

One can similarly define edge-adjacency matrices for centrally symmetric triangulations (those matrices will have entries 0, ± 1, and ± 2), and verify that cyclohedral flips translate precisely into matrix mutations.

3.4. Exchange Relations

We next introduce a set of algebraic (more precisely, *birational*) transformations that will go hand in hand with the matrix mutations. We start by explaining this construction in the case of an associahedron.

Let us fix an arbitrary *initial triangulation* $T_\circ$ of a convex $(n+3)$-gon, and introduce a variable for each diagonal of this triangulation, and also for each side of the $(n+3)$-gon. This gives $2n+3$ variables altogether. We are now going to associate a rational function in these $2n+3$ variables to *every* diagonal of the $(n+3)$-gon. This will be done in a recursive fashion. Whenever we perform a diagonal flip as the one shown in Figure 3.14, all but one rational functions associated to the current triangulation remain unchanged: the rational function x associated with the diagonal being removed gets replaced by the rational function x' associated with the "new" diagonal, where x' is determined from the *exchange relation*

$$x\,x' = a\,c + b\,d\,. \tag{8}$$

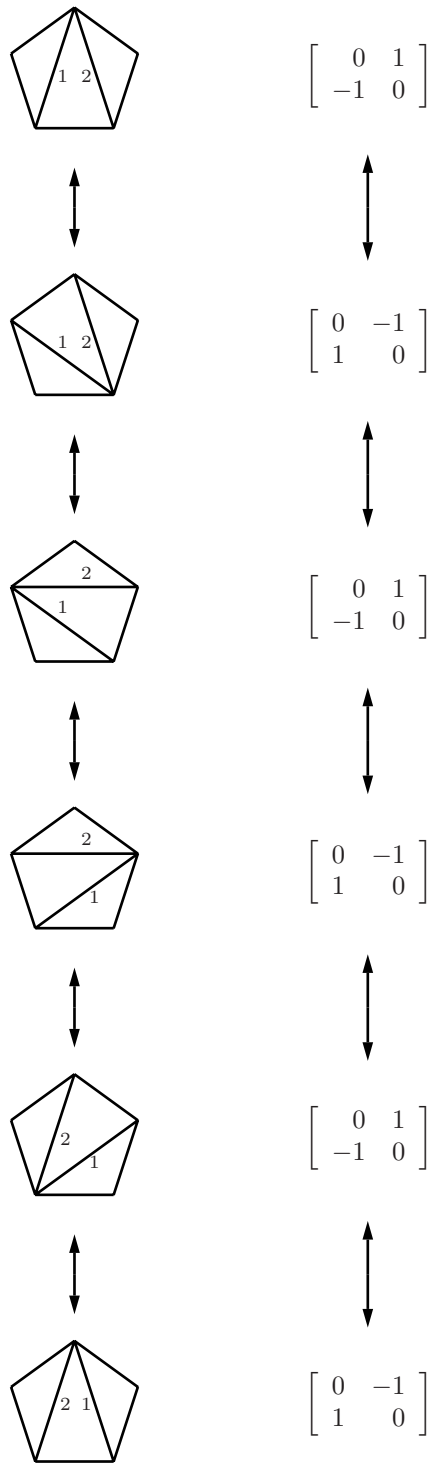


Figure 3.13. Diagonal flips in a pentagon, and the corresponding matrix mutations

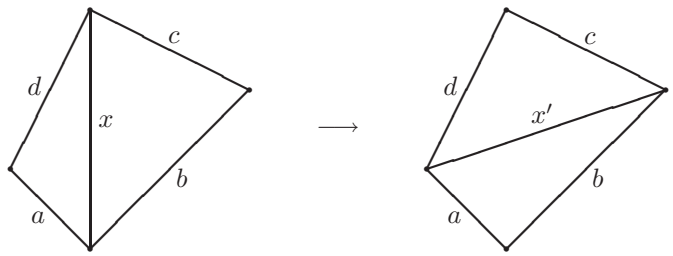


Figure 3.14. A diagonal flip

Lemma 3.9. *The rational function x_γ associated to each diagonal γ does not depend on the particular choice of a sequence of flips that connects the initial triangulation with another one containing γ.*

Lemma 3.9 can be rephrased as saying that there are no "monodromies" associated with sequences of flips that begin and end at the same triangulation.

To illustrate Lemma 3.9, consider the triangulations of a pentagon (i.e., $n = 2$). We label the sides of the pentagon by the variables q_1, q_2, q_3, q_4, q_5, as shown in Figure 3.15.

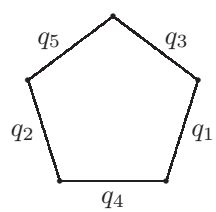


Figure 3.15. Labeling the sides of a pentagon

We then label the diagonals incident to the top vertex by the variables y_1 and y_2. Thus, our initial triangulation $T_\circ$ appears at the top of Figure 3.16. The rational functions y_3, y_4, y_5 associated with the remaining three diagonals are then computed from the exchange relations associated with the flips shown in Figure 3.16.

Starting at the top of Figure 3.16 and moving clockwise, we recursively express $y_3, y_4, \ldots$ in terms of y_1, y_2 and $q_1, \ldots, q_5$:

$$y_3 = \frac{q_2 y_2 + q_4 q_5}{y_1}\,,$$

$$y_4 = \frac{q_3 y_3 + q_5 q_1}{y_2} = \frac{q_3 q_2 y_2 + q_3 q_4 q_5 + q_5 q_1 y_1}{y_1 y_2}\,,$$

$$y_5 = \frac{q_4 y_4 + q_1 q_2}{y_3} = \cdots = \frac{q_3 q_4 + q_1 y_1}{y_2} \quad \text{(check!)},$$

and, finally,

$$y_1 = \frac{q_5 y_5 + q_2 q_3}{y_4} = \cdots = y_1\,,$$

$$y_2 = \frac{q_1 y_1 + q_3 q_4}{y_5} = \cdots = y_2\,,$$

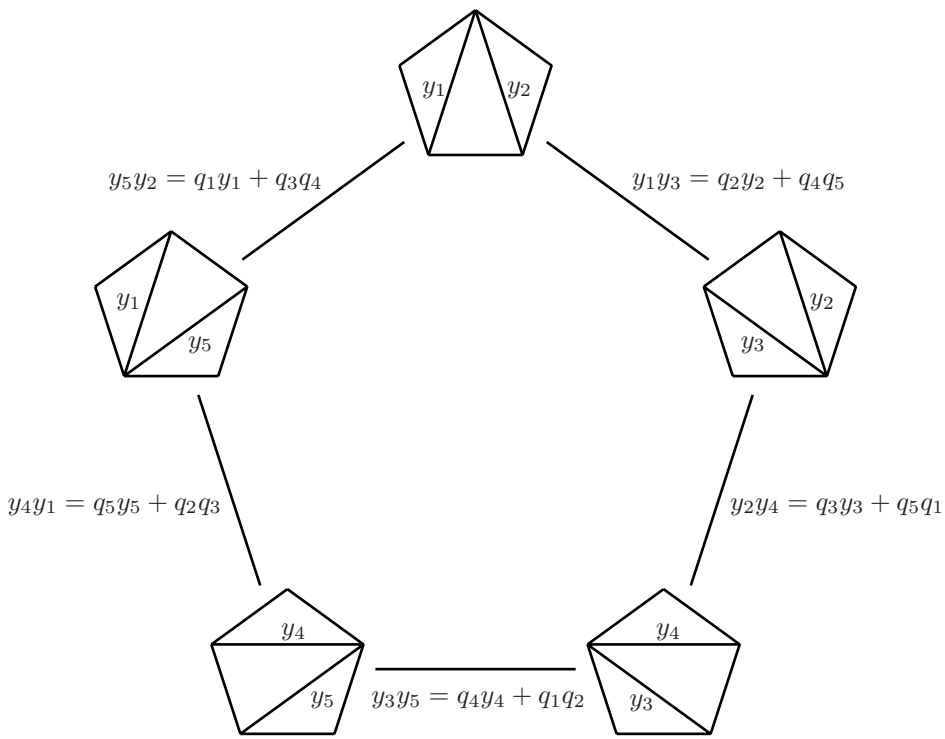


Figure 3.16. Exchange relations for the flips in a pentagon

recovering the original values and completing the cycle.

We note that under the specialization $q_1 = \cdots = q_5 = 1$, the phenomenon we just observed is nothing else but the 5-periodicity of the pentagon recurrence, which we have thus generalized.

Lemma 3.9 is a special case of a much more general result from the theory of *cluster algebras*. It can also be proved directly in at least two different ways briefly sketched below; these proofs point at connections of this subject to other areas of mathematics.

Ptolemy's Theorem and hyperbolic geometry

The classical Ptolemy's Theorem asserts that in an inscribed quadrilateral, the sum of the products of the two pairs of opposite sides equals the product of the two diagonals. This relation looks exactly like the exchange relation (8). It suggests that one can prove Lemma 3.9 simply by interpreting the rational function associated with each side or diagonal as the Euclidean length of the corresponding segment. There is however a problem with this type of argument: the space of inscribed $(n+3)$-gons (up to congruence) is $(n+3)$-dimensional, whereas we need $2n+3$ independent variables in our setup.

The problem can be resolved by passing from Euclidean to hyperbolic geometry, where an analogue of Ptolemy's Theorem holds, and where one can "cook up" the required additional degrees of freedom. For much more on this topic, see [**24, 29**].

Plücker coordinates on the Grassmannian $\mathrm{Gr}(2, n+3)$

Take a $2 \times (n+3)$ matrix $z = (z_{ij})$. For any $k, l \in [n+3]$, $k < l$, let us denote by

$$P_{kl} = \det \begin{pmatrix} z_{1k} & z_{1l} \\ z_{2k} & z_{2l} \end{pmatrix}$$

the 2×2 minor of z that occupies columns k and l. These minors are the homogeneous *Plücker coordinates* of the row span of z as an element of the *Grassmannian* $\mathrm{Gr}(2, n+3)$ of all 2-dimensional subspaces of an $(n+3)$-space. See, e.g., [**25**].

It is easy to check (the special case of) the *Grassmann-Plücker relations*:

$$P_{ik}P_{jl} = P_{ij}P_{kl} + P_{il}P_{jk}\,.$$

Once again, one recognizes the exchange relation (8). It is straightforward to construct, for a particular special choice of initial triangulation $T_\circ$, a matrix z for which the values of the minors P_{kl} corresponding to the sides and diagonals of $T_\circ$ are equal to the variables associated with these segments. It then follows by induction that the rational function associated to every diagonal is equal to the corresponding minor P_{kl}, implying Lemma 3.9.

LECTURE 4
Cluster Algebras

Our next task is to create a general axiomatic theory of mutations ("flips") and exchanges, using the above examples as prototypes. This will lead us to the basic notions and results of the theory of cluster algebras. Cluster algebras were introduced in [**20**] as a combinatorial/algebraic framework for the study of dual canonical bases and related total positivity phenomena. They since found applications in higher Teichmüller theory and representation theory of quivers, among other fields. All these motivations and applications will remain behind the scenes in these lectures.

Most of this lecture is based on [**19, 20, 21**]. Sections 4.4 and 4.5 are based on [**13**] and [**4, 23**], respectively.

4.1. Seeds and Clusters

Consider a diagonal flip that transforms a triangulation T of a convex $(n+3)$-gon into another triangulation T', as shown in Figure 3.14. The corresponding exchange relation (8) can be written entirely in terms of the edge-adjacency matrix $\tilde{B}$. To be more precise, let us assume, as before, that the diagonals of T have been labeled in some way by the numbers $1, \ldots, n$, whereas the sides of the $(n+3)$-gon have been assigned the labels going from $n+1$ through $m = 2n+3$. The labeling for T' is the same except for the one diagonal (say, labeled k) that is getting exchanged between T and T'.

This labeling of sides and diagonals of T allows us to (temporarily) denote the associated rational functions by $x_1, \ldots, x_m$. For T', we get the same rational functions except that x_k is replaced by x'_k. Then the exchange relation under consideration takes the form

$$x_k\, x'_k = \prod_{\substack{b_{ik}>0 \\ 1\le i\le m}} x_i^{b_{ik}} + \prod_{\substack{b_{ik}<0 \\ 1\le i\le m}} x_i^{-b_{ik}}\,. \tag{9}$$

In other words, the right-hand side of (9) is the sum of two monomials whose exponents are the absolute values of the entries in the kth column of $\tilde{B}$, while the sign of an entry determines which monomial the corresponding term contributes to.

Example 4.1. Let T be the triangulation of a pentagon in Figure 3.11, with its edges labeled $1, \ldots, 7$ as shown. The exchange relations corresponding to flipping

the diagonals 1 and 2 are, respectively:

$$x_1x_1' = x_6x_7 + x_2x_4\,,$$
$$x_2x_2' = x_1x_3 + x_5x_6\,,$$

in agreement with (9).

To summarize, both the combinatorics of flips and the algebra of exchange relations can be encoded entirely in terms of the matrices $\tilde{B}$ using, first, the machinery of matrix mutations and, second, the "birational dynamics" governed by (9). We shall now use this observation to lay out the axioms of a cluster algebra. The formulation of these axioms will require some technical preparation, which hopefully will make sense to the reader in light of the examples discussed above.

A cluster algebra $\mathcal{A}$ is a commutative ring contained in an *ambient field* $\mathcal{F}$ isomorphic to the field of rational functions in m variables over $\mathbb{Q}$. (Think of the rational functions in the variables associated with the sides and diagonals of a fixed initial triangulation.)

$\mathcal{A}$ is generated inside $\mathcal{F}$ by a (possibly infinite) set of generators. These generators are obtained from an initial *seed* via an iterative process of *seed mutations* which follows a set of canonical rules.

A *seed* in $\mathcal{F}$ is a pair $(\tilde{\mathbf{x}}, \tilde{B})$, where

- $\tilde{\mathbf{x}} = \{x_1, \ldots, x_m\}$ is a set[1] of m algebraically independent generators of $\mathcal{F}$, which is split into a disjoint union of an n-element *cluster* $\mathbf{x} = \{x_1, \ldots, x_n\}$ and an $(m-n)$-element set of *frozen variables* $\mathbf{c} = \{x_{n+1}, \ldots, x_m\}$;
- $\tilde{B} = (b_{ij})$ is an $m \times n$ integer matrix of rank n whose *principal part* $B = (b_{ij})_{i,j\in[n]}$ is *skew-symmetrizable*, i.e., there exists a diagonal matrix D with positive diagonal entries such that DBD^{-1} is skew-symmetric.

(Equivalently, there exist positive integers $d_1, \ldots, d_n$ such that $d_ib_{ij} = -d_jb_{ji}$ for all i and j.) The matrix B is called the *exchange matrix* of a seed.

A *seed mutation* μ_k in direction $k \in \{1, \ldots, n\}$ transforms a seed $(\tilde{\mathbf{x}}, \tilde{B})$ into another seed $(\tilde{\mathbf{x}}', \tilde{B}')$ defined as follows:

- $\tilde{\mathbf{x}}' = \tilde{\mathbf{x}} - \{x_k\} \cup \{x_k'\}$, where x_k' is found from the exchange relation (9);
- $\tilde{B}' = \mu_k(\tilde{B})$, i.e., $\tilde{B}$ undergoes a matrix mutation (hence so does B).

The following lemma justifies the definition of a seed mutation by showing that $(\tilde{\mathbf{x}}', \tilde{B}')$ is indeed a seed.

Lemma 4.2. *Matrix mutations preserve the rank of a matrix. If B is skew-symmetrizable, then so is $\mu_k(B)$, with the same skew-symmetrizing matrix D.*

Note that seed mutations do not change the frozen variables $\mathbf{c} = \{x_{n+1}, \ldots, x_m\}$.

Example 4.3. Let $\mathbf{x}$ and $\mathbf{c}$ be the sets of variables associated with the diagonals and sides, respectively, of some triangulation of a convex $(n+3)$-gon. (Thus $m = 2n+3$.) Let $\tilde{B}$ be the sign-adjacency matrix of the triangulation. The mutations of seeds $(\tilde{\mathbf{x}}, \tilde{B})$ of this kind correspond to combining the exchange relations (8) with the matrix mutations associated with diagonal flips.

[1] A diligent reader might object that we call $\tilde{\mathbf{x}}$ a set rather than a sequence. This is because we regard any two seeds obtained from each other by simultaneous relabeling of the elements x_i and the matrix entries b_{ij} as identical. That is, for any permutation $w \in \mathcal{S}_m$ such that $w(i) = i$ for $i > n$, we make no distinction between the seeds $(\tilde{\mathbf{x}}, \tilde{B})$ and $(w(\tilde{\mathbf{x}}), w(\tilde{B}))$, where $w(\tilde{\mathbf{x}}) = (x_{w(1)}, \ldots, x_{w(m)})$ and $w(\tilde{B}) = (b_{w(i),w(j)})$.

Seed mutations generate the *mutation equivalence* relation on seeds: $(\tilde{\mathbf{x}}, \tilde{B}) \sim (\tilde{\mathbf{x}}', \tilde{B}')$. Let $\mathcal{S}$ be an equivalence class for this relation. Thus, $\mathcal{S}$ is obtained by repeated mutations of an arbitrary initial seed in all possible directions. This creates an *exchange graph*. See Figure 4.1.

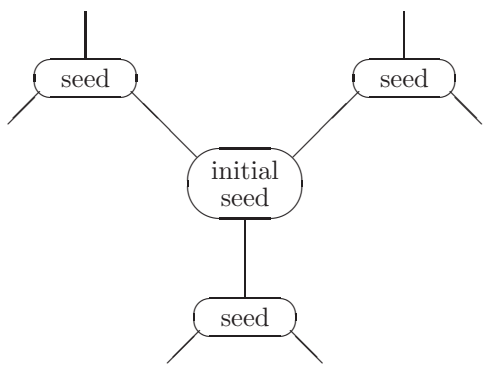


Figure 4.1. Seed mutations and the exchange graph

Let $\mathcal{X} = \mathcal{X}(\mathcal{S})$ be the union of all clusters for all the seeds in $\mathcal{S}$. The elements of $\mathcal{X}$ are called *cluster variables*. See Figure 4.2.

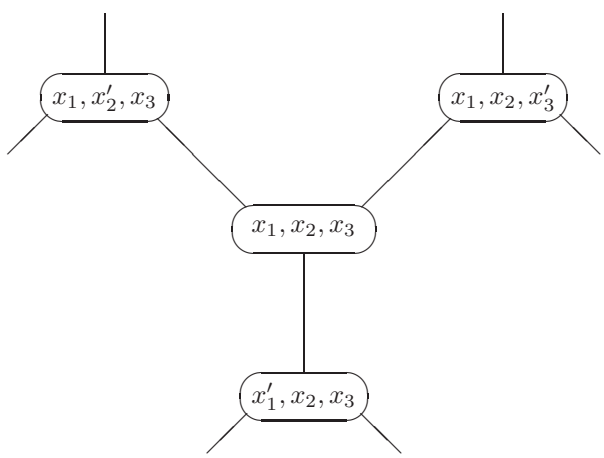


Figure 4.2. Cluster variables

The *cluster algebra*[2] $\mathcal{A} = \mathcal{A}(\mathcal{S})$ associated with $\mathcal{S}$ is generated inside $\mathcal{F}$ by the cluster variables in $\mathcal{X}$ together with the frozen variables $x_{n+1}, \ldots, x_m$ and their inverses. (A variation of this definition includes cluster and frozen variables, but none of their inverses, in the generating set.) The integer n is called the *rank* of $\mathcal{A}$.

Theorem 4.4 (The Laurent phenomenon [**20**])**.** *Any cluster variable is expressed in terms of the variables $x_1, \ldots, x_m$ of any given seed as a Laurent polynomial with integer coefficients.*

[2]Strictly speaking, this is a definition of a *skew-symmetrizable cluster algebra of geometric type*.

Conjecture 4.5 (Nonnegativity conjecture [**20**]). *Every coefficient in these Laurent polynomials is nonnegative.*

Conjecture 4.5 has been proved in a number of special cases, including our main motivating example of an associahedron, to which we return in Example 4.6.

Example 4.6. In the case of Example 4.3, the exchange graph on seeds is precisely the exchange graph on triangulations illustrated in Figures 3.3 and 3.4. The cluster algebra in this example is generated inside the ring of rational functions in $2n+3$ independent variables by the rational functions associated with *all* diagonals and sides of the $(n+3)$-gon. (Cf. Lemma 3.9.) Here we use a variation of the definition of a cluster algebra where the inverses of frozen variables are not included in the set of generators.

This cluster algebra is canonically isomorphic to the homogeneous coordinate ring of the Grassmannian $\mathrm{Gr}(2, n+3)$ with respect to its Plücker embedding. The cluster variables, together with the frozen variables, form the set of all Plücker coordinates on this Grassmannian. Theorem 4.4 and Conjecture 4.5 (proven in this special case) assert that any Plücker coordinate is written in terms of the Plücker coordinates associated with a given triangulation as a Laurent polynomial with nonnegative integer coefficients.

4.2. Finite Type Classification

All results in this section were obtained in [**21**].

A cluster algebra is said to be of *finite type* if it has finitely many distinct seeds. Amazingly, the classification of the cluster algebras of finite type turns out to be completely parallel to the Cartan-Killing classification of (finite crystallographic) root systems. Thus there is a cluster algebra of finite type for each Dynkin diagram, or each Cartan matrix of finite type. We shall now explain how.

For a Cartan matrix $A=(a_{ij})$ of finite type, we define a skew-symmetrizable matrix $B(A)=(b_{ij})$ by

$$b_{ij} = \begin{cases} 0 & \text{if } i=j; \\ a_{ij} & \text{if } i \neq j \text{ and } i \in I_+; \\ -a_{ij} & \text{if } i \neq j \text{ and } i \in I_-, \end{cases}$$

where I_+ and I_- are defined as in Section 2.5. To illustrate, in type B_4, we have (cf. Example 2.3):

$$A = \begin{bmatrix} 2 & -2 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix}, \qquad B(A) = \begin{bmatrix} 0 & -2 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & -1 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

under the convention $I_+ = \{1,3\}$, $I_- = \{2,4\}$.

Theorem 4.7 (Finite type classification). *A cluster algebra $\mathcal{A}$ is of finite type if and only if the exchange matrix at some seed of $\mathcal{A}$ is of the form $B(A)$, where A is a Cartan matrix of finite type.*

The type of A (in the Cartan-Killing nomenclature) is uniquely determined by the cluster algebra $\mathcal{A}$, and is called the "cluster type" of $\mathcal{A}$.

We note that in deciding whether a cluster algebra is of finite type, the bottom part of the matrix $\tilde{B}$ plays no role whatsoever: everything is determined by its principal part B.

In the special cases where a cluster algebra has rank $n = 2$, is of finite type (that is, one of the types A_2, B_2, and G_2), and has no frozen variables (that is, $m = 2$), Theorem 4.7 brings us back to the recurrences of Section 1.1. Indeed, these recurrences are precisely given by the exchange relations in those cluster algebras. The periodicity of the corresponding sequences is simply a reformulation of the "finite type" property for cluster algebras.

Theorem 4.8 (Combinatorial criterion for finite type)**.** *A cluster algebra $\mathcal{A}$ is of finite type if and only if the exchange matrix $B = (b_{ij})$ for any seed of $\mathcal{A}$ satisfies the inequalities $|b_{ij}b_{ji}| \le 3$ for all $i, j \in \{1, \dots, n\}$.*

To rephrase, a mutation equivalence class of skew-symmetrizable $n \times n$ matrices defines a class of cluster algebras of finite type if and only if, for each matrix $B = (b_{ij})$ in this equivalence class, the inequality $|b_{ij}b_{ji}| \le 3$ holds for all i and j.

Combining Theorems 4.8 and 2.4 yields the following completely elementary statement about integer matrices, no direct proof of which is known[3].

Corollary 4.9. *Let $\mathfrak{B}$ be a mutation equivalence class of skew-symmetrizable integer matrices, with the skew-symmetrizing matrix D. (Cf. Lemma 4.2.) The following are equivalent:*

- *any matrix $B = (b_{ij}) \in \mathfrak{B}$ satisfies the inequalities $|b_{ij}b_{ji}| \le 3$, for all i and j;*
- *there exists a matrix $B = (b_{ij}) \in \mathfrak{B}$ with the following property. Define $A = (a_{ij})$ by*
$$a_{ij} = \begin{cases} -|b_{ij}| & \text{if } i \ne j; \\ 2 & \text{if } i = j. \end{cases}$$
Then DAD^{-1} is positive definite.

Let Φ be an irreducible finite root system with Cartan matrix A, and let $\mathcal{A}$ be a cluster algebra of the corresponding cluster type. Theorem 4.7 tells us that the set $\mathcal{X}$ of cluster variables is finite. A more detailed description of this set is provided by Theorem 4.10 below.

Let $\alpha_1, \dots, \alpha_n$ be the simple roots of Φ, and let $\{x_1, \dots, x_n\}$ be the cluster at a seed in $\mathcal{A}$ with the exchange matrix $B(A)$.

Let $\Phi_{\ge -1}$ denote the set of roots in Φ which are either negative simple or positive. Theorem 4.10 shows that the cluster variables in $\mathcal{A}$ are naturally labeled by the roots in $\Phi_{\ge -1}$.

Theorem 4.10. *For any root $\alpha = c_1\alpha_1 + \cdots + c_n\alpha_n \in \Phi_{\ge -1}$, there is a unique cluster variable $x[\alpha]$ given by*

$$x[\alpha] = \frac{P_\alpha(x_1, \dots, x_m)}{x_1^{c_1} \cdots x_n^{c_n}}, \tag{10}$$

where P_α is a polynomial in $x_1, \dots, x_m$ with nonzero constant term. The map $\alpha \mapsto x[\alpha]$ is a bijection between $\Phi_{\ge -1}$ and $\mathcal{X}$.

[3] *Note added in revision.* According to A. Zelevinsky, such a proof has been recently found in his joint work with M. Barot and C. Geiss.

Note that the right-hand side of (10) is a Laurent polynomial, in agreement with Theorem 4.4.

4.3. Cluster Complexes and Generalized Associahedra

This section is based on [**19, 21**], except for the last statement in Theorem 4.11, which was proved in [**13**].

It can be shown that in a given cluster algebra of finite type, each seed is uniquely determined by its cluster. Consequently, the combinatorics of exchanges is encoded by the *cluster complex*, a simplicial complex (indeed, a pseudomanifold) on the set of all cluster variables whose maximal simplices (facets) are the clusters. See Figure 4.3. By Theorem 4.10, the cluster variables—hence the vertices of the cluster complex—can be naturally labeled by the set $\Phi_{\geq -1}$ of "almost positive roots" in the associated root system Φ.

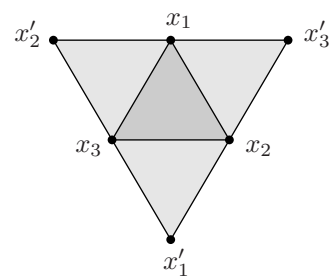


Figure 4.3. The cluster complex

This dual graph of the cluster complex is precisely the exchange graph of the cluster algebra.

Theorem 4.11 below shows that the cluster complex is always spherical, and moreover polytopal.

Recall that $Q_{\mathbb{R}}$ denotes the $\mathbb{R}$-span of Φ. The $\mathbb{Z}$-span of Φ is the *root lattice*, denoted by Q.

Theorem 4.11. *The n roots that label the cluster variables in a given cluster form a $\mathbb{Z}$-basis of the root lattice Q. The cones spanned by such n-tuples of roots form a complete simplicial fan in the ambient real vector space $Q_{\mathbb{R}}$ (the "cluster fan"). This fan is the normal fan*[4] *of a simple n-dimensional convex polytope in the dual space $Q_{\mathbb{R}}^*$.*

This polytope is called the *generalized associahedron* of the corresponding type.

Thus, the cluster complex of a cluster algebra of finite type is canonically isomorphic to the dual simplicial complex of a generalized associahedron of the corresponding type. Conversely, the dual graph of the cluster complex is the 1-skeleton of the generalized associahedron.

[4]Let $P \subset V \cong \mathbb{R}^n$ be an n-dimensional simple convex polytope. The *support function* $F : V^* \to \mathbb{R}$ of P is given by

$$F(\gamma) = \max_{\mathbf{z} \in P} \langle \mathbf{z}, \gamma \rangle.$$

The *normal fan* $\mathcal{N}(P)$ is a complete simplicial fan in the dual space V^* whose full-dimensional cones are the domains of linearity for F. More precisely, each vertex $\mathbf{z}$ of P gives rise to the cone $\{\gamma \in V^* : F(\gamma) = \langle \mathbf{z}, \gamma \rangle\}$ in $\mathcal{N}(P)$.

In type A_n, this construction recovers the n-dimensional associahedron (cf. Figure 4.4). The explanation involves an identification of the roots in $\Phi_{\geq -1}$ with diagonals of a convex $(n+3)$-gon that will be discussed later in Example 4.16. In type B_n, one obtains the n-dimensional cyclohedron. Thus the n-dimensional associahedron (resp., cyclohedron) is dual to the cluster complex of an arbitrary cluster algebra of type A_n (resp., B_n).

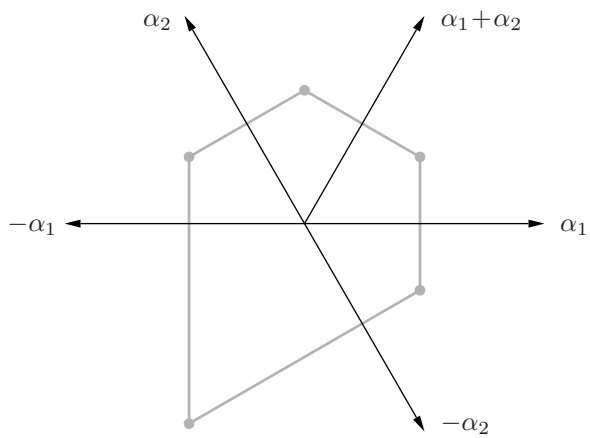


Figure 4.4. Associahedron of type A_2 and its dual fan

Theorem 4.11 leaves the following two questions unanswered:

- Which n-subsets of "almost positive" roots ("root clusters") label the clusters of the cluster algebra of finite type? (An answer to this question would make explicit the combinatorics of a generalized associahedron.)
- What are the inequalities defining a generalized associahedron inside $Q^*_{\mathbb{R}}$? (We already know they are of the form $\langle \mathbf{z}, \alpha \rangle \leq \text{const}$, for $\alpha \in \Phi_{\geq -1}$.)

We are now going to answer these questions, one after another. The answer to the first question is facilitated by the following property of a cluster complex.

Theorem 4.12. *The cluster complex is a clique complex for its* 1*-skeleton. In other words, a subset $S \subset \Phi_{\geq -1}$ is a simplex in the cluster complex if and only if every* 2*-element subset of S is a* 1*-simplex in this complex.*

In type A_n, Theorem 4.12 reflects the basic property of the dual complex of an associahedron: a collection of diagonals forms a simplex if and only if any two of them do not cross.

In order to describe the cluster complex, we therefore need only to clarify which *pairs* of roots label the edges of the cluster complex. Thus, we need to define the root-theoretic analogue of the notion of "non-intersecting diagonals" that lies at the heart of the combinatorial construction of an associahedron.

We will assume from now on that the root system Φ underlying a cluster algebra $\mathcal{A}$ is irreducible. (The general case can be obtained by taking direct products.) We retain the notation of Lecture 2. Thus, n is the rank of Φ (and $\mathcal{A}$); I is the n-element indexing set, which is partitioned into disconnected pieces I_+ and I_-;

W is the corresponding reflection group, generated as a Coxeter group by the generators s_i, for $i \in I$; $w_\circ$ is the element of maximal length in W; $A = (a_{ij})$ is the Cartan matrix; h is the Coxeter number.

Definition 4.13. Define involutions $\tau_\pm : \Phi_{\geq -1} \to \Phi_{\geq -1}$ by

$$\tau_\varepsilon(\alpha) = \begin{cases} \alpha & \text{if } \alpha = -\alpha_i \text{ , for } i \in I_{-\varepsilon}; \\ \prod_{i \in I_\varepsilon} s_i(\alpha) & \text{otherwise.} \end{cases}$$

For example, in type A_2, we get:

$$\begin{array}{ccccccccc} -\alpha_1 & \overset{\tau_+}{\longleftrightarrow} & \alpha_1 & \overset{\tau_-}{\longleftrightarrow} & \alpha_1 + \alpha_2 & \overset{\tau_+}{\longleftrightarrow} & \alpha_2 & \overset{\tau_-}{\longleftrightarrow} & -\alpha_2 \\ \circlearrowright & & & & & & & & \circlearrowright \\ \tau_- & & & & & & & & \tau_+ \end{array}$$

The product $\tau_-\tau_+$ can be viewed as a deformation of the Coxeter element. Hence, what is the counterpart of the Coxeter number?

Theorem 4.14. *The order of $\tau_-\tau_+$ is $(h+2)/2$ if $w_\circ = -1$, and is $h+2$ otherwise. Every $\langle \tau_-, \tau_+ \rangle$-orbit in $\Phi_{\geq -1}$ has a nonempty intersection with $-\Pi$. These intersections are precisely the $\langle -w_\circ \rangle$-orbits in $(-\Pi)$.*

Theorem 4.15. *There is a unique binary relation (called "compatibility") on $\Phi_{\geq -1}$ that has the following two properties:*

- *$\langle \tau_-, \tau_+ \rangle$-invariance: α and β are compatible if and only if $\tau_\varepsilon \alpha$ and $\tau_\varepsilon \beta$ are, for $\varepsilon \in \{+, -\}$;*
- *a negative simple root $-\alpha_i$ is compatible with a root β if and only if the simple root expansion of β does not involve α_i.*

This compatibility relation is symmetric. The clique complex for the compatibility relation is canonically isomorphic to the cluster complex.

In other words (cf. Theorem 4.12), a subset of roots in $\Phi_{\geq -1}$ forms a simplex in the cluster complex if and only if every pair of roots in this subset is compatible.

Example 4.16. In type A_n, the compatibility relation can be described in concrete combinatorial terms using a particular identification of the roots in $\Phi_{\geq -1}$ with the diagonals of a regular $(n+3)$-gon. Under this identification, the roots in $-\Pi$ correspond to the diagonals on the "snake" shown in Figure 4.5. Each positive root $\alpha_i + \alpha_{i+1} + \cdots + \alpha_j$ corresponds to the unique diagonal that crosses precisely the diagonals $-\alpha_i, -\alpha_{i+1}, \ldots, -\alpha_j$ from the snake (see Figure 4.6). It is easy to check that the transformations τ_+ and τ_- act on the set of diagonals as if they were the reflections generating the dihedral group of symmetries of the $(n+3)$-gon. It then follows that two roots are compatible if and only if the corresponding diagonals do not cross each other (at an interior point).

4.4. Polytopal Realizations of Generalized Associahedra

We now demonstrate how to explicitly describe each generalized associahedron by a set of linear inequalities.

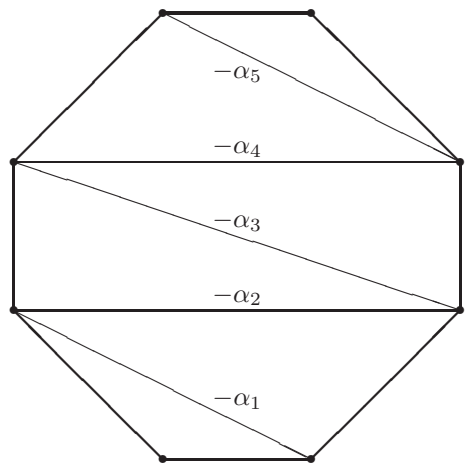


Figure 4.5. The "snake" in type A_5

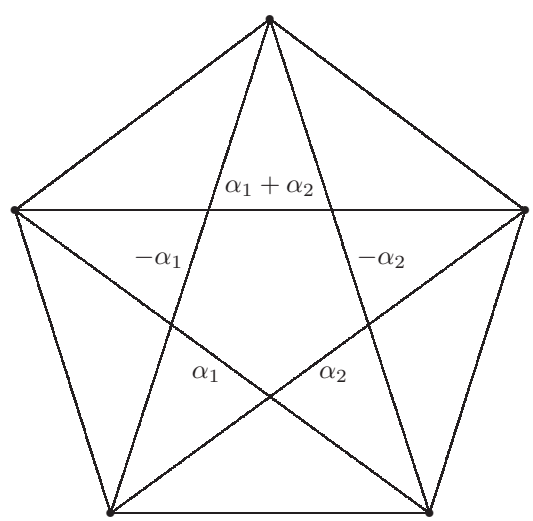


Figure 4.6. Labeling of the diagonals in type A_2

Theorem 4.17. *Suppose that a $(-w_\circ)$-invariant function $F : -\Pi \to \mathbb{R}$ satisfies the inequalities*

$$\sum_{i \in I} a_{ij} F(-\alpha_i) > 0 \quad \textit{for all } j \in I.$$

Let us extend F (uniquely) to a $\langle \tau_-, \tau_+ \rangle$-invariant function on $\Phi_{\geq -1}$. The generalized associahedron is then given in the dual space $Q^*_{\mathbb{R}}$ by the linear inequalities

$$\langle \mathbf{z}, \alpha \rangle \leq F(\alpha)\,, \ \textit{for all } \alpha \in \Phi_{\geq -1}\,.$$

An example of a function F satisfying the conditions in Theorem 4.17 is obtained by setting $F(-\alpha_i)$ equal to the coefficient of the simple coroot $\alpha_i^\vee$ in the

half-sum of all positive coroots. (Coroots are the roots of the "dual" root system; see [**9, 34**].)

Example 4.18. In type A_3, Theorem 4.17 is illustrated in Figure 4.7, which shows a 3-dimensional associahedron given by the inequalities

$$\begin{aligned} \max(-z_1\,,\,-z_3\,,\,z_1\,,\,z_3\,,\,z_1+z_2\,,\,z_2+z_3) &\le 3/2\,, \\ \max(-z_2\,,\,z_2\,,\,z_1+z_2+z_3) &\le 2\,. \end{aligned}$$

Example 4.19. In type C_3, Theorem 4.17 is illustrated in Figure 4.8 that shows a 3-dimensional cyclohedron given by the inequalities

$$\begin{aligned} \max(-z_1\,,\,z_1\,,\,z_1+z_2\,,\,z_2+z_3) &\le 5/2\,, \\ \max(-z_2\,,\,z_2\,,\,z_1+z_2+z_3\,,\,z_1+2z_2+z_3) &\le 4\,, \\ \max(-z_3\,,\,z_3\,,\,2z_2+z_3\,,\,2z_1+2z_2+z_3) &\le 9/2\,. \end{aligned}$$

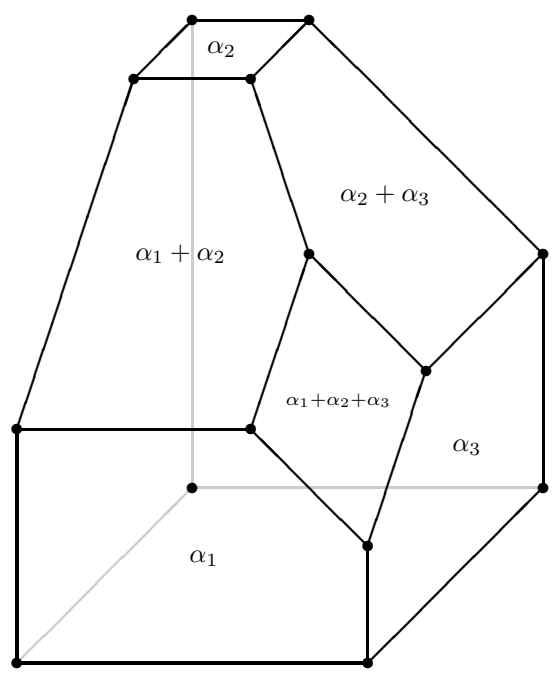


Figure 4.7. Polytopal realization of the type A_3 associahedron

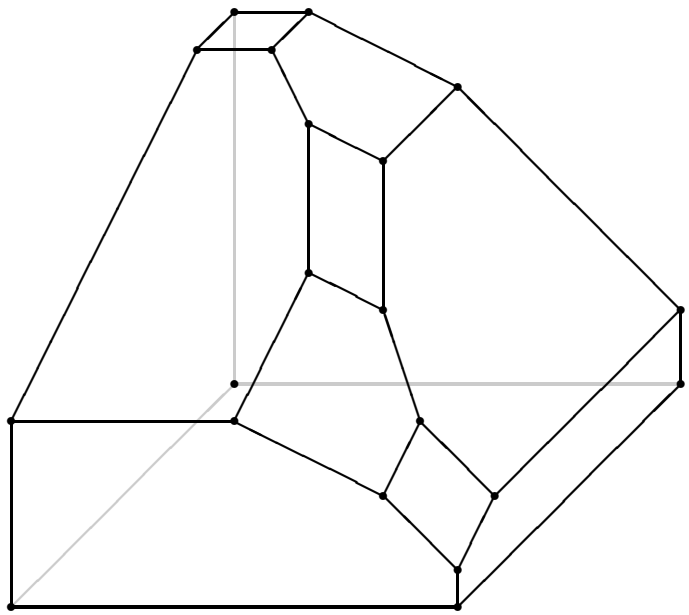


Figure 4.8. Polytopal realization of the type C_3 associahedron (cyclohedron)

4.5. Double Wiring Diagrams and Double Bruhat Cells

The goal of this section is to give a glimpse into how cluster algebras come up in "real life." We will present just one example: the coordinate ring of the open double Bruhat cell in $GL_n(\mathbb{C})$.

We will need the notion of a *double wiring diagram* (of type $(w_\circ, w_\circ)$), which is illustrated in Figure 4.9. Such a diagram consists of two families of n piecewise-straight lines, each family colored with one of two colors. The crucial requirement is that each pair of lines of like color intersect exactly once. The lines in a double wiring diagram are numbered separately within each color, as shown in Figure 4.9.

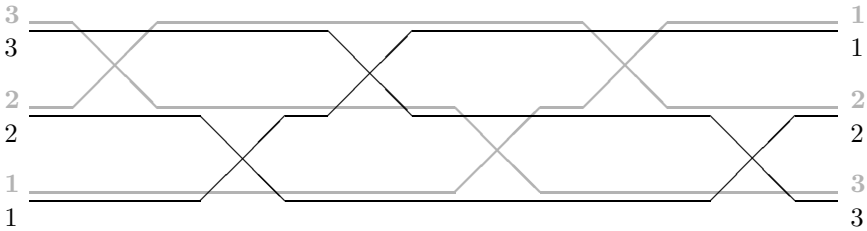


Figure 4.9. Double wiring diagram

We note in passing that double wiring diagrams correspond naturally to shuffles of two reduced words for the element $w_\circ$ in the symmetric group $\mathcal{S}_n$.

From now on, we will not distinguish between double wiring diagrams that are *isotopic*, i.e., have the same "topology." For example, the diagrams in Figures 4.9 and 4.10 are isotopic to each other. The diagram in Figure 4.10 is obtained from

Figure 4.9 by sliding the two leftmost crossings past each other, and also doing the same for the two rightmost crossings.

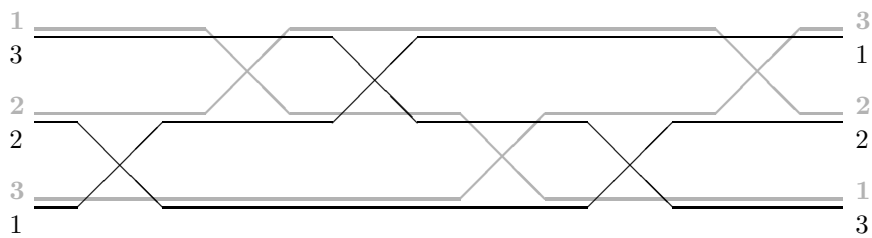


Figure 4.10. An isotopic double wiring diagram

The following lemma is a direct corollary of a theorem of G. Ringel (1956). It can also be obtained from the type A version of a classical result by J. Tits (1969) concerning the word problem in Coxeter groups.

Lemma 4.20. *Any two (isotopy classes of) double wiring diagrams can be transformed into each other by a sequence of local "moves" of three different kinds, shown in Figure 4.11. (Each of these local moves only changes a small portion of a double wiring diagram, leaving the rest of it intact.)*

The reader is asked to ignore, for now, the labels $A, B, \ldots, Z$ in Figure 4.11.

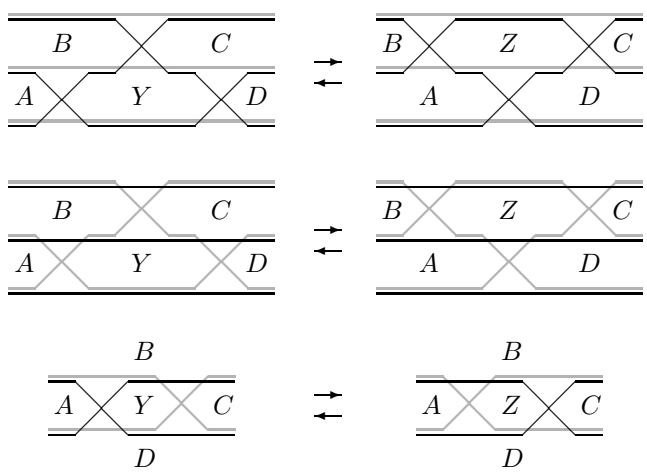


Figure 4.11. Local "moves"

To illustrate Lemma 4.20, the double wiring diagram in Figure 4.9 allows 4 different local moves, all of which are of the kind shown at the bottom of Figure 4.11. Two of these moves can be performed by first passing to the isotopic Figure 4.10. To make each of the other two moves, slide the two innermost crossings in Figure 4.9 past each other; this will create two patterns of the form shown at the bottom of Figure 4.11.

A *chamber* of a double wiring diagram is a connected component of the complement to the union of the lines, with the exception of the "crumbs" made of narrow horizontal isthmuses and small triangular regions; the large component at

the very bottom is not included either. With these conventions, there are exactly n^2 chambers altogether (e.g., 9 chambers in Figure 4.9). We then assign to every chamber a pair of subsets of the set $[1, n] = \{1, \ldots, n\}$: each subset indicates which lines of the corresponding color pass *below* that chamber; see Figure 4.12.

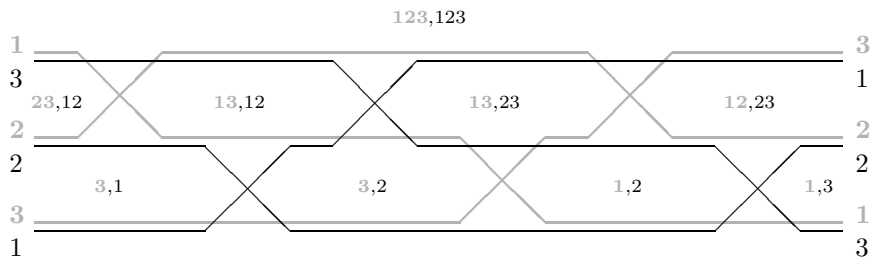


Figure 4.12. Chamber minors

Suppose we are given an $n \times n$ matrix $x = (x_{ij})$. For any subsets $I, J \subset \{1, \ldots, n\}$ of equal cardinality, we denote by $\Delta_{I,J}(x)$ the corresponding minor of x, that is, the determinant of the submatrix of x occupying the rows and columns specified by the sets I and J. Then each chamber of a double wiring diagram is naturally associated with a *chamber minor* $\Delta_{I,J}$ (viewed as a function on the general linear group $GL_n(\mathbb{C})$), where I and J are the sets written into that chamber.

We note that two double wiring diagrams have the same associated collections of chamber minors if and only if they are isotopic.

Let $\mathcal{F}$ denote the field of rational functions on $GL_n(\mathbb{C})$, i.e., the field of rational functions with complex coefficients in the matrix entries x_{ij} (viewed as indeterminates).

Lemma 4.21. *The n^2 chamber minors of an arbitrary double wiring diagram form a set of algebraically independent generators of the field $\mathcal{F}$.*

Notice that each local move in Figure 4.11 exchanges a single chamber minor Y (associated with a *bounded*, or interior, chamber) with another chamber minor Z, and keeps all other chamber minors in place. We can therefore define, by analogy with triangulations, a graph of exchanges whose vertices correspond to (isotopy classes of) double wiring diagrams, and whose edges correspond to the moves in Figure 4.11.

Example 4.22. For $n = 3$, there are 34 non-isotopic double wiring diagrams. The corresponding 34-vertex graph of exchanges can be found in [**23**, Figure 10]. It has 18 vertices of degree 4, and 16 vertices of degree 3. They correspond, respectively, to the double wiring diagrams that allow 4 local moves (as the diagram in Figure 4.12) and those allowing only 3 local moves (as the diagram in Figure 4.13).

Lemma 4.23. *Whenever two double wiring diagrams differ by a single local move of one of the three types shown in Figure 4.11, the chamber minors appearing there satisfy the identity $AC + BD = YZ$.*

Lemmas 4.21 and 4.23 suggest the existence of a cluster algebra structure associated with $n \times n$ matrices. We next present one of several versions of this structure, leaving out most of the technical details. The ambient field for our cluster algebra is the field $\mathcal{F}$ of rational functions on $GL_n(\mathbb{C})$ introduced above. Each double

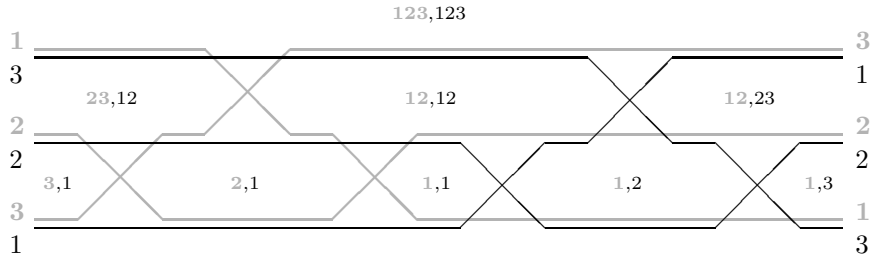


Figure 4.13. A double wiring diagram allowing 3 local moves

wiring diagram provides us with a seed whose cluster variables are the $(n-1)^2$ chamber minors associated with the bounded chambers; the frozen variables are the $2n-1$ chamber minors associated with the unbounded chambers at the edges of the diagram. It remains to define the matrices $\tilde{B}$.

Take any double wiring diagram in which every bounded chamber can be "flipped" (such a diagram can be constructed for any n). Comparing the corresponding exchange relations $AC + BD = YZ$ with (9), determine the matrix entries of $\tilde{B}$. It can be shown that exchanges associated with the local moves on double wiring diagrams are compatible with the cluster algebra axioms. Furthermore, applying these axioms uncovers hitherto hidden clusters which do not correspond to any wiring diagrams. Each variable in these clusters is a regular function on $GL_n(\mathbb{C})$ (a polynomial in the matrix entries). The resulting cluster algebra coincides with the coordinate ring of the *open double Bruhat cell* $G^{w_\circ,w_\circ}$ in $GL_n(\mathbb{C})$. We refer to [**4**] for further details.

Example 4.24. The open double Bruhat cell $G^{w_\circ,w_\circ} \subset GL_3(\mathbb{C})$ consists of all complex 3×3 matrices $x = (x_{ij})$ whose minors

$$x_{13}, \quad \left\| \begin{matrix} x_{12} & x_{13} \\ x_{22} & x_{23} \end{matrix} \right\|, \quad x_{31}, \quad \left\| \begin{matrix} x_{21} & x_{22} \\ x_{31} & x_{32} \end{matrix} \right\|, \quad \det(x) \tag{11}$$

are nonzero. (These 5 minors correspond to the unbounded chambers of any double wiring diagram for $GL_3(\mathbb{C})$.) The coordinate ring $\mathbb{C}[G^{w_\circ,w_\circ}]$ turns out to be a cluster algebra of type D_4 over the ground ring generated by the minors in (11) and their inverses. Thus, the ring of rational functions on GL_3 exhibits some quite unexpected symmetries of type D_4.

This cluster algebra has 16 cluster variables, corresponding to the 16 roots in $\Phi_{\geq -1}$. These variables are:

- 14 (among the 19 total) minors of x, namely, all except those listed in (11);
- two "hidden" variables: $x_{12}x_{21}x_{33} - x_{12}x_{23}x_{31} - x_{13}x_{21}x_{32} + x_{13}x_{22}x_{31}$ and $x_{11}x_{23}x_{32} - x_{12}x_{23}x_{31} - x_{13}x_{21}x_{32} + x_{13}x_{22}x_{31}$.

These 16 variables form 50 clusters of size 4, one for each of the 50 vertices of the type D_4 associahedron.

For any $n \geq 4$, the construction described above produces a cluster algebra of infinite type.

LECTURE 5
Enumerative Problems

5.1. Catalan Combinatorics of Arbitrary Type

Let Φ be a finite irreducible crystallographic root system of rank n, and W the corresponding reflection group. We retain the root-theoretic notation used in Lectures 2 and 4. In particular, $e_1, \ldots, e_n$ are the exponents of Φ, and h is the Coxeter number.

The number of vertices of an n-dimensional associahedron (or, equivalently, the number of clusters in a cluster algebra of type A_n) is the Catalan number $\frac{1}{n+2}\binom{2n+2}{n+1}$. It is natural to ask similar enumerative questions for other Cartan-Killing types.

Theorem 5.1 ([**19**]). *The number of clusters in a cluster algebra of finite type associated with a root system Φ (or, equivalently, the number of vertices of the corresponding generalized associahedron) is equal to*

$$N(\Phi) \stackrel{\text{def}}{=} \prod_{i=1}^{n} \frac{e_i + h + 1}{e_i + 1}\,. \tag{12}$$

Figure 5.1 shows the values of $N(\Phi)$ for all Φ. Recall that the exponents of root systems are tabulated in Figure 2.8.

A_n	B_n, C_n	D_n	E_6	E_7	E_8	F_4	G_2
$\frac{1}{n+2}\binom{2n+2}{n+1}$	$\binom{2n}{n}$	$\frac{3n-2}{n}\binom{2n-2}{n-1}$	833	4160	25080	105	8

Figure 5.1. The numbers $N(\Phi)$

As the numbers $N(\Phi)$ given by (12) can be thought of as generalizations of the Catalan numbers to an arbitrary Cartan-Killing type, it comes as no surprise that they count a host of various combinatorial objects related to the root system Φ. Below in this section, we briefly describe several families of objects counted by $N(\Phi)$. We refer the reader to the introductory sections of [**1, 3, 2, 12, 39**] for the history of research in this area, for further details and references, and for numerous generalizations and connections.

The numbers $N(\Phi)$ seem to have first appeared in D. Djoković's work [**18**] on enumeration of conjugacy classes of elements of finite order in Lie groups.

Antichains in the root poset (non-nesting partitions)

The *root poset* of Φ is the partial order on the set of positive roots Φ_+ such that $\beta \leq \gamma$ if and only if $\gamma - \beta$ is a nonnegative (integer) linear combination of simple roots. See Figures 5.2 and 5.3.

Theorem 5.2 ([**11, 43, 46**]). *The number of antichains (i.e., sets of pairwise non-comparable elements) in the root poset of Φ is equal to $N(\Phi)$.*

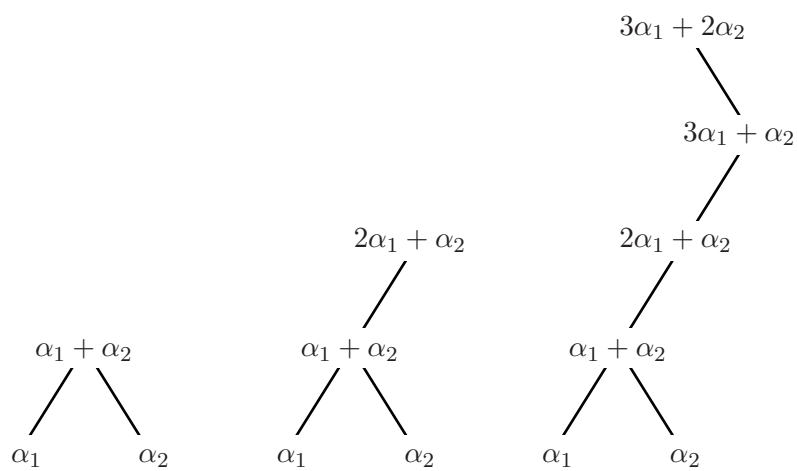


Figure 5.2. The root posets of types A_2, B_2 and G_2.

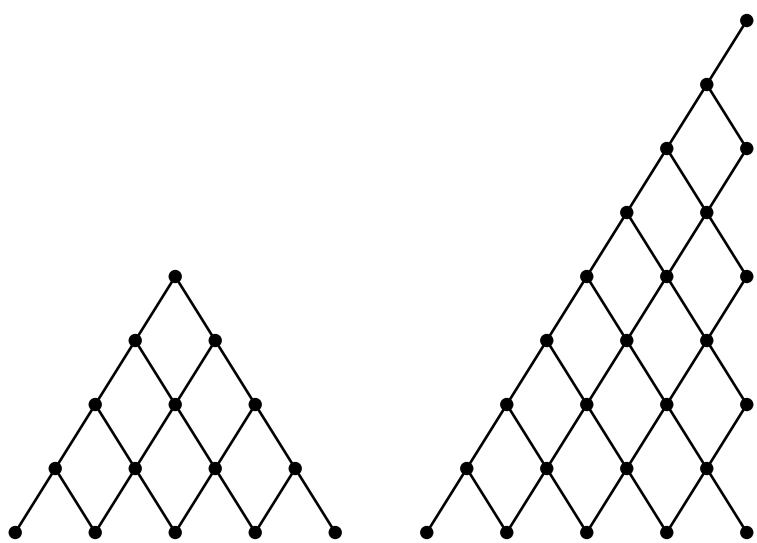


Figure 5.3. The root posets of types A_5 and B_5.

Positive regions of the Shi arrangement

The *Shi arrangement* is the arrangement of affine hyperplanes defined by the equations

$$\begin{array}{l}\langle\beta, x\rangle = 0\\ \langle\beta, x\rangle = 1\end{array} \quad \text{for all } \beta \in \Phi_+.$$

(Thus, the number of hyperplanes in the Shi arrangement is equal to the number of roots in the root system Φ.) The *positive regions* of this arrangement are the regions contained in the *positive cone*, which consists of the points x such that $\langle\beta, x\rangle > 0$ for any $\beta \in \Phi_+$.

Theorem 5.3 ([**46**]). *The number of positive regions in the Shi arrangement is equal to $N(\Phi)$.*

Figure 5.4 shows the Shi arrangements of types A_2, B_2 and G_2, oriented so as to agree with the root systems as drawn in Figure 1.6.

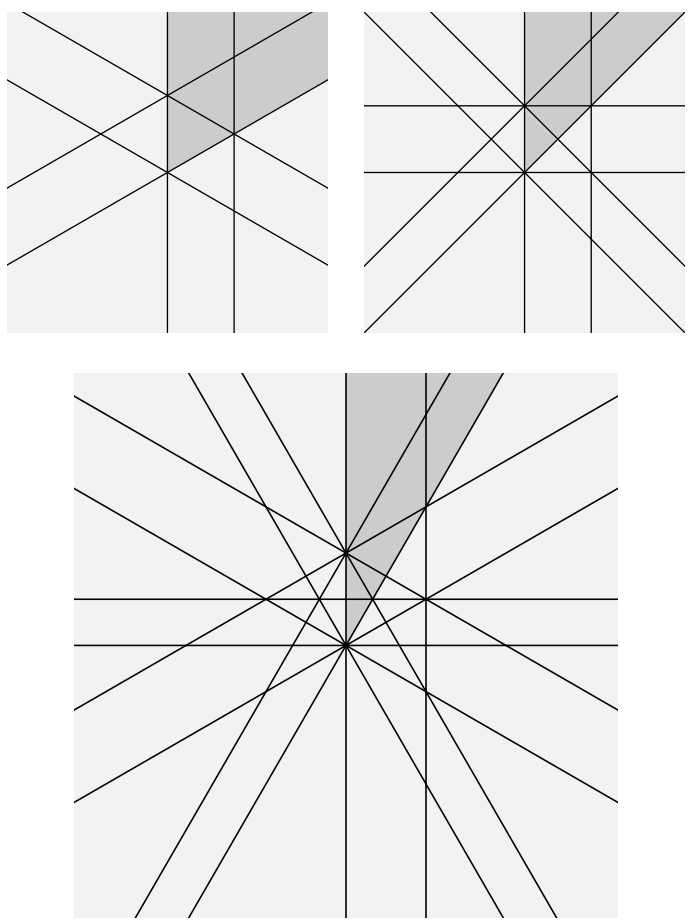


Figure 5.4. The Shi arrangements of types A_2, B_2 and G_2. The positive cone is shaded.

W-orbits in a discrete torus

The reflection group W acts on the root lattice $Q = \mathbb{Z}\Phi$, hence on the "discrete torus" $Q/(h+1)Q$ obtained as a quotient of Q by its subgroup $(h+1)Q$.

Theorem 5.4 ([**32**]). *The number of W-orbits in $Q/(h+1)Q$ is equal to $N(\Phi)$.*

Figures 5.5 and 5.6 illustrate these orbits in types A_2 and B_2, where $h = 3$ and $h = 4$, respectively. Each figure shows the reflection lines of the Coxeter arrangement; the shaded region is a fundamental domain for the translations in $(h+1)Q$.

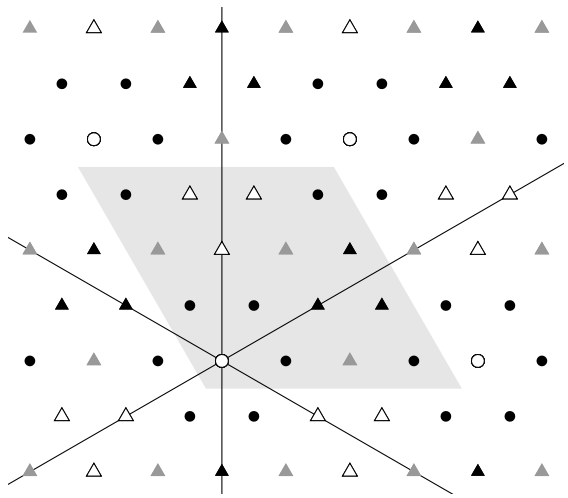


Figure 5.5. A_2-orbits in $Q/4Q$. Each orbit is labeled by a different symbol.

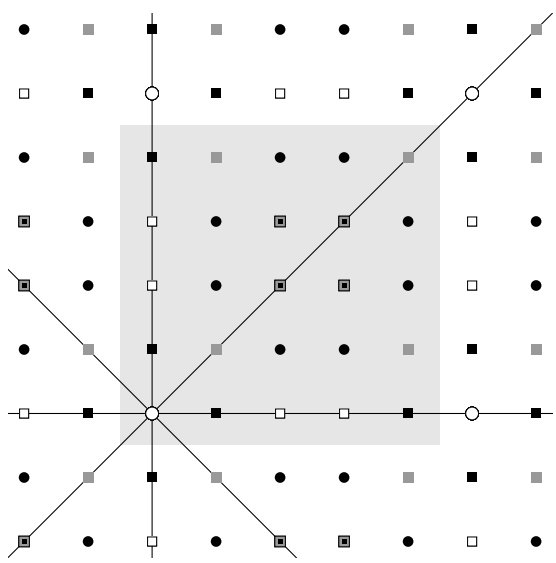


Figure 5.6. B_2-orbits in $Q/5Q$.

Non-crossing partitions

The classical *non-crossing partitions* introduced by Kreweras are (unordered) partitions of the set $[n+1] = \{1, \ldots, n+1\}$ into non-empty subsets called *blocks* which satisfy the following "non-crossing" condition:

- there does not exist an ordered quadruple $(a < b < c < d)$ such that the two-element sets $\{a, c\}$ and $\{b, d\}$ are contained in different blocks.

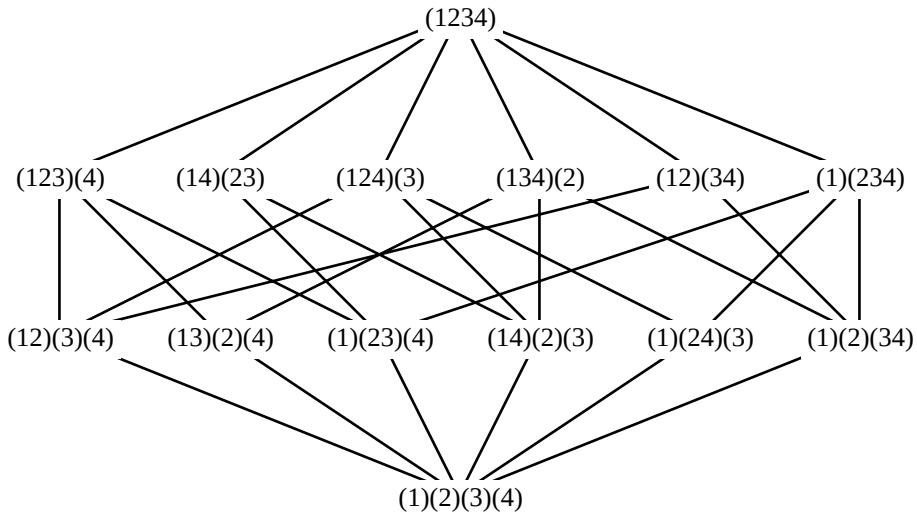


Figure 5.7. The non-crossing partition lattice of type A_3

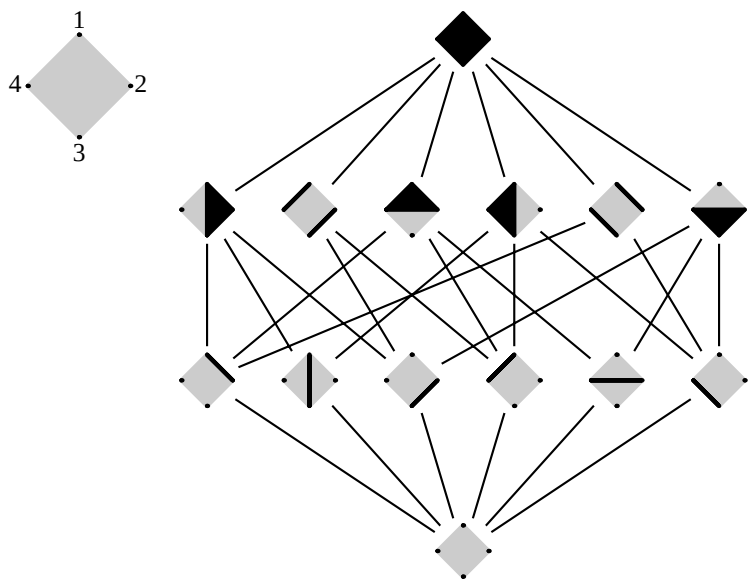


Figure 5.8. Planar representation of non-crossing partitions

Figure 5.7 shows the 14 non-crossing partitions for $n = 3$, partially ordered by refinement. Such partial order is in fact a lattice for any n; the number of non-crossing partitions is a Catalan number.

An alternative way of representing non-crossing partitions is shown in Figure 5.8. Place the elements of $[n + 1]$ around a circle. Then the non-crossing partitions are those set partitions in which the convex hulls of blocks do not intersect.

We will now explain how this construction arises as a type-A special case of a general construction valid for *any* (possibly infinite) Coxeter system (W, S).

A *reflection* in a Coxeter group W is an element conjugate to a generator $s \in S$. Any element $w \in W$ can be written as a product of reflections. Let $L(w)$ denote the length (i.e., number of factors) of a shortest such factorization. We then partially order W by setting $u \preceq uv$ whenever $L(uv) = L(u) + L(v)$, i.e., whenever concatenating shortest factorizations for u and v gives a shortest factorization for uv. Equivalently, w covers u in this partial order if and only if $L(w) = L(u) + 1$ and there is a reflection t such that $w = ut$.

Let c be a product (in an arbitrary order) of the generators in S. Thus, c is a *Coxeter element* in W, in the broader sense of the notion alluded to in a footnote in Section 2.5. The *non-crossing partition lattice* for W (see [**7, 10**]) is the interval $[1, c]$ in the partial order $(W, \preceq)$ defined above. It is a classical result that all Coxeter elements are conjugate to each other. Since the set of all reflections is fixed under conjugation, it follows that different choices of c yield isomorphic posets. (These posets are lattices, which is a non-trivial theorem.)

The following theorem was obtained in [**7, 40**]. A version for the classical types $ABCD$ appeared earlier in [**43**].

Theorem 5.5. *Let W be the reflection group associated with a finite root system Φ. Then the non-crossing partition lattice for W has $N(\Phi)$ elements.*

In type A_n, the general construction presented above recovers the ordinary non-crossing partition lattice. To realize why, look again at Figure 5.7, and interpret each element of the poset as a permutation in $\mathcal{S}_4$ written in cycle notation.

The non-crossing partition lattice of type B_n can also be given a direct combinatorial description. Let us take the ordinary lattice of non-crossing partitions of a $2n$-element set in its representation illustrated in Figure 5.8. Then consider the sublattice consisting of those partitions whose planar representations are centrally symmetric. The result (for $n = 3$) is shown in Figure 5.9.

5.2. Generalized Narayana Numbers

For any enumerative problem whose answer is a Catalan number, replacing a simple count by a generating function with respect to some combinatorial statistic results in a q-analogue of a Catalan number. There are at least three such q-analogues that routinely pop up in various contexts. One is obtained from the usual formula $\frac{1}{n+2}\binom{2n+2}{n+1}$ by replacing $n + 2$ and $\binom{2n+2}{n+1}$ with their standard q-analogues. A different answer is obtained while counting order ideals in the root poset of type A_n by the cardinality of an ideal. For more on these q-analogues, see [**26, 27, 51**].

We will focus on a third q-analogue that is related to the Narayana numbers, defined by the formula $\frac{1}{n+1}\binom{n+1}{k}\binom{n+1}{k+1}$. The Narayana numbers form a triangle shown on the right in Figure 5.10. Thus, the numbers in each row of this triangle

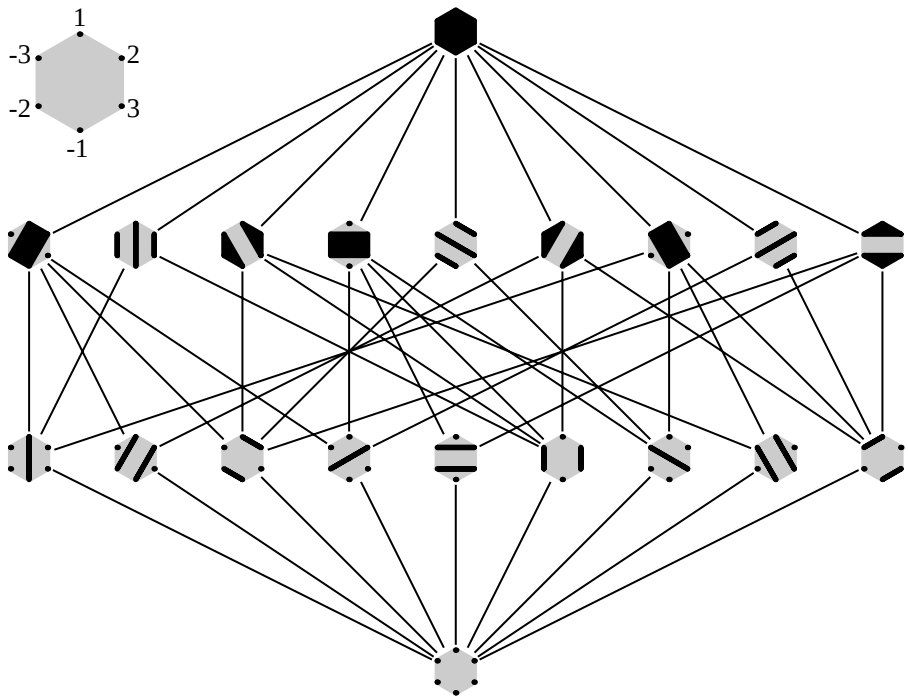


Figure 5.9. The non-crossing partitions of type B_3.

are obtained by looking at the corresponding row of Pascal's triangle on the left, computing products of consecutive pairs of entries, and dividing them by $n+1$.

$$\begin{array}{ccccccccccc}
 & & & & & 1 & & & & & \\
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 2 & & 1 & & & \\
 & & 1 & & 3 & & 3 & & 1 & & \\
 & 1 & & 4 & & 6 & & 4 & & 1 & \\
1 & & 5 & & 10 & & 10 & & 5 & & 1
\end{array}
\qquad\qquad
\begin{array}{ccccccccc}
 & & & & 1 & & & & \\
 & & & 1 & & 1 & & & \\
 & & 1 & & 3 & & 1 & & \\
 & 1 & & 6 & & 6 & & 1 & \\
1 & & 10 & & 20 & & 10 & & 1
\end{array}$$

Figure 5.10. The Pascal triangle and the Narayana numbers

Remarkably, the row sums in the triangle of Narayana numbers are the Catalan numbers:

$$\sum_{k=0}^{n} \frac{1}{n+1}\binom{n+1}{k}\binom{n+1}{k+1} = \frac{1}{n+2}\binom{2n+2}{n+1}.$$

This suggests introducing a q-analogue of the Catalan numbers given by

$$\sum_{k=0}^{n} \frac{1}{n+1}\binom{n+1}{k}\binom{n+1}{k+1} q^k. \tag{13}$$

We will now explain the connection between this q-analogue and the classical (type A) associahedron. This connection will lead us to an extension of the definition to other root systems.

We will need the notions of the f-vector and h-vector of an $(n-1)$-dimensional simplicial complex. The *f-vector* is $(f_{-1}, f_0, \ldots, f_{n-1})$ where f_i denotes the number of i-dimensional faces. The unique "(-1)-dimensional" face is the empty face. The *h-vector* $(h_0, h_1, \ldots, h_n)$ is determined from the f-vector by the "reverse Pascal's triangle" recursion which we illustrate by an example.

Example 5.6. The f-vector of the simplicial complex dual to the associahedron of type A_3 is $(1, 9, 21, 14)$. (See Figure 3.6.) To calculate the h-vector, we place the f-vector and a row of 1's in a triangular array as shown in Figure 5.11 on the left, with most of the entries as yet undetermined. The remaining entries are then filled in by applying the following rule: each entry is the difference between the entry preceding it in its row and the entry directly southwest of it. Thus, we get $9-1=8$, $21-8=13$, *etc.* Finally, we obtain the h-vector $(1, 6, 6, 1)$ by reading the rightmost entries in every row. Notice that these are exactly the Narayana numbers appearing in the third row in Figure 5.10.

$$\begin{array}{ccccccc}
 & & & 14 & & ? & \\
 & & 21 & & ? & & ? \\
 & 9 & & ? & & ? & & ? \\
1 & & 1 & & 1 & & 1 & & 1
\end{array}
\qquad\qquad
\begin{array}{ccccccc}
 & & & 14 & & 1 & \\
 & & 21 & & 13 & & 6 \\
 & 9 & & 8 & & 7 & & 6 \\
1 & & 1 & & 1 & & 1 & & 1
\end{array}$$

Figure 5.11. Computing the h-vector

Lemma 5.7. *The components of the h-vector of the simplicial complex dual to an n-dimensional associahedron are the Narayana numbers $\frac{1}{n+1}\binom{n+1}{k}\binom{n+1}{k+1}$.*

Motivated by Lemma 5.7, we define the (generalized) Narayana numbers $N_k(\Phi)$ $(k = 0, \ldots, n)$ for an arbitrary root system Φ as the entries of the h-vector of the simplicial complex dual to the corresponding generalized associahedron.

Example 5.8. The f-vector of the simplicial complex dual to the 3-dimensional cyclohedron (the associahedron of type B_3) is $(1, 12, 30, 20)$. The corresponding h-vector is $(1, 9, 9, 1)$. In general, the Narayana numbers of type B_n are the squares of entries of Pascal's triangle: $N_k(B_n) = \binom{n}{k}^2$.

It is easy to see that the entries of an h-vector always add up to f_{n-1}, the number of top-dimensional faces in the simplicial complex. Thus, $\sum_k N_k(\Phi) = N(\Phi)$. Consequently, the generating function for the Narayana numbers of type Φ

$$N(\Phi, q) = \textstyle\sum_{k=0}^{n} N_k(\Phi) q^k$$

provides a q-analogue of $N(\Phi)$ which generalizes (13). These generating functions for the finite crystallographic root systems are tabulated in Figure 5.12.

The Narayana numbers provide refined counts for the various interpretations of $N(\Phi)$ given in Section 5.1. These enumerative results are listed in Theorem 5.9 below; we elaborate on the items in the theorem in subsequent comments.

Theorem 5.9 is a combination of results in [**2, 19, 39, 44, 48**]; see [**2**] for a historical overview, and for further generalizations.

$$
\begin{aligned}
N(A_n,q) &= \sum_{k=0}^{n} \frac{1}{n+1}\binom{n+1}{k}\binom{n+1}{k+1} q^k \\
N(B_n,q) &= \sum_{k=0}^{n} \binom{n}{k}^2 q^k \\
N(D_n,q) &= 1+q^n+\sum_{k=1}^{n-1}\left[\binom{n}{k}^2 - \frac{n}{n-1}\binom{n-1}{k-1}\binom{n-1}{k}\right] q^k \\
N(E_6,q) &= 1+36q+204q^2+351q^3+204q^4+36q^5+q^6 \\
N(E_7,q) &= 1+63q+546q^2+1470q^3+1470q^4+546q^5+63q^6+q^7 \\
N(E_8,q) &= 1+120q+1540q^2+6120q^3+9518q^4 \\
&\qquad\qquad +6120q^5+1540q^6+120q^7+q^8 \\
N(F_4,q) &= 1+24q+55q^2+24q^3+q^4 \\
N(G_2,q) &= 1+6q+q^2
\end{aligned}
$$

Figure 5.12. Generating functions for generalized Narayana numbers

Theorem 5.9. *The following numbers are equal to each other, and to $N_k(\Phi)$:*

(i) *the kth component of the h-vector for the dual complex of a generalized associahedron of type Φ;*

(ii) *the number of elements of rank k in the non-crossing partition lattice for W;*

(iii) *the number of antichains of size k in the root poset for Φ;*

(iv) *the number of W-orbits in $Q/(h+1)Q$ consisting of elements whose stabilizer has rank k;*

(v) *the components of the h-vector for the dual cell complex of the positive part of the Shi arrangement.*

Remark 5.10 (Comments on Theorem 5.9)**.**

(i) This was our definition of $N_k(\Phi)$.

(ii) The lattice of non-crossing partitions of type Φ is graded, and $N_k(\Phi)$ is the number of elements of rank k.

(iii) The h-vector of any simplicial polytope satisfies the *Dehn-Sommerville equations* $h_i = h_{d-i}$. Thus interpretation (i) implies that $N_k(\Phi) = N_{n-k}(\Phi)$. This symmetry of the Narayana numbers is also apparent in the interpretation (ii) because the non-crossing partition lattices are self-dual. However, this symmetry is not at all obvious in the interpretations (iii)–(v). In particular, no direct combinatorial explanation is known for why the number of antichains of size k in the root poset is the same as the number of antichains of size $n-k$.

(iv) The stabilizer of an element in $Q/(h+1)Q$ is a reflection subgroup of W. The stabilizers of elements in the same W-orbit are conjugate, and therefore have the same rank. $N_k(\Phi)$ is the number of orbits in which the stabilizers have rank k. For example, in type A_2 there is 1 orbit whose stabilizer has rank 2 (the unfilled circle in Figure 5.5), 3 orbits whose stabilizers have rank 1 (each symbolized by a triangle) and 1 orbit whose stabilizers have rank 0 (the filled circles), in agreement with $N(A_2,q) = 1+3q+q^2$.

(v) The positive regions of the Shi arrangement can be used to define a "dual" cell complex. The vertices of this complex correspond to the positive regions of the Shi arrangement. The faces of the complex correspond to those faces of the closures of these regions that are not contained in the boundary of the positive cone. Accordingly, the maximal faces correspond to the vertices of the arrangement which lie in the interior of the positive cone. See Figure 5.13. Amazingly, this cell complex has the same f-vector (hence the same h-vector) as the corresponding associahedron. In the example of Figure 5.13, we get 5 vertices, 5 faces, and 1 two-dimensional face, matching the numbers for the pentagon (the type A_2 associahedron).

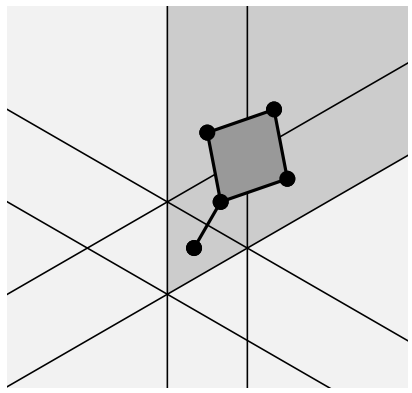


Figure 5.13. The dual complex for the positive part of the Shi arrangement of type A_2.

5.3. Non-crystallographic Types

The construction of generalized associahedra via Definition 4.13 and Theorems 4.14 and 4.15 can be carried out verbatim for the non-crystallographic root systems $I_2(m)$, H_3 and H_4. (However, the last sentence of Theorem 4.15 must be ignored, since no "cluster complex" exists for non-crystallographic root systems.) The associahedron of type $I_2(m)$ is an $(m+2)$-gon. The 1-skeleton of the associahedron for H_3 is shown in Figure 5.14. (The vertex at infinity completes the three unbounded regions to heptagons.)

The analogue of Theorem 5.1 holds true in types H_3, H_4, and $I_2(m)$: the number of vertices of a generalized associahedron is equal to $N(\Phi)$. The latter number is still given by (12), with the exponents taken from Figure 2.8. Figure 5.15 shows these values of $N(\Phi)$ explicitly.

The corresponding h-vectors ("Narayana numbers") are given by

$$\begin{aligned} N(I_2(m),q) &= 1+mq+q^2, \\ N(H_3,q) &= 1+15q+15q^2+q^3, \\ N(H_4,q) &= 1+60q+158q^2+60q^3+q^4. \end{aligned}$$

The construction of the non-crossing partition lattice does not require a crystallographic Coxeter group. Theorem 5.5 and Theorem 5.9(ii) remain valid for the finite non-crystallographic root systems. At present, the other manifestations of $N(\Phi)$ and $N_k(\Phi)$ presented in Sections 5.1 and 5.2 (including Parts (iii)–(v) of Theorem 5.9) do not appear to extend to the non-crystallographic cases.

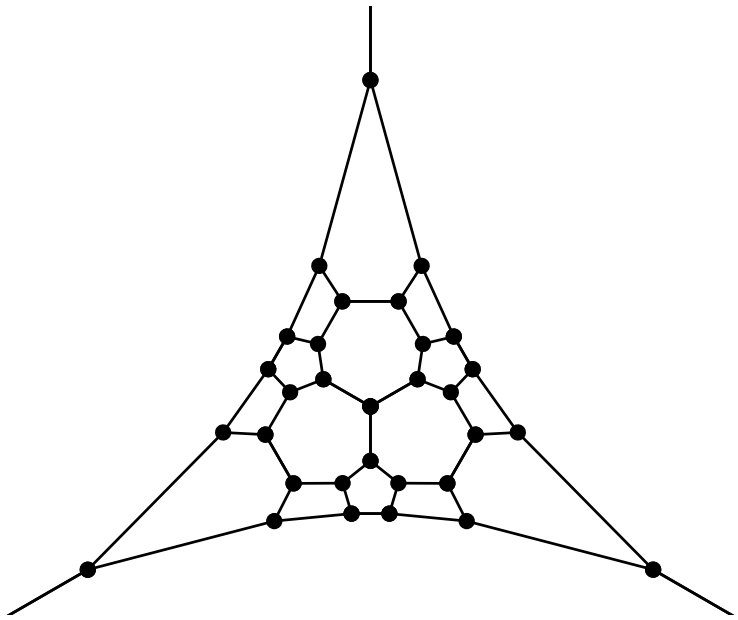


Figure 5.14. The associahedron of type H_3

H_3	H_4	$I_2(m)$
32	280	$m+2$

Figure 5.15. The numbers $N(\Phi)$ in non-crystallographic cases

5.4. Lattice Congruences and the Weak Order

This section is based on [**41**]. Its main goal is to establish a relationship between two fans associated with a root system Φ and the corresponding reflection group W:

- the *Coxeter fan* created by (the regions of) the Coxeter arrangement, and
- the *cluster fan* described in Theorem 4.11.

These fans are the normal fans of a permutahedron and an associahedron of the corresponding type, respectively.

Let ω_i denote the *fundamental weight* [**9**] corresponding to α_i. For $i \in I$, we set

$$\varepsilon(i) = \begin{cases} +1 & \text{if } i \in I_+ \,, \\ -1 & \text{if } i \in I_- \,. \end{cases}$$

Theorem 5.11. *The linear automorphism $Q_{\mathbb{R}} \to Q_{\mathbb{R}}$ defined by $\alpha_i \mapsto \varepsilon(i)\omega_i$ moves the cluster fan to a fan refined by the Coxeter fan.*

The gluing of maximal cones of the Coxeter fan corresponds to contraction of edges in the 1-skeleton of a permutahedron. By Theorem 5.11, this can be done in such a way that the result of the contraction is the 1-skeleton of a generalized associahedron. We have thus extended Theorem 3.6 to all types.

The statement of Theorem 5.11 does not specify which regions of the Coxeter arrangement should be combined together to produce the maximal cones of the transformed cluster fan. We next present a lattice-theoretic construction that, conjecturally, answers this question.

The *weak order* on W is the partial order in which $u \leq v$ if and only if some reduced word for u occurs as an initial segment of a reduced word for v. In particular, v covers u in the weak order if and only if $u^{-1}v$ is a simple reflection, and the length of v is greater than the length of u (necessarily by 1). Lemma 2.13 (see also the paragraph that follows it) implies that the Hasse diagram of the weak order can be identified with the 1-skeleton of a W-permutohedron.

Theorem 5.12 ([**6**]). *The weak order on a finite Coxeter group is a lattice.*

Example 5.13. The weak order of type A_n can be described in the language of permutations of $[n{+}1]$, written in one-line notation. Permutation $v = (v_1, \ldots, v_{n+1})$ covers $u = (u_1, \ldots, u_{n+1})$ if v is obtained from u by exchanging two entries u_i and u_{i+1} with $u_i < u_{i+1}$. Figure 5.16 shows the weak order on A_3. (Cf. Figure 2.3.)

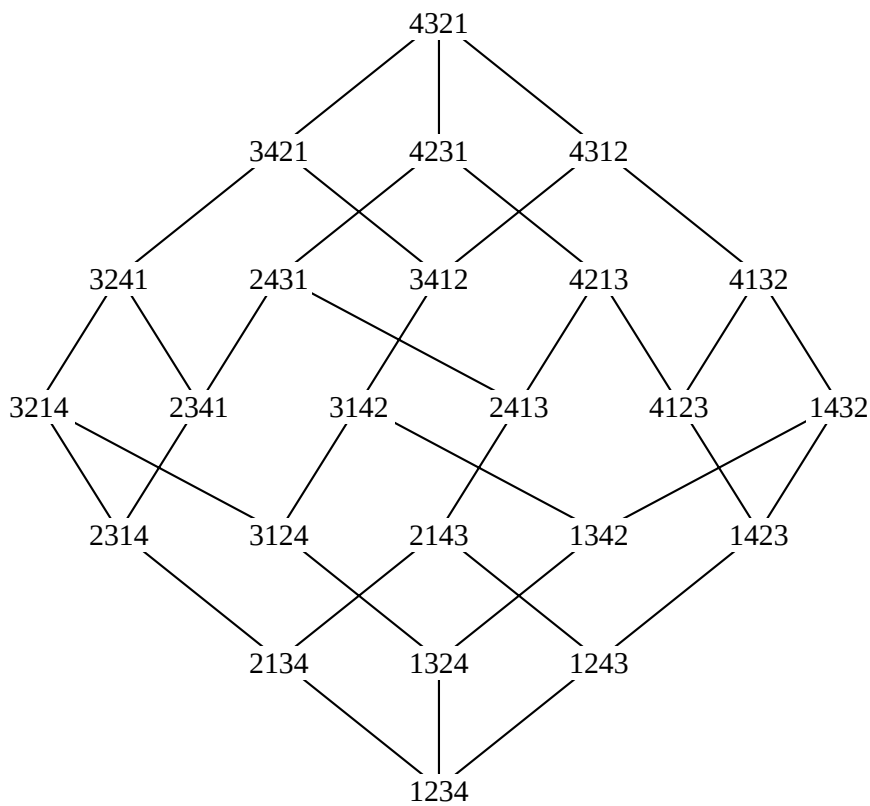


Figure 5.16. The weak order of type A_3

A *congruence* on a lattice is an equivalence relation which respects the meet and join operations. A (bipartite) *Cambrian congruence* on the weak order of W is defined as the (unique) coarsest congruence "$\equiv$" such that, for each edge (s, t) in

the Coxeter diagram, with $t \in I_-$, we have

$$t \equiv tsts \cdots \quad (m_{st} - 1 \text{ factors}).$$

Example 5.14. Figure 5.17 shows the bipartite Cambrian congruence for W of type A_3, i.e., the coarsest congruence on the weak order of the symmetric group $\mathcal{S}_4$ such that $1324 \equiv 3124$ and $1324 \equiv 1342$. The congruence classes are shaded.

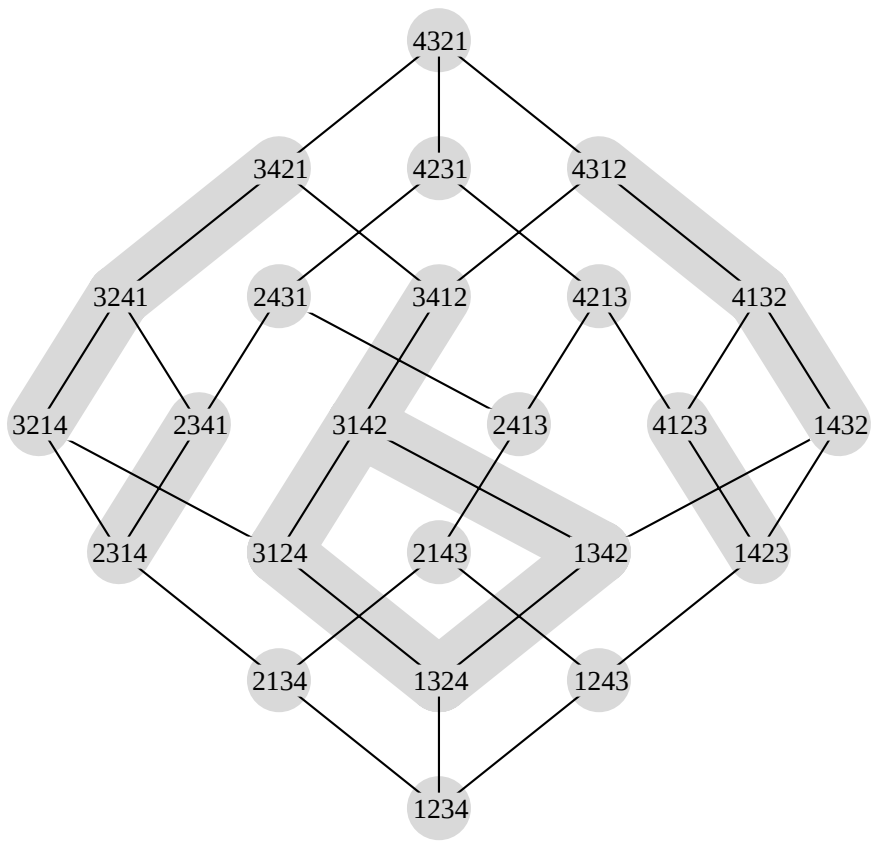


Figure 5.17. A bipartite Cambrian congruence of type A_3

Conjecture 5.15. *Two regions R_u and R_v of the Coxeter arrangement are contained in the same maximal cone of the transformed cluster fan (see Theorem 5.11) if and only if $u \equiv v$ under the bipartite Cambrian congruence.*

Conjecture 5.15 has been proved in types A_n and B_n. The proof makes explicit the combinatorics of the Cambrian congruence and connects it to constructions given by Billera and Sturmfels [**5**] (type A) and Reiner [**42**] (type B). The conjecture implies in particular that the Hasse diagram of the quotient of the weak order by the Cambrian congruence (called the *Cambrian lattice*) is isomorphic to the 1-skeleton of the generalized associahedron.

More concretely, the Cambrian lattice is obtained as the induced subposet of the weak order formed by taking the (unique) smallest element in each (Cambrian) congruence class; see Figure 5.18. We omit the description of the bijection used to

translate the top picture in Figure 5.18 (the Cambrian lattice labeled by permutations) into the bottom one (the associahedron labeled by triangulations).

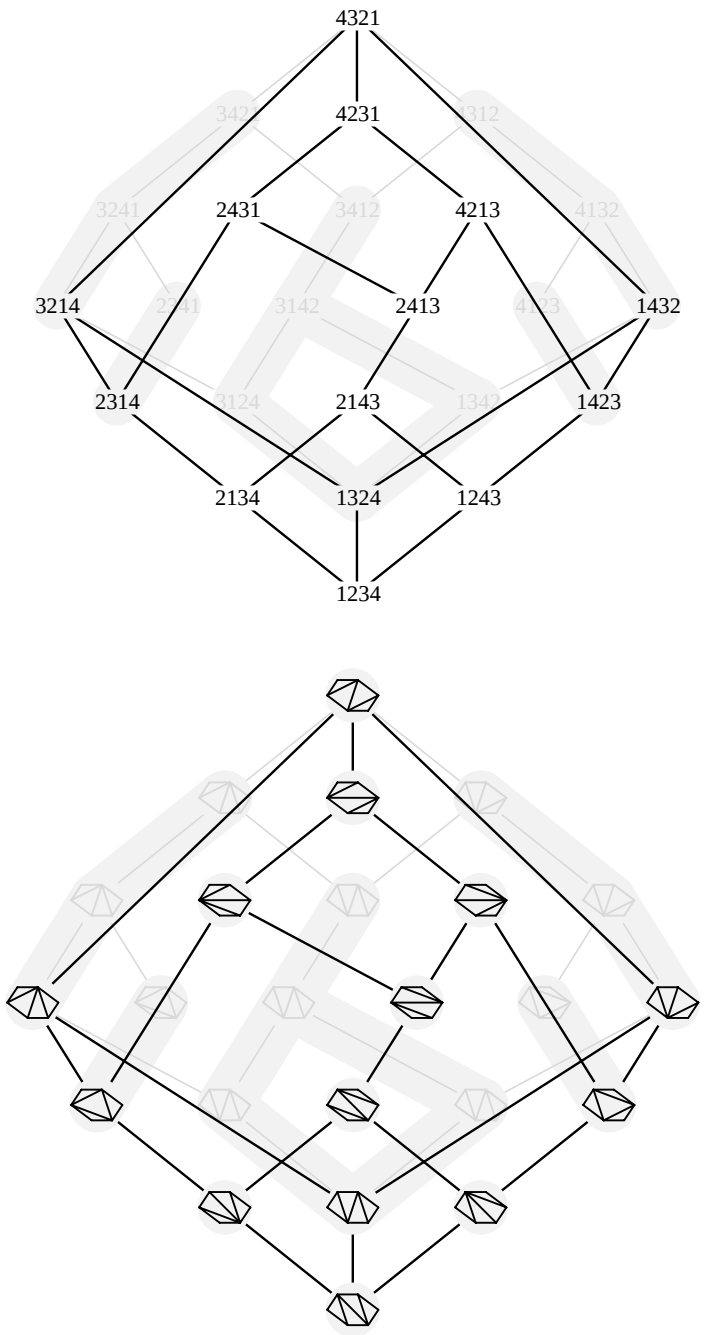


Figure 5.18. A bipartite Cambrian lattice of type A_3

BIBLIOGRAPHY

1. C. A. Athanasiadis, Generalized Catalan numbers, Weyl groups and arrangements of hyperplanes, *Bull. London Math. Soc.* **36** (2004), 294–302.
2. C. A. Athanasiadis, On a refinement of the generalized Catalan numbers for Weyl groups, *Tran. Amer. Math. Soc.* **357** (2005), 179-196.
3. C. A. Athanasiadis and V. Reiner, Noncrossing partitions for the group D_n, *SIAM J. Discrete Math.* **18** (2004), 397–417.
4. A. Berenstein, S. Fomin and A. Zelevinsky, Cluster algebras III: Upper bounds and double Bruhat cells, *Duke Math. J.* **126** (2005), 1-52.
5. L. Billera and B. Sturmfels, Iterated fiber polytopes, *Mathematika* **41** (1994), 348–363.
6. A. Björner, Orderings of Coxeter groups, *Combinatorics and algebra* (Boulder, Colo., 1983), 175–195, *Contemp. Math.* **34**, Amer. Math. Soc., Providence, RI, 1984.
7. D. Bessis, The dual braid monoid, *Ann. Sci. Ecole Norm. Sup.* **36** (2003), 647-683.
8. R. Bott and C. Taubes, On the self-linking of knots. Topology and physics, *J. Math. Phys.* **35** (1994), no. 10, 5247–5287.
9. N. Bourbaki, *Lie groups and Lie algebras, Chapters 4–6*, Springer-Verlag, Berlin, 2002.
10. T. Brady and C. Watt, $K(\pi, 1)$'s for Artin groups of finite type, *Geom. Dedicata* **94** (2002), 225–250.
11. P. Cellini and P. Papi, *ad*-nilpotent ideals of a Borel subalgebra II, *J. Algebra* **258** (2002), 112–121.
12. F. Chapoton, Enumerative properties of generalized associahedra, *Séminaire Lotharingien de Combinatoire*, B51b (2004), 16 pp.
13. F. Chapoton, S. Fomin, and A. Zelevinsky, Polytopal realizations of generalized associahedra, *Canad. Math. Bull.* **45** (2002), 537–566.
14. H. S. M. Coxeter, *Regular polytopes*, Dover Publications, Inc., New York, 1973.
15. M. Csörnyei and M. Laczkovich, Some periodic and non-periodic recursions, *Monatsh. Math.* **132** (2001), 215–236.
16. S. L. Devadoss, Combinatorial equivalence of real moduli spaces, *Notices Amer. Math. Soc.* **51** (2004), no. 6, 620–628.

17. S. L. Devadoss, A space of cyclohedra, *Discrete Comput. Geom.* **29** (2003), 61–75.
18. D. Ž. Djoković, On conjugacy classes of elements of finite order in compact or complex semisimple Lie groups, *Proc. Amer. Math. Soc.* **80** (1980), 181–184.
19. S. Fomin and A. Zelevinsky, Y-systems and generalized associahedra, *Ann. of Math.* **158** (2003), 977–1018.
20. S. Fomin and A. Zelevinsky, Cluster algebras I: Foundations, *J. Amer. Math. Soc.* **15** (2002), 497–529.
21. S. Fomin and A. Zelevinsky, Cluster algebras II: Finite type classification, *Invent. Math.* **154** (2003), 63–121.
22. S. Fomin and A. Zelevinsky, Cluster algebras: Notes for the CDM-03 conference, *CDM 2003: Current Developments in Mathematics*, International Press, 2004, 1–34.
23. S. Fomin and A. Zelevinsky, Total positivity: tests and parametrizations, *Math. Intelligencer* **22** (2000), 23–33.
24. V. Fock and A. Goncharov, Moduli spaces of local systems and higher Teichmuller theory, `math.AG/0311149`.
25. W. Fulton and J. Harris, *Representation theory. A first course*, Springer-Verlag, New York, 1991.
26. J. Fürlinger and J. Hofbauer, q-Catalan numbers. *J. Combin. Theory Ser. A* **40** (1985), 248–264.
27. A. Garsia and M. Haiman, A remarkable q, t-Catalan sequence and q-Lagrange inversion, *J. Algebraic Combin.* **5** (1996), 191–244.
28. M. Geck and G. Malle, Reflection groups. A contribution to the Handbook of Algebra, `math.RT/0311012`.
29. M. Gekhtman, M. Shapiro, and A. Vainshtein, Cluster algebras and Weil-Petersson forms, `math.QA/0309138`.
30. I. Gelfand, M. Kapranov, and A. Zelevinsky, *Discriminants, Resultants and Multidimensional Determinants*, Birkhäuser Boston, 1994.
31. G. Grätzer, *General lattice theory*, 2nd edition, Birkhäuser Verlag, Basel, 1998.
32. M. D. Haiman, Conjectures on the quotient ring by diagonal invariants, *J. Algebraic Combin.* **3** (1994) 17–76.
33. M. Hazewinkel, W. Hesselink, D. Siersma, and F. D. Veldkamp, The ubiquity of Coxeter-Dynkin diagrams (an introduction to the A-D-E problem). *Nieuw Arch. Wisk. (3)* **25** (1977), no. 3, 257–307.
34. J. Humphreys, *Reflection Groups and Coxeter Groups*, Cambridge Univ. Press, 1990.
35. V. Kac, *Infinite dimensional Lie algebras*, 3rd edition, Cambridge University Press, 1990.
36. C. W. Lee, The associahedron and triangulations of the n-gon, *European J. Combin.* **10** (1989), no. 6, 551–560.
37. L. Lewin, *Polylogarithms and associated functions*, North-Holland Publishing Co., New York-Amsterdam, 1981.
38. M. Markl, Simplex, associahedron, and cyclohedron, *Contemp. Math.* **227** (1999), 235–265.
39. D. I. Panyushev, ad-nilpotent ideals of a Borel subalgebra: generators and duality, *J. Algebra* **274** (2004), 822–846.

40. M. Picantin, Explicit presentations for the dual braid monoids, *C. R. Math. Acad. Sci. Paris* **334** (2002), 843–848.
41. N. Reading, Cambrian Lattices, *Adv. Math.*, **205** (2006), no. 2, 313-353.
42. V. Reiner, Equivariant fiber polytopes. *Doc. Math.* **7** (2002), 113–132.
43. V. Reiner, Non-crossing partitions for classical reflection groups, *Discrete Math.* **177** (1997), 195–222.
44. V. Reiner and V. Welker, On the Charney-Davis and Neggers-Stanley Conjectures, *J. Combin. Theory Ser. A* **109** (2005), 247–280.
45. I. Reiten, Dynkin diagrams and the representation theory of algebras, *Notices Amer. Math. Soc.* **44** (1997), no. 5, 546–556.
46. J.-Y. Shi, The number of $\oplus$-sign types, *Quart. J. Math. Oxford* **48** (1997), 93–105.
47. R. Simion, A type-B associahedron, *Adv. in Appl. Math.* **30** (2003), 2–25.
48. E. Sommers, B-stable ideals in the nilradical of a Borel subalgebra, *Canad. Math. Bull.*, to appear.
49. R. P. Stanley, On the number of reduced decompositions of elements of Coxeter groups, *European J. Combin.* **5** (1984), 359–372.
50. R. P. Stanley, *Enumerative Combinatorics, vol.2*, Cambridge University Press, 1999, Exercise 6.19. See also the "Catalan addendum" posted at `http://www-math.mit.edu/~rstan/ec/`.
51. R. P. Stanley, *ibid.*, Exercise 6.34.
52. J. D. Stasheff, Homotopy associativity of H-spaces. I, II, *Trans. Amer. Math. Soc.* **108** (1963), 275–292, 293–312.
53. J. Stasheff, What is ... an operad? *Notices Amer. Math. Soc.* **51** (2004), no. 6, 630–631.
54. A. Tonks, Relating the associahedron and the permutohedron, *in:* Operads: Proceedings of Renaissance Conferences (Hartford, CT/Luminy, 1995), 33–36, *Contemp. Math.* **202**, Amer. Math. Soc., Providence, RI, 1997.
55. E. W. Weisstein, Archimedean Solid, in: *MathWorld–A Wolfram Web Resource*, `http://mathworld.wolfram.com/ArchimedeanSolid.html`.
56. A. Zelevinsky, From Littlewood-Richardson coefficients to cluster algebras in three lectures, *Symmetric Functions 2001: Surveys of Developments and Perspectives,* S. Fomin, Ed., NATO Science Series II: Mathematics, Physics and Chemistry, 74. Kluwer Academic Publishers, Dordrecht, 2002.
57. A. Zelevinsky, Cluster algebras: notes for 2004 IMCC (Chonju, Korea, August 2004), `math.RT/0407414`.
58. G. Ziegler, *Lectures on Polytopes*, Springer-Verlag, 1995.
59. J.-B. Zuber, CFT, BCFT, ADE and all that, *in:* Quantum symmetries in theoretical physics and mathematics (Bariloche, 2000), 233–266, *Contemp. Math.* **294**, Amer. Math. Soc., Providence, RI, 2002.

Topics in Combinatorial Differential Topology and Geometry

Robin Forman

IAS/Park City Mathematics Series
Volume 14, 2004

Topics in Combinatorial Differential Topology and Geometry

Robin Forman

Many questions from a variety of areas of mathematics lead one to the problem of analyzing the topology or the combinatorial geometry of a simplicial complex. We will see a number of examples in these notes. Some very general theories have been developed for the investigation of similar questions for smooth manifolds. Our goal in these lectures is to show that there is much to be gained in the world of combinatorics from borrowing questions, tools, motivation, and even inspiration from the theory of smooth manifolds.

These lectures center on two main topics which illustrate the dramatic impact that ideas from the study of smooth manifolds have had on the study of combinatorial spaces. The first topic has its origins in differential topology, and the second in differential geometry.

One of the most powerful and useful tools in the study of the topology of smooth manifolds is Morse theory. In the first three lectures we present a combinatorial Morse theory that posesses many of the desirable properties of the smooth theory, and which can be usefully applied to the study of very general combinatorial spaces. In the first two lectures we present the basic theory along with numerous examples. In the third lecture, we show that discrete Morse theory is a very natural tool for the study of some questions in complexity theory.

Much of the study of global differential geometry is concerned with the relationship between the geometry of a Riemannian manifold and its topology. One long conjectured, still unproved, relationship is the Hopf conjecture, which states that if a manifold has nonpositive sectional curvature, then the sign of its Euler characteristic depends only on its dimension. (See Lecture 4 for a more precise statement.) In [**15**] Charney and Davis formulated a combinatorial analogue of

[1]The Department of Mathematics, Rice University, Houston, TX, USA 77251.
E-mail address: `forman@rice.edu`.
The author was supported in part by the National Science Foundation. The author would also like to thank Carsten Lange, who served as the TA for these lectures, created most of the figures in these notes, and whose enthusiasm and attention to detail dramatically increased the comprehensibility of the text. The author expresses his gratitude to the organizers of the IAS/PC Summer Institute for their tireless dedication and enthusiasm for all things organizational and mathematical. Their support greatly improved the lectures and these notes.

©2007 American Mathematical Society

this conjecture, and then observed that their conjecture is related to some of the central questions in geometric combinatorics. There has been some fascinating recent work on this subject, which has resulted in some very tantalizing, more general conjectures. In Lectures 4 and 5 we present an introduction to the conjectures of Charney and Davis, discuss some of the known partial results, and survey the most recent developments.

LECTURE 1
Discrete Morse Theory

1. Introduction

There is a very close relationship between the topology of a smooth manifold M and the critical points of a smooth function f on M. For example, if M is compact, then f must achieve a maximum and a minimum. Morse theory is a far-reaching extension of this fact. Milnor's beautiful book [**71**] is the standard reference on this subject. In these notes we present an adaptation of Morse theory that may be applied to any simplicial complex (or more general cell complex). There have been other adaptations of Morse Theory that can be applied to combinatorial spaces. For example, a Morse Theory of piecewise linear functions appears in [**59**] and the very powerful "Stratified Morse Theory" was developed by Goresky and MacPherson [**46**],[**47**]. These theories, especially the latter, have each been successfully applied to prove some very striking results.

We take a slightly different approach than that taken in these references. Rather than choosing a suitable class of continuous functions on our spaces to play the role of Morse functions, we will be working entirely with discrete structures. Hence, we have chosen the name *discrete Morse theory* for the ideas we will describe. Moreover, in these notes, we will describe the theory entirely in terms of the (discrete) gradient vector field, rather than an underlying function. We show that even without introducing any continuity, one can recreate, in the category of combinatorial spaces, a complete theory that captures many of the intricacies of the smooth theory, and can be used as an effective tool for a wide variety of combinatorial and topological problems.

The goal of these lectures is to present an overview of the subject of discrete Morse theory that is sufficient both to understand the major applications of the theory to combinatorics, and to apply the the theory to new problems. We will not be presenting theorems in their most recent or most general form, and simple examples will often take the place of proofs. Those interested in a more complete presentation of the theory can consult the reference [**32**]. Earlier surveys of this work have appeared in [**31**] and [**35**], and earlier, and similar, versions of some of the sections in these notes appeared in [**39**] and [**40**].

2. Cell Complexes and CW Complexes

The main theorems of discrete (and smooth) Morse theory are best stated in the language of CW complexes, so we begin with an overview of the basics of such complexes. J. H. C. Whitehead introduced CW complexes in his foundational work on homotopy theory, and all of the results in this section are due to him. The reader should consult [**68**] for a very complete introduction to this topic. In these notes we will consider only finite CW complexes, so many of the subtleties of the subject will not appear.

The building blocks of cell complexes are cells. Let B^d denote the closed unit ball in d-dimensional Euclidean space. That is, $B^d = \{x \in \mathbb{E}^d \text{ s.t. } |x| \leq 1\}$. The boundary of B^d is the unit $(d-1)$-sphere $S^{(d-1)} = \{x \in \mathbb{E}^d \text{ s.t. } |x| = 1\}$. A *$d$-cell* is a topological space which is homeomorphic to B^d. If σ is d-cell, then we denote by $\dot{\sigma}$ the subset of σ corresponding to $S^{(d-1)} \subset B^d$ under any homeomorphism between B^d and σ. A *cell* is a topological space which is a d-cell for some d.

The basic operation of cell complexes is the notion of *attaching a cell.* Let X be a topological space, σ a d-cell and $f : \dot{\sigma} \to X$ a continuous map. We let $X \cup_f \sigma$ denote the disjoint union of X and σ quotiented out by the equivalence relation that each point $s \in \dot{\sigma}$ is identified with $f(s) \in X$. We refer to this operation by saying that $X \cup_f \sigma$ is the result of attaching the cell σ to X. The map f is called *the attaching map.*

We emphasize that the attaching map must be defined on all of $\dot{\sigma}$. That is, the entire boundary of σ must be "glued" to X. For example, if X is a circle, then Figure 1(i) shows one possible result of attaching a 1-cell to X. Attaching a 1-cell to X cannot lead to the space illustrated in Figure 1(ii) since the entire boundary of the 1-cell has not been "glued" to X.

We are now ready for our main definition. A *finite cell complex* is any topological space X such that there exists a finite nested sequence

$$\emptyset \subset X_0 \subset X_1 \subset \cdots \subset X_n = X \tag{1}$$

such that for each $i = 0, 1, 2, \ldots, n$, X_i is the result of attaching a cell to $X_{(i-1)}$.

Note that this definition requires that X_0 be a 0-cell. If X is a cell complex, we refer to any sequence of spaces as in (1) as a *cell decomposition of X*. Suppose that

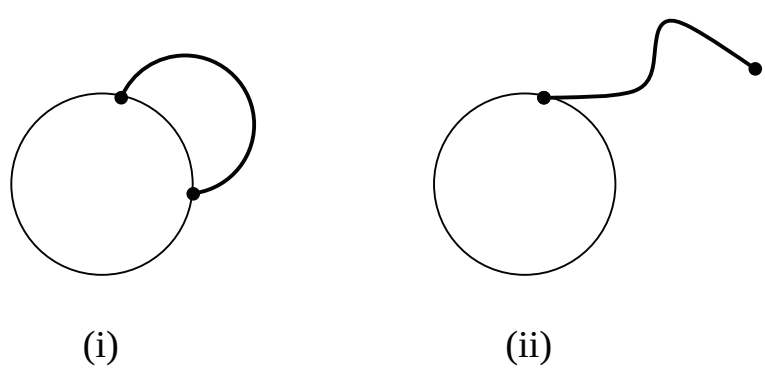


Figure 1. On the left a 1-cell is attached to a circle. This is not true for the picture on the right.

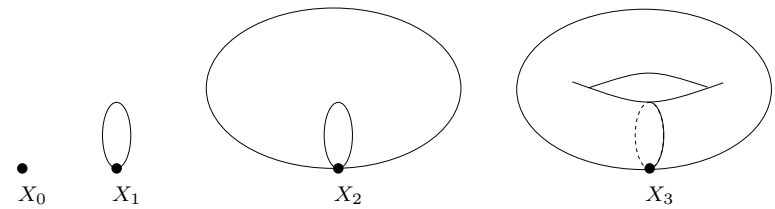


Figure 2. A cell decomposition of the torus.

in the cell decomposition (1), of the $n+1$ cells that are attached, exactly c_d are d-cells. Then we say that the cell complex X has a cell decomposition consisting of c_d d-cells for every d.

We note that a (closed) d-simplex is a d-cell. Thus a finite simplicial complex is a cell complex, and has a cell decomposition in which the cells are precisely the closed simplices.

In Figure 2 we demonstrate a cell decomposition of a 2-dimensional torus which, beginning with the 0-cell, requires attaching two 1-cells and then one 2-cell. Here we can see one of the most compelling reasons for expanding our view from simplicial complexes to more general cell complexes. Every simplicial decomposition of the 2-torus has at least 7 vertices, 21 edges and 14 triangles.

It may seem that quite a bit has been lost in the transition from simplicial complexes to general cell complexes. After all, a simplicial complex is completely described by a finite amount of combinatorial data. On the other hand, the construction of a cell decomposition requires the choice of a number of continuous maps. However, if one is only concerned with the homotopy type of the resulting cell complex, then things begin to look a bit more manageable. Namely, the homotopy type of $X \cup_f \sigma$ depends only on the homotopy type of X and the homotopy class of f.

Theorem 1. *Let $h : X \to X'$ denote a homotopy equivalence, σ a cell, and $f_1 : \dot{\sigma} \to X$, $f_2 : \dot{\sigma} \to X'$ two continuous maps. If $h \circ f_1$ is homotopic to f_2, then $X \cup_{f_1} \sigma$ and $X' \cup_{f_2} \sigma$ are homotopy equivalent.*

(See Theorem 2.3 on page 120 of [**68**].) An important special case is when h is the identity map. We state this case separately for future reference.

Corollary 2. *Let X be a topological space, σ a cell, and $f_1, f_2 : \dot{\sigma} \to X$ two continuous maps. If f_1 and f_2 are homotopic, then $X \cup_{f_1} \sigma$ and $X \cup_{f_2} \sigma$ are homotopy equivalent.*

Therefore, the homotopy type of a cell complex is determined by the homotopy classes of the attaching maps. Since homotopy clases are discrete objects, we have now recaptured a bit of the combinatorial atmosphere that we seemingly lost when generalizing from simplicial complexes to cell complexes.

Let us now present some examples.

1) Suppose X is a topological space which has a cell decomposition consisting of exactly one 0-cell and one d-cell. Then X has a cell decomposition $\emptyset \subset X_0 \subset X_1 = X$. The space X_0 must be the 0-cell, and $X = X_1$ is the result of attaching the d-cell to X_0. Since X_0 consists of a single point, the only possible attaching map is the constant map. Thus X is constructed from taking a closed d-ball and

identifying all of the points on its boundary. One can easily see that this implies that the resulting space is a d-sphere.

2) Suppose X is a topological space which has a cell decomposition consisting of exactly one 0-cell and n d-cells. Then X has a cell decomposition as in (1) such that X_0 is the 0-cell, and for each $i = 1, 2, \ldots, n$ the space X_i is the result of attaching a d-cell to $X_{(i-1)}$. From the previous example, we know that X_1 is a d-sphere. The space X_2 is constructed by attaching a d-cell to X_1. The attaching map is a continuous map from a $(d-1)$-sphere to X_1. Every map of the $(d-1)$-sphere into X_1 is homotopic to a constant map (since $\pi_{(d-1)}(X_1) \cong \pi_{(d-1)}(S^d) \cong 0$). If the attaching map is actually a constant map, then it is easy to see that the space X_2 is the wedge of two d-spheres, denoted by $S^d \vee S^d$. (The wedge of a collection of topological spaces is the space resulting from choosing a point in each space, taking the disjoint union of the spaces, and identifying all of the chosen points.) Since the attaching map must be homotopic to a constant map, Corollary 2 implies that X_2 is homotopy equivalent to a wedge of two d-spheres.

When constructing X_3 by attaching a d-cell to X_2, the relevant information is a map from S^{d-1} to X_2, and the homotopy type of the resulting space is determined by the homotopy class of this map. All such maps are homotopic to a constant map (since $\pi_{d-1}(X_2) \cong \pi_{d-1}(S^d \vee S^d) \cong 0$). Since X_2 is homotopy equivalent to a wedge of two d-spheres, and the attaching map is homotopic to a constant map, it follows from Theorem 1 that X_3 is homotopy equivalent to the space that would result from attaching a d-cell to $S^d \vee S^d$ via a constant map, i.e. X_3 is homotopy equivalent to a wedge of three d-spheres.

Continuing in this fashion, we can see that X must be homotopy equivalent to a wedge of n d-spheres.

The reader should not get the impression that the homotopy type of a cell complex is determined by the number of cells of each dimension. This is true only for very few spaces (and the reader might enjoy coming up with some other examples). The fact that wedges of spheres can, in fact, be identified by this numerical data partly explains why the main theorem of many papers in combinatorial topology is that a certain simplicial complex is homotopy equivalent to a wedge of spheres. Namely such complexes are the easiest to recognize. However, that does not explain why so many simplicial complexes that arise in combinatorics are homotopy equivalent to a wedge of spheres. I have often wondered if perhaps there is some deeper explanation for this.

3) Suppose that X is a cell complex which has a cell decomposition consisting of exactly one 0-cell, one 1-cell and one 2-cell. Let us consider a cell decomposition for X with these cells: $\emptyset \subset X_0 \subset X_1 \subset X_2 = X$. We know that X_0 is the 0-cell. Suppose that X_1 is the result of attaching the 1-cell to X_0. Then X_1 must be a circle, and X_2 arises from attaching a 2-cell to X_1. The attaching map is a map from the boundary of the 2-cell, i.e. a circle, to X_1 which is also a circle. Up to homotopy, such a map is determined by its winding number, which can be taken to be a nonnegative integer. If the winding number is 0, then without altering the homotopy type of X we may assume that the attaching map is a constant map, which yields that $X \sim S^1 \vee S^2$ (where $\sim$ denotes homotopy equivalence). If the winding number is 1 then without altering the homotopy type of X we may assume that the attaching map is a homeomorphism, in which case X is a 2-dimensional disc. If the winding number is 2, then without altering the homotopy type of X

we may assume that the attaching map is a standard degree 2 mapping (i.e. that wraps one circle around the other twice, with no backtracking). The reader should convince him/herself that the result in this case is that X is the 2-dimensional projective space $\mathbb{P}^2$. In fact, each winding number results in a homotopically distinct space. These spaces can be distinguished by their homology, since $H_1(X, \mathbb{Z})$ for the space X resulting from an attaching map with winding number n is isomorphic to $\mathbb{Z}/n\mathbb{Z}$.

It seems that we are not quite done with this example, because we assumed that the 1-cell was attached before the 2-cell, and we must consider the alternative order, in which X_1 is the result of attaching a 2-cell to X_0. In this case, X_1 is a 2-sphere, and $X = X_2$ is the result of attaching a 1-cell to X_1. The attaching map is a map of S^0 into S^2. Since S^2 is connected (i.e. $\pi_0(S^2) = 0$) all such maps are homotopic to a constant map. Taking the attaching map to be a constant map yields that $X = S^1 \vee S^2$. Thus adding the cells in this order merely resulted in fewer possibilities for the homotopy type of X. This is a general phenomenon. Generalizing the argument we just presented, using the fact that $\pi_i(S^d) = 0$ for $i < d$, yields the following statement.

Proposition 3. *Let*

$$\emptyset \subset X_0 \subset X_1 \subset \cdots \subset X_n = X \tag{2}$$

be a cell decomposition of a finite cell complex X. Then X is homotopy equivalent to a finite cell decomposition with precisely the same number of cells of each dimension as in (2), and with the cells attached so that their dimensions form a nondecreasing sequence.

A CW complex is one that can be constructed in this fashion. In fact, even more is required.

Definition 4. A *CW complex* is a cell complex with the property that the boundary of each cell is mapped into the union of the cells of lower dimension.

In some sense, this is a merely technical requirement, as every cell complex is homotopy equivalent to a CW complex. However, there are certain advantages to working with CW complexes, and all of the cell complexes which arise in these notes will be CW complexes.

I first learned of simplicial complexes in a course on algebraic topology. They were introduced as a category of topological spaces for which it was rather easy to define homology and cohomology, i.e. in terms of the simplical chain- and cochain-complexes. One might be concerned that in the transition from simplicial complexes to cell complexes we have lost this ability to easily compute these topological invariants. In fact, much of this computability remains. Let X be a cell complex with a fixed cell decomposition. Suppose that in this decomposition X is constructed from exactly c_d cells of dimension d for each $d = 0, 1, 2, \ldots, n = \dim(K)$, and let $C_d(X, \mathbb{Z})$ denote the space $\mathbb{Z}^{c_d}$ (more precisely, $C_d(X, \mathbb{Z})$ denotes the free abelian group generated by the d-cells of X, each endowed with an orientation). The following is one of the fundamental results in the theory of cell complexes.

Theorem 5. *There are boundary maps $\partial_d : C_d(X, \mathbb{Z}) \to C_{d-1}(X, \mathbb{Z})$, for each d, so that*

$$\partial_{d-1} \circ \partial_d = 0$$

and such that the resulting differential complex

$$0 \longrightarrow C_n(X,\mathbb{Z}) \xrightarrow{\partial_n} \dots \xrightarrow{\partial_1} C_0(X,\mathbb{Z}) \longrightarrow 0$$

calculates the homology of X. That is, if we define

$$H_d(C,\partial) = \frac{\mathrm{Ker}(\partial_d)}{\mathrm{Im}(\partial_{d+1})},$$

then for each d

$$H_d(C,\partial) \cong H_d(X,\mathbb{Z}),$$

where $H_d(X,\mathbb{Z})$ denotes the singular homology of X.

The actual definition of the boundary map ∂ is slightly nontrivial and we will not go into it here (see [**68**, Ch. V, Sec. 2] for the details). In fact, it is here that we see the main distinction between general cell complexes and CW complexes. There may exist multiple choices for the boundary map for a general cell complex, but the boundary map is canonical for a CW complex. At first it may seem that without knowing this boundary map, there is little to be gained from Theorem 5. In fact, much can be learned from just knowing of the existence of such a boundary map. For example, let us choose a coefficient field $\mathbb{F}$, and tensor everything with $\mathbb{F}$ to get a differential complex

$$0 \longrightarrow C_n(X,\mathbb{F}) \xrightarrow{\partial_n} \dots \xrightarrow{\partial_1} C_0(X,\mathbb{F}) \longrightarrow 0$$

which calculates $H_*(X,\mathbb{F})$, where now $C_d(X,\mathbb{F}) \cong \mathbb{F}^{c_d}$.

From basic linear algebra we can deduce the following inequalities.

Theorem 6. *Let X be a cell complex with a fixed cell decomposition with c_d cells of dimension d for each d. Fix a coefficient field $\mathbb{F}$ and let b_* denote the Betti numbers of X with respect to $\mathbb{F}$, i.e. $b_d = \dim(H_d(X,\mathbb{F}))$.*

(i) (The Weak Morse Inequalities) For each d

$$c_d \geq b_d.$$

(ii) Let $\chi(X)$ denote the Euler characteristic of X, i.e.

$$\chi(X) = b_0 - b_1 + b_2 - \dots.$$

Then we also have

$$\chi(X) = c_0 - c_1 + c_2 - \dots.$$

As the name "Weak Morse Inequalities" implies, this theorem can be strengthened. The following inequalities, known as the "Strong Morse Inequalities", also follow from standard linear algebra.

Theorem 7 (The Strong Morse Inequalities)**.** *With all notation as in Theorem 6, for each $d = 0, 1, 2, \dots$*

$$c_d - c_{d-1} + c_{d-2} - \dots + (-1)^d c_0 \geq b_d - b_{d-1} + b_{d-2} - \dots + (-1)^d b_0.$$

As the names imply, Theorem 7 does directly imply Theorem 6, as one can see by comparing Strong Morse Inequalities for consecutive values of d, and using the fact that $b_i = 0$ for i larger than the dimension of K.

We mentioned earlier that a great benefit of passing from simplicial complexes to the more general cell complexes is that one often can use many fewer cells. Let us take another look at this phenomenon in light of the Morse inequalities. Consider the case where X is a two-dimensional torus, so that with respect to any coefficient field $b_0 = 1, b_1 = 2, b_2 = 1$. From the weak Morse inequalities, we have that for any cell decomposition,

$$c_0 \geq b_0 = 1$$

$$c_1 \geq b_1 = 2$$

$$c_2 \geq b_2 = 1.$$

A simplicial decomposition is a special case of a CW decomposition, so these inequalities are satisfied when c_d denotes the number of d-simplices in a fixed simplicial decomposition. However, every simplicial decomposition has at least seven 0-simplices, twenty-one 1-simplices and fourteen 2-simplices, so these inequalities are far from equality. It is generally the case that for a simplicial decomposition these inequalities are very far from optimal, and hence are generally of little interest. On the other hand, earlier we demonstrated a CW decomposition of the two-torus with exactly one 0-cell, two 1-cells and one 2-cell. The inequalities tell us, in particular, that one cannot build the torus using fewer cells.

3. The Morse Theory

In this section we introduce the main topic of the first three lectures, namely discrete Morse theory. Morse theory, in the standard setting of smooth manifolds, is usually described in the language of smooth functions on smooth manifolds (e.g. [**71**]). In practice, though, it is often useful to work with gradient vector fields rather than functions (e.g. [**72**], [**82**]). In the discrete setting, too, one can follow either path. In these notes, we will focus on the notion of a (discrete) gradient vector field. To see how discrete Morse theory can be presented from the function point of view, see [**31**] or [**32**],

Let K be a CW complex. (Most of our examples will be simplicial complexes, but in a few places, even when our object of study is a simplicial complex, it will be convenient to allow more general cell complexes.)

Definition 8. Let β be a $(p+1)$-cell of K, with attaching map $h : S^p \to K_p$, where K_p denotes the union of the cells of dimension $\leq p$.

(i) A cell α is a *face* of β, denoted by $\alpha < \beta$ (or $\beta > \alpha$) if $\beta \neq \alpha \subset \beta$ (where here we are identifying a cell with its image in K).

(ii) A face α of β is said to be *regular* if
   (a) $h^{-1}(\alpha)$ is homeomorphic to a ball, and
   (b) h restricted to $h^{-1}(\alpha)$ is a homeomorphism onto α.

(iii) A *regular CW complex* is a CW complex in which every face is regular. We note that every simplicial complex or polyhedron is a regular CW complex.

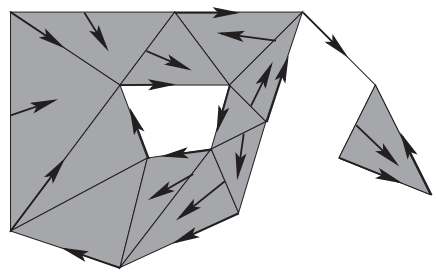


Figure 3. A discrete vector field.

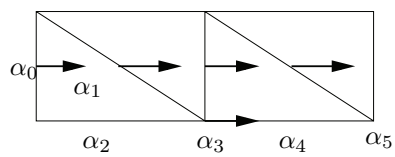


Figure 4. A V-path.

Definition 9. A *discrete vector field* V on K is a collection of pairs $\{\alpha^{(p)} < \beta^{(p+1)}\}$ of cells of K such that each cell is in at most one pair of V, and such that if $\{\alpha^{(p)} < \beta^{(p+1)}\}$ is in V then α is a regular face of β.

We picture such vector fields by drawing, for each pair $\{\alpha^{(p)} < \beta^{(p+1)}\} \in V$, an arrow whose tail lies in α and whose head lies in β (Figure 3). Such pairings were studied in the case of a simplicial complex in [**85**] and [**27**] as a tool for investigating the possible f-vectors for a such complexes. Here we take a different point of view. Our first step is to introduce a special class of vector fields which will play the role of gradient vector fields.

Definition 10.

(1) Given a discrete vector field V on a cell complex K, a *V-path* is a sequence of cells $\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_r$ such that for each $i = 1, 2, \ldots, r$, either $\{\alpha_{i-1} < \alpha_i\} \in V$ or α_i is a codimesion-one face of α_{i-1} and $\{\alpha_i < \alpha_{i-1}\} \notin V$ (Figure 4). We say such a path is a *non-trivial closed path* if $r > 0$ and $\alpha_0 = \alpha_r$.

(2) A discrete vector field V is a *gradient vector field* if there are no non-trivial closed V-paths.

(3) If V is a gradient vector field on a cell complex K and α is a cell of K which is not contained in any pair in V, then we say that α is a *critical cell of V*.

The main theorem of discrete Morse theory is the following.

Theorem 11. *Let K be a CW complex with a discrete gradient field V. Then K is (simple-)homotopy equivalent to a CW complex with precisely one cell of dimension p for each critical cell of V of dimension p.*

Before presenting the very simple proof, we will recall the notion of simple-homotopy. This idea was introduced by J.H.C. Whitehead in an effort to establish a combinatorial basis for homotopy theory. Let K be a CW complex.

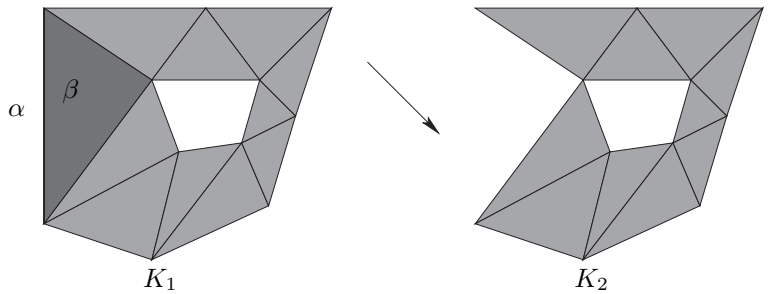


Figure 5. An elementary collapse.

Definition 12. Let β be a $(p+1)$-cell of K, and α a regular face of β. We say that α is a *free face* of β if α is not the face of any other cell of K. (This implies that β is maximal, i.e. is not the face of any cell in K, and that $\dim(\alpha) = p$.)

If α is a free face of β then $K - (int(\alpha) \cup int(\beta))$ is a deformation retract of K. Such a deformation retract is called an *elementary collapse* (and in the category of simplicial complexes, an *elementary simplicial collapse*). See Figure 5. Simple-homotopy is the equivalence relation generated by elementary collapse.

We are now ready to present the proof of Theorem 11. (Many essentially equivalent proofs have appeared since the original proof in [**32**]. Here we present the very short proof that can be found in [**59**].)

Proof. Since V has no closed paths, we can find a cell α of K which has no predecessors, i.e. such that there is no cell β such that β, α is a V-path. There are two possibilities, either (i) α is a maximal face, and is critical for V, or (ii) α is a free face of a cell β, and $\{\alpha < \beta\} \in V$, see Figure 6. In case (i), K is the result of attaching the cell α to $K' = K - int(\alpha)$. In case (ii), K collapses onto $K' = K - (int(\alpha) \cup int(\beta))$. The proof now follows by induction. $\square$

Combining Theorems 11, 6, and 7, and the fact that homotopy equivalent spaces have isomorphic homology, yields the following theorem.

Theorem 13. *Let K be a simplicial complex with a discrete gradient vector field. Let m_p denote the number of critical simplices of dimension p. Let $\mathbb{F}$ be any field, and $b_p = \dim H_p(K, \mathbb{F})$ the pth Betti number with respect to $\mathbb{F}$. Then we have the following relationships.*

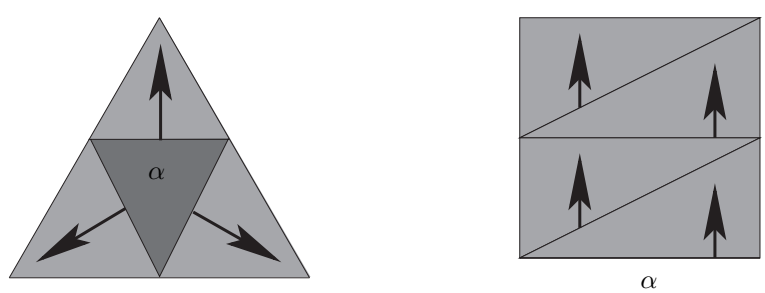


Figure 6. Two possibilities for a cell α with no predecessors.

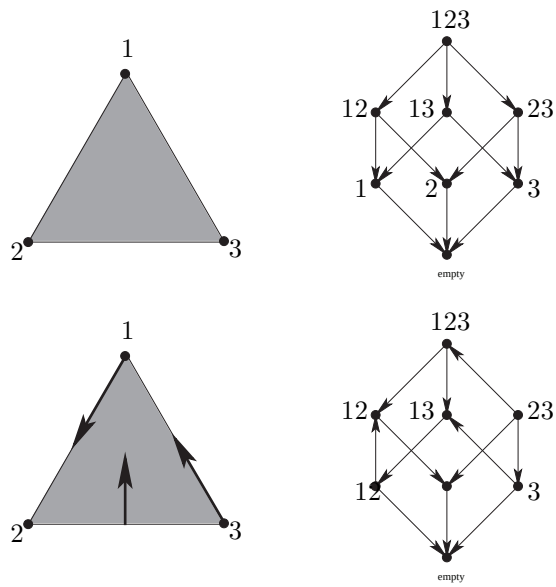


Figure 7. A 2-simplex and its Hasse diagram as a directed graph. A discrete vector field defines a modified Hasse diagram.

I. The Weak Morse Inequalities.

(i) For each $p = 0, 1, 2, \ldots, n$ (where n is the dimension of K)

$$m_p \geq b_p.$$

(ii) $m_0 - m_1 + m_2 - \cdots + (-1)^n m_n = b_0 - b_1 + b_2 - \cdots + (-1)^n b_n \ \ [= \chi(K)].$

II. The Strong Morse Inequalities.

For each $p = 0, 1, 2, \ldots, n, n+1$,

$$m_p - m_{p-1} + \cdots \pm m_0 \geq b_p - b_{p-1} + \cdots \pm b_0.$$

4. A More Combinatorial Language

The notion of a gradient vector field has a very nice purely combinatorial description due to Chari [**14**], with which we can recast the Morse theory in an appealing form. Let K be a regular CW complex. The *Hasse diagram* of K is defined to be the partially ordered set of cells of K ordered by the face relation. Consider the Hasse diagram as a directed graph, directed downward. That is, the vertices of the graph are in 1-1 correspondence with the cells of K, and there is a directed edge from β to α if and only if α is a codimension-one face of β. Now let V be a combinatorial vector field. We modify the directed graph as follows. If $\{\alpha < \beta\} \in V$ then reverse the orientation of the edge between α and β, so that it now goes from α to β. A V-path is precisely a directed path in this modified graph.

Thus, in this combinatorial language, a discrete vector field is a partial matching of the Hasse diagram, and a discrete vector field is a gradient vector field if the partial matching is acyclic, in the sense that the resulting directed graph has no directed loops.

When using this language, there is one possible minor source of confusion. When working with a simplicial complex, one usually includes the empty set as an

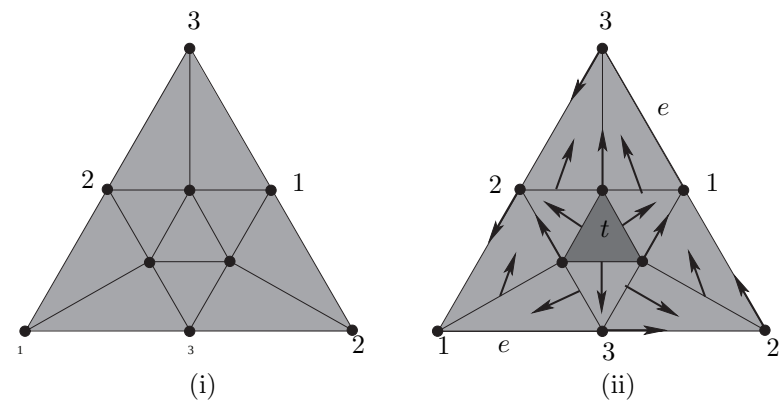


Figure 8. (i) A triangulation of the projective plane. (ii) A discrete vector field on the projective plane.

element of the Hasse diagram (considered as a simplex of dimension -1), while we have not considered the empty set previously. This issue will appear repeatedly in these lectures.

5. Our First Example: The Real Projective Plane

Figure 8(i) shows a triangulation of the real projective plane $\mathbb{P}^2$. The vertices along the boundary with the same labels are to be identified, as are the edges whose endpoints have the same labels. In Figure 8(ii) we illustrate a discrete vector field V on this simplicial complex. One can easily see that there are no closed V-paths (since all V-paths go to the boundary of the figure and there are no closed V-paths on the boundary), and hence is a gradient vector field. The only cells which are neither the head nor the tail of an arrow are the vertex label 1, the edge e, and the triangle t. Thus, by Theorem 11, the projective plane is homotopy equivalent to a CW complex with exactly one 0-cell, one 1-cell and one 2-cell. (Of course, we already knew this from our discussion of Example 3 in Section 2.)

This example gives rise to two potential concerns. The first is that from the main theorem we learn only a statement about "homotopy equivalence". This is sufficient if one is only interested in calculating homology or homotopy groups. However, one might be interested in determining the (PL-)homeomorphism type of the complex. This is possible, in some cases, using deep results of J. H. C. Whitehead. We revisit this topic briefly in the next section.

The second potential point of concern is that as we saw in Section 2 there are an infinite number of different homotopy types of CW complexes which can be built from exactly one 0-cell, one 1-cell and one 2-cell. One might wonder if Morse theory can give us any additional information as to how the cells are attached. In fact, one can deduce much of this information if one has enough information about the gradient paths of the gradient vector field. This point is discussed further in Section 3 of Lecture 2, where we will return to this example of the triangulated projective plane.

6. Sphere Theorems

As mentioned in our discussion at the end of Section 5, one can sometimes use discrete Morse theory to make statements about more than just the homotopy type of the simplicial complex. One can sometimes classify the complex up to homeomorphism or combinatorial equivalence. In this section we give some examples of such arguments. An interesting application of these ideas is presented in the next section. So far, we have not placed any restrictions on the simplicial complexes under consideration. The main idea of this section is that if our simplicial complex has some additional structure, then one may be able to strengthen the conclusion. This idea rests on some very deep work of J. H. C. Whitehead [**95**].

A simplicial complex K is a *combinatorial d-ball* if K and the standard d-simplex σ_d have isomorphic subdivisions. A simplicial complex K is a *combinatorial $(d-1)$-sphere* if K and $\dot{\sigma}_d$ have isomorphic subdivisions (where $\dot{\sigma}_d$ denotes the boundary of σ_d with its induced simplicial structure). A simplicial complex K is a *combinatorial d-manifold with boundary* if the link of every vertex is either a combinatorial $(d-1)$-sphere or a combinatorial $(d-1)$-ball. The following is a special case of the powerful main theorem of [**95**].

Theorem 14. *Let K be a combinatorial d-manifold with boundary which simplicially collapses to a vertex. (That is, K can be a reduced to a vertex by a sequence of elementary simplicial collapses.) Then K is a combinatorial d-ball.*

With this theorem, and its generalizations, one can sometimes strengthen the conclusion of Theorem 11 beyond homotopy equivalence. We present just one example.

Theorem 15. *Let X be a combinatorial d-manifold with a discrete gradient vector field with exactly two critical simplices. Then X is a combinatorial d-sphere.*

The proof is quite simple (given Theorem 14). The statement is trivial for $d = 0$, so we assume that $d \geq 1$. Suppose that X is a combinatorial d-manifold with a discrete gradient vector field V with exactly two critical simplices. Let x_0 be a vertex of X. If x_0 is not critical, then $\{x_0 < e\}$ is an element of V, for some edge e. Let x_1 be the other endpoint of e. Then x_0, e, x_1 is a V-path. If x_1 is not critical, we can follow the V-path to the next vertex x_2, etc. Since there are only a finite number of vertices, and there are no loops, we must eventually reach a critical vertex. We can run this argument in reverse for d-simplices. That is, if α_0 is a d-simplex, and α_0 is not critical, then $\{\beta < \alpha_0\}$ is an element of V for some $(d-1)$-simplex β. Let α_1 denote the other d-simplex incident to β. Then $\alpha_1, \beta, \alpha_0$ is a V-path, and we can follow this path backwards until reaching a critical d-simplex. Thus, there must be precisely one critical vertex x, and one critical d-simplex α. Then $X - \alpha$ is a combinatorial d-manifold with boundary with a discrete gradient vector field with only a single critical simplex, namely the vertex x. It follows that $X - \alpha$ collapses to x. Whitehead's theorem now implies that $X - \alpha$ is a combinatorial d-ball, which implies that X is a combinatorial d-sphere.

7. Our Second Example

In this section we demonstrate some of the ideas of the previous sections with a simple example from algebra. Fix a positive integer n, and consider the following

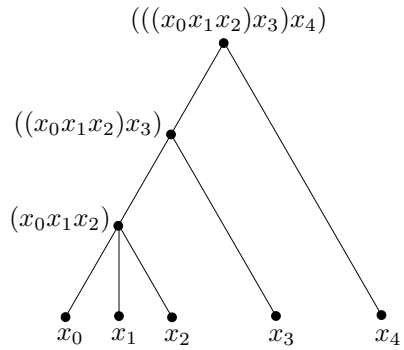


Figure 9. The planar rooted tree corresponding to $(((x_0x_1x_2)x_3)x_4)$.

$(n-2)$-dimensional simplicial complex, which we denote M_n. Starting with the following expression

$$(x_0x_1x_2\dots x_n)$$

consider all ways of adding legal pairs of parentheses. An expression resulting from adding $p+1$ pairs of parentheses will be a p-simplex in our complex. The faces of this p-simplex are all expressions that result from removing corresponding pairs of parentheses.

For example, consider the case $n=3$. The vertices of M_3 are the expressions

$$v_1 = ((x_0x_1)x_2x_3), \qquad v_2 = ((x_0x_1x_2)x_3), \qquad v_3 = (x_0(x_1x_2)x_3),$$
$$v_4 = (x_0(x_1x_2x_3)), \qquad v_5 = (x_0x_1(x_2x_3)),$$

and the edges are the expressions

$$e_1 = (((x_0x_1)x_2)x_3), \qquad e_2 = ((x_0(x_1x_2))x_3), \qquad e_3 = (x_0((x_1x_2)x_3)),$$
$$e_4 = (x_0(x_1(x_2x_3))), \qquad e_5 = ((x_0x_1)(x_2x_3)).$$

One can easily check the relations

$$e_1 = \{v_1, v_2\}, \qquad e_2 = \{v_2, v_3\}, \qquad e_3 = \{v_3, v_4\},$$
$$e_4 = \{v_4, v_5\}, \qquad e_5 = \{v_5, v_1\},$$

so that M_3 is a circle triangulated with 5 edges and 5 vertices.

These complexes arise in a number of different settings. For example, they arise in the study of planar rooted trees. To illustrate by an example, the edge $(((x_0x_1x_2)x_3)x_4)$ of M_4 can naturally be associated with the planar rooted tree shown in Figure 9. From this point of view, the top dimensional simplices correspond to binary trees. (See [**10**] and the references therein for an extensive discussion of such issues.) Moreover, the complexes M_n arise in geometry, as they are closely related to the simplicial complex of subdivisions of an $(n+1)$-gon into subpolygons (see, e.g. [**60**]). In the study of homotopy associative algebras ([**86**], [**87**]) one studies an algebra which is associative only up to homotopy. In that case, M_2, for example, arises from studying all ways of multiplying 3 elements, with $(x_0x_1x_2)$ representing a homotopy between $((x_0x_1)x_2)$ and $(x_0(x_1x_2))$. Note that here we see a slight difference. From this point of view, one would like to think of $((x_0x_1)x_2)$ and $(x_0(x_1x_2))$ as vertices, and $(x_0x_1x_2)$ as an edge between them. Thus, in this context, one is essentially working with the dual of the complex we have defined. We will say more about this a bit later (see the remarks following Theorem 17).

The main goal of this section is to use discrete Morse theory to give a simple proof of the following result.

Theorem 16. *The complex M_n is homotopy equivalent to an $(n-2)$-sphere.*

This result is well known, and it is only our proof that is new. We will prove this theorem by showing that one can easily construct a discrete gradient vector field on M_n which has precisely two critical simplices, namely one critical vertex and one critical $(n-2)$-simplex. The theorem then follows from Theorem 11. In fact, one can deduce more. We saw above that M_3 is not just a homotopy circle, but rather it is an actual combinatorial circle. One can easily see that the link of every vertex of M_n is isomorphic to a complex of the form $M_p * M_{n-p}$ (where $*$ denotes join). By induction, M_p and M_{n-p} are combinatorial spheres of dimension $p-2$ and $n-p-2$, respectively, so the link is a combinatorial sphere of dimension $n-3$ (see Proposition II.1 of [**45**]). Since the link of every vertex of M_n is a combinatorial $(n-3)$-sphere, it follows that M_n is a combinatorial $(n-2)$-manifold (see page 19 of [**45**]). Therefore we can apply Theorem 6 to learn the following stronger result.

Theorem 17. *The complex M_n is a combinatorial $(n-2)$-sphere.*

Before beginning our proof, we return to our earlier comments about the complex arising in the study of homotopy associative algebras. As remarked above, in that case one considers what is essentially the dual of the complex M_n. However, there is a slight modification. Let M_n^* denote a combinatorial $(n-2)$-sphere endowed with the cell decomposition which is dual to that of M_n. In M_n the trivial expression $(x_0x_1\dots x_n)$ corresponds to a simplex of dimension -1, i.e. the empty set. In the dual setting, $(x_0x_1,\dots x_n)$ corresponds to a cell of dimension $n-1$, whose boundary sphere is identified with all of M_n^*. Adding in this cell to form the cone on M_n^* results in a complex, introduced in [**86**] (see also [**87**]) called the associahedron (or Stasheff polytope), and which is often denoted A_{n+1}. Thus we learn

Corollary 18. *The associahedron A_{n+1} is a combinatorial $(n-1)$-ball.*

A proof of this appears in [**86**], by very different methods, and numerous alternative proofs have also been presented. In fact, A_{n+1} is a polytope ([**60**]). For more about the associahedron, from many points of view, one should certainly consult Fomin and Reading's wonderful lecture notes in this volume [**30**].

Let us now describe the construction of the desired gradient vector field V on M_n. Let s be a simplex of M_n. Suppose that there is not a pair of parenthesis around x_0 and x_1. If it is possible to legally add a pair of parentheses around x_0 and x_1 do so and call the resulting simplex t. We then add the pair $\{s \prec t\}$ to V. For example, in M_4 the expression $((x_0x_1x_2)(x_3x_4))$ is paired with $(((x_0x_1)x_2)(x_3x_4))$. After this step, the expressions which have not been paired with any other expression are those that have at least one parenthesis between x_0 and x_1, and it is simple to see that any such parenthesis must be a left parenthesis. There is one additional unpaired expression, namely the expression $s^* = ((x_0x_1)x_2x_3\dots x_n)$. According to our rule, this should be paired with the original expression $(x_0, x_1\dots x_n)$ with no added parentheses, but this is not permitted.

If s is any expression other than s^* that is currently unpaired, and a pair of parentheses can legally be added around the elements x_1 and x_2, do so and call the resulting simplex t. We then add the pair $\{s \prec t\}$ to V. After this step, the expressions which have not been paired with any other expression are s^* and those that have at least one left parenthesis between x_0 and x_1, and at least one left

parenthesis between x_1 and x_2. Pair such an expression with the one resulting from adding a pair of parentheses around x_2 and x_3 if possible. Continue this process as long as possible. When it has terminated, the only expressions that have not been paired up with any other expression are s^* and the one that has a left parenthesis between every consecutive pair x_1 and x_{i+1} for $i = 0, 1, \dots, n-1$, i.e. the expression $t^* = (x_0(x_1(x_2(\dots(x_{n-2}(x_{n-1}x_n))))\dots)$. Note that t^* is an $(n-2)$-simplex of the complex M_n.

This completes our construction of the vector field V. All that needs to be checked is there are no closed V-paths. Denote by V_k the discrete vector field that has been constructed after the kth step in the construction, i.e. after consideration of the pair x_{k-1}, x_k. It is simple to check that V_1 has no closed orbits. Let $s_0^{(p)}, t_0^{(p+1)}, s_1^{(p)}$ denote a V-path. This requires that s_0 and t_0 be paired in V. Suppose that s_0 and t_0 are paired in V_k The reader can check that this implies that either s_1 is the head of an arrow in V_k (and hence the V-path cannot be continued) or s_1 is paired in V_{k-1}. Thus, by induction, there can be no closed V-paths.

8. Exercises for Lecture 1

(1) (a) Prove the strong Morse inequalities. That is, suppose that

$$V: \ 0 \to V_n \xrightarrow{\partial_n} V_{n-1} \xrightarrow{\partial_{n-1}} V_{n-2} \xrightarrow{\partial_{n-2}} \cdots \xrightarrow{\partial_1} V_0 \to 0$$

is a differential complex (i.e. $\partial_{i+1} \circ \partial_i = 0$ for all i). Let m_i denote the dimension of V_i, and b_i the dimension of the ith homology (=$\mathrm{Ker}(\partial_i)/\mathrm{Im}(\partial_{i+1})$). Prove that for each i

$$m_i - m_{i-1} + m_{i-2} - \cdots \pm m_0 \geq b_i - b_{i-1} + b_{i-2} - \cdots \pm b_0.$$

Make sure you see how these inequalities imply the Weak Morse Inequalities.

(b) Now prove the converse of the Morse inequalities. That is, suppose that we are given finite lists of nonnegative integers $m_0, \dots, m_n$, and $b_0, \dots, b_n$ which satisfy the above inequality for each i. Prove that there is a complex V as above with $m_i = \dim(V_i)$ for each i, and such that b_i is the dimension of the ith homology. [This shows that one cannot deduce anything stronger than the strong Morse inequalities using only the abstract existence of a complex which calculates the desired homology.]

(2) Prove that every triangulated disc is collapsible (i.e. collapses to a vertex).

(3) Triangulate a torus (more precisely, construct a simplicial complex which is homeomorphic to the torus) and find a discrete gradient field on the resulting simplicial complex with as few critical simplices as possible.

(4) Prove that every triangulated surface has a perfect gradient vector field. That is, let M be a connected simplicial complex which is homeomorphic to a compact surface. Prove that there is a gradient vector field on M with precisely 1 critical vertex, 1 critical 2-simplex, and g critical edges, where g denotes the genus of M. (Hint: Use the Morse inequalities to see that it is sufficient to find a discrete gradient vector field with exactly one critical vertex, and exactly one critical triangle.)

(5) One can also present discrete Morse theory using the language of functions, rather than gradient vector fields. Let K be a finite simplicial complex.

A function $f : K \to \mathbb{R}$ (i.e. f assigns a single real number to each simplex) is called a *discrete Morse function* if for each p-simplex α

$$\#\{\beta^{(p+1)} > \alpha \ s.t. \ f(\beta) \leq f(\alpha)\} \leq 1$$

and

$$\#\{\gamma^{(p-1)} < \alpha \ s.t. \ f(\gamma) \geq f(\alpha)\} \leq 1.$$

Given such a function f, define a set of pairs V_f by declaring that $\{\alpha < \beta\} \in V_f$ if α is a codimension-one face of β and $f(\beta) \leq f(\alpha)$.

(a) Show that V_f is actually a discrete vector field (i.e. that each simplex is contained in at most one pair in V).
(b) Show that V_f is a gradient vector field.
(c) Show that every gradient vector field arises in this way. That is, if V is a gradient vector field, then there is a discrete Morse function f such that $V = V_f$.

LECTURE 2
Discrete Morse Theory, continued

1. Suspensions and Discrete Morse Theory

Let K be a simplicial complex, and let x and y be two points not in K. Then the *suspension* of K is defined to be the join of K and the set $\{x, y\}$. More geometrically, embed K in some $\mathbb{R}^d$, and embed $\mathbb{R}^d$ in $\mathbb{R}^{d+1}$ by adding a final coordinate. Let x be the point $(0, \ldots, 0, 1)$ and y the point $(0, \ldots, 0, -1)$. Then the suspension of K is the union of all of the closed line segments connecting x to a point in K and all of the closed line segments connecting the point y to a point in K. This space comes with a natural simplicial decomposition induced from that of K.

Let S be a simplex, and M a nonempty proper subcomplex of S. There are two interesting topological spaces to consider in this setting. One is M itself, and the other is S/M, the result of identifying all of the points in M to a single point. While S/M is not a simplicial complex, it does have a canonical cell decomposition giving S/M the structure of a CW complex. Moreover, if $\alpha < \beta$ are two faces of S which are not in M, and α^* and β^* are their images in S/M, then $\alpha^* < \beta^*$, and moreover, α^* is a regular face of β^*.

In fact, the two spaces M and S/M are closely related, and one can deduce essentially the entire topological structure of either one from a knowledge of the other. More precisely, we have the following statement.

Theorem 19. *S/M is homotopy equivalent to the suspension of M.*

Of particular interest to us is the following result.

Corollary 20. *For any p, $\widetilde{H}_{p+1}(S/M, \mathbb{Z}) \cong \widetilde{H}_p(M, \mathbb{Z})$.*

These results are not hard to prove using standard methods, but we present a discrete Morse theory proof of Corollary 20, as the technique (more than the result) will prove useful later (see the next section). In fact, a more careful analysis of this proof allows one to deduce Theorem 19, but we will leave that to the reader. Our approach is to simultaneously construct gradient vector fields U and V on M and S/M, respectively. Let v be any vertex of M. If α is a nonempty simplex of M which does not contain v and which has the property that $v * \alpha$ is also in M, then

pair α with $v * \alpha$. Let U_1 denote this collection of pairs. That is

$$U_1 = \{\{\alpha < v * \alpha\} \text{ s.t. } \emptyset \neq \alpha \text{ and } v * \alpha \subset M\}.$$

It is a simple observation that U_1 is a gradient vector field. Similarly, define a gradient vector field V_1 on S/M by setting

$$V_1 = \{\{\alpha < v * \alpha\} \text{ s.t. } \emptyset \neq \alpha \text{ and } \alpha \nsubseteq M\}.$$

(We are now identifying a simplex α in S, $\alpha \nsubseteq M$, with its image in S/M.) The simplices of M which are critical for U_1 are the vertex v and any nonempty simplex α of M with the property that $v * \alpha \nsubseteq M$. Let C_U denote this collection of critical simplices of U_1. The cells of S/M which are critical for V_1 are the special 0-cell m in S/M resulting from identifying all of the points in M, along with any nonempty simplex $\beta \nsubseteq M$ which has the property that $v \in \beta$, and $\beta - v \subset M$. Let C_V denote this collection of critical simplices of V_1. We observe that there is a canonical identification of the elements of C_U with those of C_V. Namely, identify $v \in M$ with $m \in S/M$, and identify α with $v * \alpha$ whenever $\alpha \subset M$ and $v * \alpha \nsubseteq M$. Let U denote any vector field on M which is an extension of U_1, and let $U_2 = U - U_1$ (so that U_2 consists of pairs of elements in C_U). Define $V_2 = \{\{v * \alpha < v * \beta\}$ *s.t.* $\{\alpha < \beta\} \in U_2\}$, and let $V = V_1 \cup V_2$.

Lemma 21.

(i) Let $A = \alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_k$ be a sequence of elements in C_U, and let $B = v * \alpha_1, v * \alpha_2, v * \alpha_3, \ldots, v * \alpha_k$ be the corresponding sequence of elements in C_V. Then A is a U-path if and only if B is a V-path

(ii) V is a gradient vector field if and only if U is.

Proof. Part (i) follows immediately from the construction of V. To prove part (ii) let $M' = M - C_U$, and $S' = S/M - C_V$. (These are the cells that are paired in U_1 and V_1, respectively.) It is easy to see that any U-path that begins in M' stays in M', and hence is a U_1-path. Since U_1 is a gradient vector field, none of these U-paths are closed. Similarly, any V-path that begins in S' stays in S', and none of these are closed. Hence any closed U-path must lie entirely in C_U. Now the result follows from part (i). □

If U is a vector field on M which contains U_1, we say that U *collapses towards* v. If V is a vector field on S/M which contains V_1, then we say that V *collapses towards* v^*. Then Lemma 21 leads to the following result.

Theorem 22. *For any vertex v of M, there is a canonical identification of gradient vector fields of M which collapse towards v, and those of S/M which collapse towards v^*. If U is a gradient vector field on M which collapses towards v, and V is the corresponding gradient vector field on S/M, then v is critical for U, and m is critical for V. For every additional critical simplex α of U, $v * \alpha$ is critical for V, and every critical cell of V arises in this manner.*

In Section 3 we will introduce the Morse complex, a method of calculating the homology of a cell complex exactly using a knowledge of the critical cells and the gradient paths. The preceeding discussion is sufficient, modulo some minor details which can be supplied by the reader, to deduce that the Morse complex for the relative pair (M, v), which computes the reduced homology of M, is isomorphic to the Morse complex of the relative pair $(S/M, m)$, which computes the reduced

homology of S/M, with the isomporphism shifting all degrees up by 1. This suffices to prove Theorem 20. A more careful consideration of the implications of Theorem 22 yields Theorem 19.

2. Monotone Graph Properties

A number of fascinating simplicial complexes arise from the study of monotone graph properties. Let K_n denote the complete graph on n vertices, and suppose we have label the vertices 1,2,...,n. Let $\mathcal{G}_n$ denote the set of spanning subgraphs of K_n, that is, the subgraphs of K_n that contain all n vertices. (Elements of $\mathcal{G}_n$ are permitted to be disconnected and to have isolated vertices.) A subset $\mathcal{P} \subset \mathcal{G}_n$ is called a *graph property* of graphs with n vertices if inclusion in $\mathcal{P}$ only depends on the isomorphism type of the graph. That is, $\mathcal{P}$ is a graph property if for all pairs of graphs $G_1, G_2 \in \mathcal{G}_n$, if G_1 and G_2 are isomorphic (ignoring the labelings on the vertices) then $G_1 \in \mathcal{P}$ if and only if $G_2 \in \mathcal{P}$. A graph property $\mathcal{P}$ of graphs with n vertices is said to be *monotone decreasing* if for any graphs $G_1 \subset G_2 \in \mathcal{G}_n$, if $G_2 \in \mathcal{P}$ then $G_1 \in \mathcal{P}$.

Monotone decreasing properties abound in the study of graph theory. Here are some typical examples: graphs having no more than k edges (for any fixed k), graphs such that the degree of every vertex is less that δ (for any fixed δ), graphs which are not connected, graphs which are not i-connected (for any fixed i), graphs which do not have a Hamiltonian cycle, graphs which do not contain a minor isomorphic to H (for any fixed graph H), graphs which are r-colorable (for any fixed r), and bipartite graphs.

Any monotone decreasing graph property $\mathcal{P}$ gives rise to a simplicial complex $\mathcal{K}$ where the d-simplices of $\mathcal{K}$ are the graphs $G \in \mathcal{P}$ which have $d+1$ edges. In particular, if G is a d-simplex in $\mathcal{K}$, then the faces of G are all of the nontrivial spanning subgraphs of G (the monotonicity of $\mathcal{P}$ implies that each of these graphs is in $\mathcal{K}$). Said in another way, if $\mathcal{P}$ is nonempty, then the vertices of $\mathcal{K}$ are the edges of K_n (more precisely, the spanning subgraphs of K_n which include all n vertices and precisely one edge), and a collection of vertices in $\mathcal{K}$ span a simplex if the spanning subgraph of K_n consisting of all edges which correspond to these vertices lies in $\mathcal{P}$.

The simplicial complexes induced by many of the above-mentioned monotone decreasing graph properties have been studied using the techniques of these notes. See for example [**14**], [**25**], [**52**], [**53**], [**65**], and [**79**]. These papers contain some beautiful mathematics in which the authors construct "by hand" explicit discrete gradient vector fields, along the way illuminating some of the intricate finer structures of the graph properties.

Some monotone graph properties have recently been the focus of intense interest because of their relation to knot theory. Unfortunately this is probably not a good time for an in depth discussion of this fascinating topic. We will mention only that Vassiliev has shown how one can derive finite type knot invariants from the study of the space of "singular knots" (i.e. maps from S^1 to $\mathbb{R}^3$ which are not embeddings). The homology of the simplicial complexes of disconnected and not-2-connected graphs show up in his spectral sequence calculation of the homology of this space. This is explained in [**93**], where Vassiliev derives the homotopy type of the complex of disconnected graphs. In [**92**] and [**6**], the topology of the space of

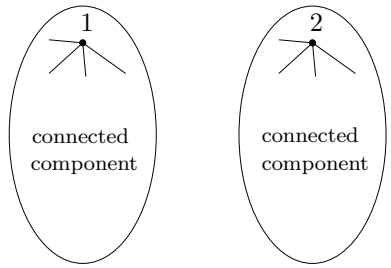


Figure 10. Graphs which are critical for V_{12} have two components.

not-2-connected graphs is determined, with discrete Morse Theory playing a minor role in the latter reference. This topic is reexamined in [**79**], in which the entire investigation is framed in the language of discrete Morse theory. We examine this topic in Section 2.2. Discrete Morse theory is used to determine the topology of not-3-connected graphs in [**52**].

2.1. The Complex of Disconnected Graphs

In this section, we will provide an introduction to this work by taking a look at the simpler case of the complex of disconnected graphs. We will show how the ideas of these lectures may be used to determine the topology of Δ_n, the simplicial compex of disconnected graphs on n vertices. Let me begin by pointing out that this complex can be well studied by more classical methods, and the answer has also been found by Vassiliev in [**93**]. The only novelty of this section is our use of discrete Morse theory.

Our goal is to construct a discrete gradient vector field V on Δ_n, the simplicial complex of all disconnected graphs with the vertex set $\{1, 2, 3, \ldots, n\}$. The construction will be in steps. Let V_{12} denote the discrete vector field consisting of all pairs $\{G, G + (1, 2)\}$, where G is any graph in Δ_n which does not contain the edge $(1, 2)$ and such that $G + (1, 2) \in \Delta_n$. Another way of describing V_{12} is that if G is any graph in Δ_n which contains the edge $(1, 2)$, then $G - (1, 2)$ and G are paired in V_{12}. Actually, there is one exception to this rule. Let g denote the graph consisting of only the single edge $(1, 2)$. Then $g - (1, 2)$ is the empty graph, which corresponds to the empty simplex in Δ_n, and may not be paired in a discrete vector field. Thus, g is unpaired in V_{12}.

The graphs in Δ_n other than g which are unpaired in V_{12} are those that do not contain the edge $(1, 2)$ and have the property that $G + (1, 2) \notin \Delta_n$. That is, those disconnected graphs G with the property that $G + (1, 2)$ is connected. Such a graph must have exactly two connected components, one of which contains the vertex labeled 1, and one which contains the vertex labeled 2. We denote these connected components by G_1 and G_2, resp. See Figure 10.

Let G be a graph other than g which is unpaired in V_{12}, and consider vertex 3. This vertex must either be in G_1 or G_2. Suppose that vertex 3 is in G_1. If G does not contain the edge $(1, 3)$ then $G + (1, 3)$ is also unpaired in V_{12}, so we can pair G with $G + (1, 3)$. If vertex 3 is in G_1, then the graph G is still unpaired if and only if G contains the edge (1,3) and $G - (1, 3)$ is the union of three connected components, one containing vertex 1, one containing vertex 2, and one containing vertex 3.

Similarly, if vertex 3 is in G_2 and G does not contain the edge $(2,3)$, then pair G with $G+(2,3)$. Let V_3 denote the resulting discrete vector field.

The unpaired graphs in V_3 are g and those that either contain the edge (1,3) and have the property that $G-(1,3)$ is the union of three connected components, one containing vertex 1, one containing vertex 2, and one containing vertex 3, or contain the edge (2,3) and have the property that $G-(2,3)$ is the union of three connected components, one containing vertex 1, one containing vertex 2, and one containing vertex 3. We illustrate these graphs in Figure 11. The circles in this figure indicate connected subgraphs.

Now consider the location of the vertex label 4, and pair any graph G which is unpaired in V_3 with $G+(1,4)$, $G+(2,4)$, or $G+(3,4)$ if possible (at most one of these graphs is unpaired in V_3). Call the resulting discrete vector field V_4. We continue in this fashion, considering in turn the vertices label $5,6,\ldots,n$. Let V_i denote the discrete vector field that has been constructed after the consideration of vertex i, and $V=V_n$ the final discrete vector field. When we are done the only unpaired graphs in V will be g and those graphs that are the union of two connected trees, one containing the vertex 1 and one containing the vertex 2. In addition, both trees have the property that the vertex labels are increasing along every ray starting from the vertex 1 or the vertex 2. There are precisely $(n-1)!$ such graphs, and they each have $n-2$ edges, and hence correspond to an $(n-3)$-simplex in Δ_n.

It remains to see that the discrete vector field V is a gradient vector field, i.e. that there are no closed V-paths. We first check that V_{12} is a gradient vector field. Let $\gamma=\alpha_0^{(p)},\beta_0^{(p+1)},\alpha_1^{(p)}$ denote a V_{12}-path. Then α_0 must be the "tail of an arrow", i.e. the smaller graph of some pair in V_{12}, with β_0 being the head of the arrow, i.e. $\beta_0=\alpha_0+(1,2)$. The simplex α_1 is a codimension-one face of β_0 other than α_0. Thus, α_1 corresponds to a graph of the form $\alpha_0+(1,2)-e$, where e is an edge of α_0 other than (1,2). Since α_1 contains the edge $(1,2)$ it is the "head of an arrow" in V_{12}, i.e. the larger graph of some pair in V_{12}, which implies that γ cannot be continued to a longer V_{12}-path. This certainly implies that there are no closed V_{12}-paths.

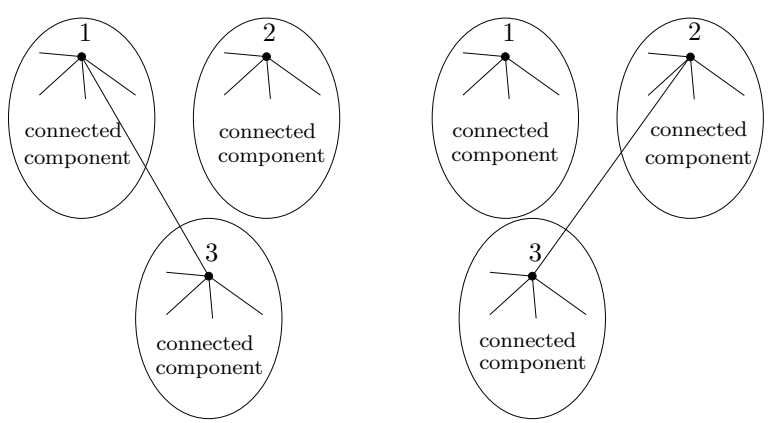


Figure 11. The two types of graphs which are critical for V_3.

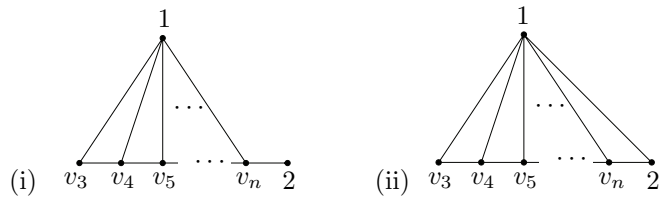


Figure 12. (i) Critical Graphs in Δ_n^2. (ii) Critical Graphs in N_n^2.

The same sort of argument will work for V. Recall that V is constructed in stages, by first considering the edge $(1,2)$ and then the vertices $3,4,5,\ldots$ in order. Let $\gamma = \alpha_0, \beta_0, \alpha_1$ denote a V-path. In particular, α_0 and β_0 must be paired in V. The reader can check that if α_0 and β_0 are first paired in V_i, $i \geq 3$, then either α_1 is the head of an arrow in V_i, in which case the V-path cannot be continued, or α_1 is paired in V_{i-1}. It follows by induction that there can be no closed V-paths.

In summary, V is a discrete gradient vector field on N_n with exactly one unpaired vertex, and $(n-1)!$ unpaired $(n-3)$-simplices. We can now apply Theorem 11 to conclude

Theorem 23 ([**93**])**.** *The complex Δ_n of disconnected graphs on n vertices is homotopy equivalent to the wedge of $(n-1)!$ spheres of dimension $(n-3)$.*

2.2. Not-2-connected Graphs

Recall that a graph G is *2-connected* if the removal of any vertex (along with all incident edges) results in a connected graph. If G is not 2-connected, we call any vertex v a *cut vertex* if $G - v$ is not connected. Let Δ_n^2 denote the complex of not-2-connected graphs on the vertex set $\{1, 2, \ldots, n\}$.

In this section, we will describe a proof of the following result.

Theorem 24. *For $n \geq 3$, the space Δ_n^2 is homotopy equivalent to a wedge of $(n-2)!$ spheres of dimension $2n-5$.*

This result was first established in [**6**] and [**92**], but we will follow (with only cosmetic changes) the proof, via discrete Morse theory, presented in [**79**]. Let g denote the graph on the vertex set $\{1, 2, \ldots, n\}$ containing only the single edge $(1,2)$. Theorem 24 follows from the following result.

Proposition 25. *There is a discrete gradient vector field on Δ_n^2 whose critical simplices are g along with all graphs of the form shown in Figure 12(i), where $v_3, v_4, \ldots, v_n$ is any permutation of $3, 4, \ldots, n$.*

Let $C_n^2 = \mathcal{G}_n / \Delta_n^2$. Then the cells of C_n^2, with the exception of the distinguished point, correspond to the 2-connected graphs, so we call C_n^2 the *complex of 2-connected graphs*. Our construction of the gradient vector field in Theorem 24 first begins by collapsing towards g. Hence, following the discussion in the previous section, Theorem 25 implies the following result.

Corollary 26. *There is a discrete gradient vector field on C_n^2 whose critical simplices are the special point p in N_n^2 corresponding to Δ_n^2, and all graphs of the form shown in Figure 12(ii).*

Proposition 25 and Corollary 26 will be proved simultaneously, inductively on n. For $n = 3$, the set $\mathcal{G}_3$ of graphs on the vertex set $\{1, 2, 3\}$, is a 2-dimensional simplex on the vertex set consisting of the 3 possible edges $\{(1, 2), (2, 3), (1, 3)\}$. The only graph on 3 vertices which is 2-connected is K_3, which corresponds to the maximal face of $\mathcal{G}_3$. That is, Δ_3^2 is a circle, and the gradient vector field which collapses towards $(1, 2)$, has critical vertex $\{[1, 2]\}$ and critical edge $\{(2, 3), (1, 3)\}$. The space C_n^2, resulting from collapsing Δ_3^2 to a point, is a 2-sphere consisting of the point p and the 2-cell K_3. The only possible gradient vector field in C_n^2 is empty so that both cells are critical. These gradient vector fields satisfy the conclusions of Proposition 25 and Corollary 26.

Now let us begin to construct a gradient vector field on Δ_n^2 for general n (assuming the construction of such a gradient vector field on C_{n-1}^2). First, we collapse towards g. That is, set

$$V_1 = \{\{G - (1, 2), G\}\}$$

where G ranges over all graphs which are not 2-connected and contain the edge $(1, 2)$. Let M_1 denote the graphs which remain unpaired. Then M_1 consists of all graphs G which are not 2-connected, and do not contain the edge $(1, 2)$, and have the property that $G + (1, 2)$ is 2-connected.

To describe the next step in our construction of V, we must take a closer look at such graphs. Such a graph G must be connected (as otherwise $G + (1, 2)$ cannot be 2-connected). Let us now recall the basic structure of connected, not-2-connected graphs. Let H be such a graph, and let H_1 be an induced 2-connected subgraph which is maximal among all induced 2-connected subgraphs. (A subgraph H of a graph G is said to be *induced* if H contains all edges of G which connect two vertices of H.) Let H_2 denote another maximal induced 2-connected subgraph. Then $H_1 \cap H_2$ can contain at most one vertex (as otherwise the induced graph on $V(H_1) \cup V(H_2)$ would be 2-connected, and larger than H_1 and H_2). If $H_1 \cap H_2$ contains a vertex, then that vertex must be a cut vertex of H. Conversely, any cut vertex of H is of the form $H_1 \cap H_2$ for some maximal induced 2-connected subgraphs H_1 and H_2. Now let $H(2)$ denote the graph whose vertices are the maximal induced 2-connected subgraphs of H, with the property that if H_1 and H_2 are maximal induced 2-connected subgraphs of H, then the corresponding vertices of $H(2)$ are adjacent if and only if $H_1 \cap H_2$ is not empty. Clearly every vertex of H is contained in some maximal induced 2-connected subgraph of H. Moreover, H is not 2-connected, which implies that $H(2)$ has at least 2 vertices. Lastly, we observe that every minimal loop in H is contained in some maximal induced 2-connected subgraph of H, and hence appears as a vertex in $H(2)$, from which one can deduce that $H(2)$ has no loops, i.e. $H(2)$ is a tree. Now let G be a connected, not-2-connected graph with the property that $G + (1, 2)$ is 2-connected. Note that vertices 1 and 2 cannot be contained in the same maximal induced 2-connected subgraph of G, as otherwise the blocks of $G + (1, 2)$ would be the same as those of G, and hence $G + (1, 2)$ would not be 2-connected. Let G_1 denote the maximal induced 2-connected subgraph of G that contains the vertex 1, and G_2 the maximal induced 2-connected subgraph of G that contains the vertex 2. It is easy to see that $G_1 \neq G_2$ (as otherwise $G + (1, 2)$ cannot be 2-connected). In fact, the following result is easily established.

Lemma 27. *With all notation as above, $G(2)$ is a path from G_1 to G_2.*

Now let G_3 denote the induced maximal 2-connected subgraph which is adjacent to G_1 in $G(2)$, and let $v(G)$ be the vertex $G_1 \cap G_3$. It is clear that $v(G) \neq 2$. Moreover, if $v(G) = 1$ then vertex 1 would be a cut vertex of $G+(1,2)$, contradicting the assumption that $G + (1,2)$ is 2-connected. Therefore $v(G) \notin \{1,2\}$. Suppose $G \in M_1$ and $(v(G), 2) \notin G$. It is easy to see that $v(G)$ is a cut vertex of $G+(v(G),2)$, and hence $G + (v(G), 2)$ is not 2-connected. Moreover, $[G + (v(G), 2)] + (1,2) = [G + (1,2)] + (v(G), 2)$ is 2-connected (since $[G + (1,2)]$ is), so $G + (v(G), 2) \in M_1$. Now define

$$V_2 = \{ \ \{G, G + (v(G), 2)\}\}$$

where G ranges over all graphs in M_1 which do not contain $(v(G), 2)$.

Let M_2 contain the graphs which are not paired in V_1 or V_2. Then M_2 consists of those graphs G in M_1 which contain $(v(G), 2)$, and which have the property that $G - (v(G), 2) \notin M_1$. First note that since $(v(G), 2) \in G$, $v(G)$ and 2 are contained in an induced 2-connected subgraph, which implies that $v(G) \in G_2$, and hence G_1 and G_2 are connected in $G(2)$. From the previous lemma, we learn that $G(2)$ must consist of only the two vertices G_1 and G_2 and the edge between them. The only way $G - (v(G), 2)$ could fail to be in M_1 is if $G - (v(G), 2)$ failed to be connected. This can happen only if $G_2 - (v(G), 2)$ is not connected. However, since G_2 is 2-connected, this can happen only if G_2 consists entirely of the vertices 2 and $v(G)$ and the edge between them. Thus, the graphs G in M_2 are precisely those that can be constructed by taking a 2-connected graph G_1 on the vertex set $\{1, 2, \ldots, n\} - \{2\}$, adding the vertex 2, and adding the edge $(i, 2)$ for some $i \notin \{1, 2\}$ (in which case $v(G) = i$).

Let $M_2(i), i = 3, 4, \ldots, n$, denote those graphs G in M_2 with $v(G) = i$. Then M_2 is the disjoint union of the $M_2(i)$'s. Each $M_2(i)$ can be canonically identified with the complex Γ of 2-connected graphs on the $n - 1$ vertices $\{1, 3, 4, \ldots, n\}$. By induction, there is a gradient vector field on Γ with precisely $(n - 3)!$ critical simplices of dimension $2(n - 1) - 5. = 2n - 7$. Using the identification, we get a gradient vector field $V_3(i)$ on $M_2(i)$ with $(n - 3)!$ critical simplices of dimension $2n - 6$. Let

$$V = V_1 \cup V_2 \cup (\cup_{i=3}^{n} V_3(i)).$$

Since there are $n - 2$ such $M_2(i)$'s, the total number of unmatched simplices in V is $(n - 2)(n - 3)! = (n - 2)!$, each of dimension $2n - 6$. The theorem now follows once we know that V is a gradient vector field.

Lemma 28. *The vector field V constructed above is a gradient vector field.*

The proof is left as a (rather non-trivial) exercise.

It is, in fact, quite easy to identify more explicitly the critical simplices in the above gradient vector field. To find the critical graphs in $M_2(i)$, $i = 3, 4, \ldots, n$, we take the critical graphs in the complex of 2-connected graphs on the vertex set $\{1, 3, 4, \ldots, n\}$ with respect to some optimal gradient vector field add the vertex 2 and the edge $(i, 2)$ for some $i = 3, 4, \ldots, n$. Fixing i, identify $\{1, 3, 4, \ldots, n\}$ with $\{1, 2, \ldots, n-1\}$ via a correspondence that identifies 1 with 1, and identifies i with 2. By induction, there is a gradient vector field on the 2-connected graphs on the vertex set $\{1, 2, \ldots, n - 1\}$ whose critical simplices have the form shown in Figure 12(ii). Using the identification, we get a gradient vector field on 2-connected graphs on the

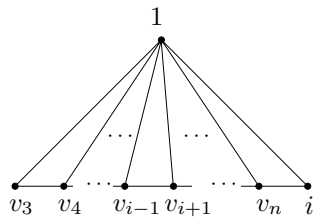


Figure 13. Critical 2-connected graphs on the vertex set $\{1, 2, 3, \ldots, n\}$.

vertex set $\{1, 3, 4, \ldots, n\}$ whose critical simplices are of the form shown in Figure 13 (where $v_3, v_4, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n$ is any permutation of $3, 4, \ldots, i-1, i+1, \ldots, n$).

Adding a vertex 2 to each such graph, and adding an edge between vertex i and vertex 2 yields the desired collection of graphs shown in Figure 12(i). Corollary 26 now follows from Theorem 22.

2.3. Some further thoughts

The reader may wonder why we stopped with not-2-connected graphs. In fact, with quite a bit of hard work, it is possible to go further. In [**52**] J. Jonsson used discrete Morse theory to prove the following result.

Theorem 29. *The simplicial complex Δ_n^3 of not-3-connected graphs is homotopy equivalent to a wedge of $(n-3)\cdot(n-2)!/2$ spheres of dimension $(2n-1)$.*

Many of the gradient vector fields presented in these notes, including the two examples in this section, follow a similar pattern, in that one constructs the gradient vector field in several stages, following distinct rules for each stage. In this way, a user of discrete Morse theory generally discovers the following useful observation, which appeared implicitly earlier, but seems to have been first explicitly stated in [**52**] and [**50**].

Lemma 30. *Let $K = \sqcup_{i\in I} K_i$ be a partition of the faces of K, where I is some partially ordered set. Suppose that for every $i \in I$, $\cup_{j\le i} K_j$ is a subcomplex of K. Now suppose we have a discrete vector field V_i on each K_i (that is, a partial pairing of the simplices in K_i) with the property that there are no closed V_i-paths in K_i. Then $V = \cup_{i\in I} V_i$ is a gradient vector field on K.*

3. The Morse Complex

In this section we will see how more precise knowledge of the gradient vector field on a simplicial complex K allows one to strengthen the conclusions of the main theorems of discrete Morse theory. In particular, rather than just knowing the number of cells in a CW decomposition for K, one can calculate the homology exactly.

Let K be a simplicial complex with a gradient vector field V. In keeping with the standard terminology in the smooth category, we will refer to V-paths (see Section 3) as gradient paths. Let $C_p(K, \mathbb{Z})$ denote the space of simplicial p-chains, and $M_p \subseteq C_p(K, \mathbb{Z})$ the span of the critical p-simplices of V. We refer to M_* as the space of Morse chains. If we let m_p denote the number of critical p-simplices, then we obviously have

$$\mathcal{M}_p \cong \mathbb{Z}^{m_p}.$$

Since homotopy equivalent spaces have isomorphic homology, the following theorem follows from Theorems 11 and 5.

Theorem 31. *There are boundary maps $\tilde{\partial}_d : \mathcal{M}_d \to \mathcal{M}_{d-1}$, for each d, so that*

$$\tilde{\partial}_{d-1} \circ \tilde{\partial}_d = 0$$

and such that the resulting differential complex

$$0 \longrightarrow \mathcal{M}_n \xrightarrow{\tilde{\partial}_n} \dots \xrightarrow{\tilde{\partial}_1} \mathcal{M}_0 \longrightarrow 0 \tag{3}$$

calculates the homology of K. That is, if we define

$$H_d(\mathcal{M}, \tilde{\partial}) = \frac{\mathrm{Ker}(\tilde{\partial}_d)}{\mathrm{Im}(\tilde{\partial}_{d+1})}$$

then for each d

$$H_d(\mathcal{M}, \tilde{\partial}) \cong H_d(K, \mathbb{Z}).$$

In fact, this statement is equivalent to the Strong Morse inequalities (see Exercise 1 of Lecture 1). The main goal of this section is to present an explicit formula for the boundary operator $\tilde{\partial}$. This requires a closer look at the notion of a gradient path. Let β be a critical $(p+1)$ simplex, and and α a critical p-simplex. Then it is easy to check that any gradient-path from β to α has the form

$$\beta = \beta_0^{(p+1)}, \alpha_1^{(p)}, \beta_1^{(p+1)}, \alpha_2^{(p)}, \dots, \beta_r^{(p+1)}, \alpha_{r+1}^{(p)} = \alpha$$

such that for each $i = 0, 1, 2, \dots, r$, $\{\alpha_{i+1} < \beta_{i+1}\} \in V$, and $\alpha_{i+1} < \beta_i$, but $\{\alpha_{i+1} < \beta_i\} \notin V$. In Figure 14 we show a single gradient path from the boundary of a critical 2-simplex β to a critical edge α, where the arrows pointing from an edge to a 2-cell indicate the gradient vector field V.

Given a gradient path as shown in Figure 14, an orientation on β induces an orientation on α. We will not state the precise definition here. The idea is that one "slides" the orientation from β along the gradient path to α. For example, for the gradient path shown in Figure 14, the indicated orientation on β induces the indicated orientation on α.

We are now ready to state the desired formula.

Theorem 32. *Choose an orientation for each simplex. Then for any critical $(p+1)$-simplex β set*

$$\widetilde{\partial}\beta = \sum_{critical\ \alpha^{(p)}} c_{\alpha,\beta}\alpha \tag{4}$$

where

$$c_{\alpha,\beta} = \sum_{\gamma \in \Gamma(\beta,\alpha)} m(\gamma)$$

where $\Gamma(\beta, \alpha)$ is the set of gradient paths which go from β to α. The multiplicity $m(\gamma)$ of any gradient path γ is equal to ± 1, depending on whether, given γ, the orientation on β induces the chosen orientation on α, or the opposite orientation. With this differential, the complex (3) computes the homology of K.

We refer to the complex (3) with the differential (4) as the Morse complex (it goes by many different names in the literature). An extensive study of the Morse complex in the smooth category appears in [**78**]. In is section, we have focused our attention on simplicial complexes. However, it is worth noting that this entire

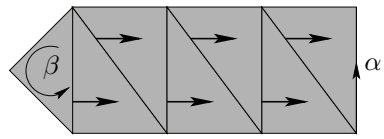


Figure 14. The flow of the edge e.

discussion applies, without any change, to any regular CW-complex, and, after some refinement of the notion of the multiplicity $m(\gamma)$, to all CW complexes. See [**32**] for details.

We only have time to present the main ideas the proof of Theorem 32. For the details, consult Sections 7 and 8 of [**32**]. The key ingredient in the proof is the notion of a (discrete time) flow associated to a discrete vector field V. In the case of smooth manifolds, the gradient vector field defines a dynamical system, namely the flow along the vector field. Viewing the Morse function from the point of view of this dynamical system leads to important new insights [**83**]. The same is true in the combinatorial category.

Up to this point in the notes, we have been thinking of V as a collection of pairs of simplices. Now it is better to think of V as a map of oriented simplices. Namely, choose an orientation for each simplex of M. If $\{\beta^{(p)} < \alpha^{(p+1)}\}$ is an element of V, then we set $V(\beta) = -i\alpha$ where $i = \pm 1$ is the incidence number of β and α (i.e $i = 1$ if the orientations agree, and -1 otherwise). Set $V(\beta^{(p)}) = 0$ if there is no such $\alpha^{(p+1)}$, i.e. if β is not the tail of any arrow in V. Now extend V linearly to a map

$$V : C_p(M, \mathbb{Z}) \to C_{p+1}(M, \mathbb{Z}),$$

and do this for each p.

The flow Φ along the gradient vector field V is a map

$$\Phi : C_p(M, \mathbb{Z}) \to C_p(M, \mathbb{Z}),$$

for each p, defined by the formula

$$\Phi \;=\; 1 + \partial V + V\partial.$$

See Figure 15 for the flow of an oriented edge e. In this figure, we indicate the orientation of e, and just enough of the vector field V in order to determine $\Phi(e)$. We observe that the map Φ commutes with the boundary operator. The other main fact is that for a finite simplicial complex, the map Φ stabilizes in finite time. That is, there is an N such that $\Phi^N = \Phi^{N+1} = \Phi^{N+2} = \dots$ (it is only here that it is necessary that the vector field V be a gradient vector field), and we denote this map by Φ^∞.

Now let us return to the analysis of the Morse complex. Let

$$C_* : 0 \longrightarrow C_n(K, \mathbb{Z}) \xrightarrow{\partial_n} \dots \xrightarrow{\partial_1} C_0(K, \mathbb{Z}) \longrightarrow 0$$

denote the usual simplicial chain complex of K. Let $C_p^\Phi(K, \mathbb{Z}) \subset C_p(K, \mathbb{Z})$ denote the subspace of Φ-invariant chains (i.e. the chains c such that $\Phi(c) = c$). Then, since Φ commutes with the boundary operator ∂, the boundary map takes $C_p^\Phi(K, \mathbb{Z})$ to $C_{p-1}^\Phi(K, \mathbb{Z})$. Now consider the complex of Φ-invariant chains.

$$C_*^\Phi : 0 \longrightarrow C_n^\Phi(K, \mathbb{Z}) \xrightarrow{\partial_n} \dots \xrightarrow{\partial_1} C_0^\Phi(K, \mathbb{Z}) \longrightarrow 0.$$

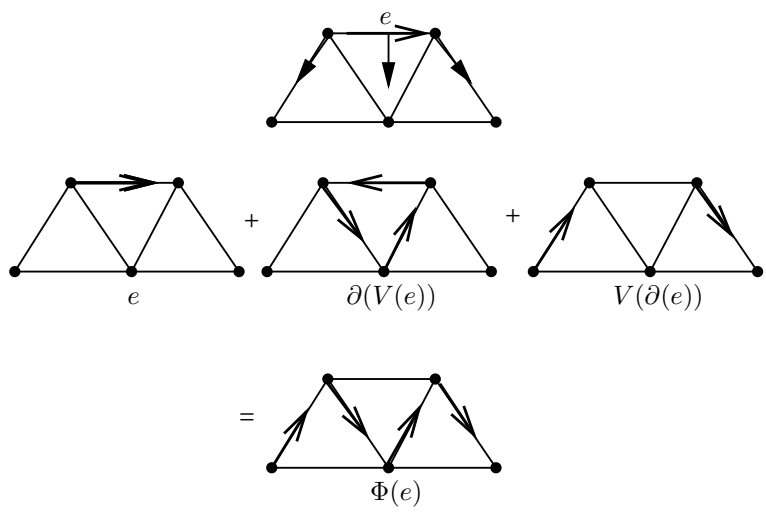


Figure 15. The flow of an oriented edge.

The first step is to see that this complex has the same homology as C_*. There are obvious maps between the two complexes, since C_*^Φ injects into C_*, and Φ^∞ maps C_* onto C_*^Φ. The composition yields the identity map on C_*^Φ. Thus, it is sufficient to show that the map $\Phi^\infty : C_* \to C_*$ induces an isomorphism on homology. For this, it is sufficient to find a homotopy operator. That is, an operator $L : C_*(K, \mathbb{F}) \to C_{*-1}(K, \mathbb{F})$ with the property that $\Phi^\infty - 1 = L\partial + \partial L$. If $\Phi^\infty = \Phi$, then one could take $L = V$. The general case of $\Phi^\infty = \Phi^N$ is similar.

To make the transition to critical simplices, one can establish that

$$\Phi^\infty : \mathcal{M}_p \to C_p(K, \mathbb{Z})$$

is an isomorphism for each p, with inverse the restriction map $r : C_p(K, \mathbb{Z}) \to \mathcal{M}_p$. Theorem 32 now follows if we take $\widetilde{\partial} = r \circ \partial \circ \Phi^\infty$. One must then calculate that this is precisely the operator defined in the statement of the theorem.

A different proof of Theorem 32 is suggested in the exercises.

Example 33. We end this section with a demonstration of how the ideas of this section may be applied to the example of the real projective plane $\mathbb{P}^2$ as illustrated in Figure 8(ii). We saw in Section 11 how discrete Morse Theory can help us see that $\mathbb{P}^2$ has a CW decomposition with exactly one 0-cell, one 1-cell and one 2-cell. Here we will see how Morse theory can distinguish between the spaces which have such a CW decomposition. Let us now calculate the boundary map in the Morse complex corresponding to the gradient vector field illustrated in Figure 8(ii). Choose an orientation for the edge e. To calculate $\widetilde{\partial}(e)$, we must count all of the gradient paths from e to v. There are precisely two such paths, since the unique gradient path beginning at each endpoint of e leads to v. (The gradient path beginning at vertex 1 is the trivial path of 0 steps.) Since the orientation of e induces a $+$ on one endpoint of e, and a $-$ orientation on the other, adding these two paths with their corresponding signs leads us to the formula that $\widetilde{\partial}(e) = 0$. Now choose an orientation for t. It can be seen from Figure 8(ii) that there are precisely two gradient paths from t to e, and both induce the same orientation on

e, so that $\widetilde{\partial}(t) = \pm 2e$. By reversing the chosen orientation on t if necessary, we may assume that $\tilde{\partial}(t) = 2e$. Therefore the homology of the real projective plane can be calculated from the following differential complex.

$$\mathbb{Z} \xrightarrow{\times 2} \mathbb{Z} \xrightarrow{0} \mathbb{Z} \longrightarrow 0.$$

Thus we see that

$$H_0(\mathbb{P}^2, \mathbb{Z}) \cong \mathbb{Z}, \qquad H_1(\mathbb{P}^2, \mathbb{Z}) \cong \mathbb{Z}/2\mathbb{Z}, \qquad H_2(\mathbb{P}^2, \mathbb{Z}) \cong 0.$$

4. Canceling Critical Points

One of the main problems in Morse theory, whether in the combinatorial or smooth setting, is to find a Morse function, or equivalently a gradient vector field, for a given space with the fewest possible critical points (much of the book [**80**] is devoted to this topic). In general this is a very difficult problem, since, in particular, it contains the Poincaré conjecture – spheres can be recognized as those spaces which have a Morse function with precisely 2 critical points. In [**72**], Milnor presents Smale's proof [**83**] of the higher dimensional Poincaré conjecture (in fact, a proof is presented of the more general h-cobordism theorem) completely in the language of Morse theory. Drastically oversimplifying matters, the proof of the higher Poincaré conjecture can be described as follows. Let M be a smooth manifold of dimension ≥ 5 which is homotopy equivalent to a sphere. Endow M with a (smooth) Morse function f. One then proceeds to show that the critical points of f can be canceled out in pairs until one is left with a Morse function with exactly two critical points, which implies that M is a (topological) sphere.

A key step in this proof is the "cancellation theorem" which provides a sufficient condition for two critical points to be canceled (see Theorem 5.4 in [**72**], which Milnor calls "The First Cancellation Theorem", or the original proof in [**74**]). In this section we will see that the analogous theorem holds for discrete Morse functions. Moreover, in the combinatorial setting the proof is much simpler. The main result is that if $\alpha^{(p)}$ and $\beta^{(p+1)}$ are 2 critical simplices, and if there is exactly 1 gradient path from β to α, then α and β can be canceled. More precisely,

Theorem 34. *Suppose V is a discrete gradient vector field on M such that $\beta^{(p+1)}$ and $\alpha^{(p)}$ are critical, and there is exactly one V-path from β to α. Then there is another gradient vector field W on M with the same critical simplices except that α and β are no longer critical. Moreover, W is equal to V except along the unique gradient path from β to α.*

In the smooth case, the proof, either as presented originally by Morse in [**74**] or as presented in [**72**], is rather technical. In our discrete case the proof is simple. If, in the top drawing in Figure 16, the indicated gradient path is the only V-path from β to α, then we can reverse the gradient vector field along this path, replacing V by the vector field W shown in the bottom drawing in Figure 14.

The uniqueness of the V-path implies that the resulting discrete vector field has no closed orbits, and hence, by Theorem 2, is a gradient vector field. Moreover, α and β are not critical for this new gradient vector field, while the criticality of all other simplices is unchanged. This completes the proof.

The proof in the smooth case proceeds along the same lines. However, in addition to turning around those vectors along the unique gradient path from β to α, one must also adjust all nearby vectors so that the resulting vector field is

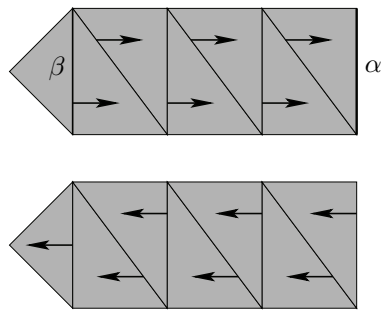


Figure 16. Canceling critical simplices.

smooth. Moreover, one must check that the new vector field is a gradient vector field, so that, in particular, modifying the vectors did not result in the creation of a closed orbit. This is an example of the sort of complications which arise in the smooth setting, but which do not make an appearance in the discrete theory.

This theorem was recently put to very good use in [**7**], in which discrete Morse theory is used to determine the homotopy type of some simplicial complexes arising in the study of partitions. It is fascinating, and quite pleasing, to see the same idea play a central role in two subjects, the Poincaré conjecture and the study of partitions, which seem to have so little to do with one another. In [**50**], Hersh generalizes this cancellation technique and investigates, among other ideas, when families of pairs of critical simplices can be canceled simultaneously. The main theorem of this section is also used extensively in [**54**] as a basic computational tool for searching for optimal gradient vector fields. To see other computational approaches to finding optimal gradient vector fields, the reader can take a look at [**62**], [**63**], [**64**].

5. Exercises for Lecture 2

(1) In the lecture we found a perfect gradient vector field on the complex Δ_n of disconnected graphs on n vertices. Since our construction began by collapsing everything towards the graph g containing only the edge $[1,2]$ we saw that this is equivalent to the construction of a perfect gradient vector field on the complex of connected graphs on n vertices. The critical connected graphs are precisely the critical disconnected graphs with the edge $[1,2]$ added. The result is the set of increasing trees on n-vertices. That is, the trees with vertex set $\{1,2,\ldots,n\}$ with the property that the labels increase along every ray starting at vertex 1. Note that there are $(n-1)!$ of these, and each contains $(n-1)$ edges (and thus corresponds to a simplex of dimension $(n-2)$). Consider the collection $\mathcal{P}_n$ of graphs on n-vertices which are paths with one endpoint labeled 1. That is, graphs of the form

$$1 - v_2 - v_3 - \cdots - v_n$$

where $\{v_2, v_3, \ldots, v_n\} = \{2,3,\ldots,n\}$. We observe that there are precisely $(n-1)!$ of these graphs, and each has $(n-1)$ edges. Your job is to construct a gradient vector field on the simplicial complex of connected graphs on

n vertices for which the critical graphs are precisely the graphs in $\mathcal{P}_n$. (In Vassiliev's original work on this complex, this is the form in which he presented the answer.)

(2) Let G be any graph, and let P be any monotone decreasing graph property. Then we can consider the simplicial complex of all spanning subgraphs of G that satisfy P.

In the lecture we only considered the case where G is a complete graph. For other graphs G, these complexes are quite interesting and largely unexplored.

(a) Examine this in the case where P is the property of being disconnected. What is the homotopy type of the resulting complex? That is, given a graph G, what is the homotopy type of the simplicial complex of disconnected spanning subgraphs of G?

(b) Pick your favorite monotone graph property and your favorite graph and examine the resulting simplicial complex.

(3) Let M be a simplicial complex with a gradient vector field V. Prove that the homology of the Morse complex (as defined in this lecture) is isomorphic to the homology of M by following these steps:

(a) Suppose that V is the empty gradient vector field. Then the Morse complex is just the standard simplicial chain complex of M.

(b) Now prove that the homology of the Morse complex of V does not change if one pair is removed from V (i.e. if one arrow is erased). Do this by showing that if

$$M: \ 0 \to M_n \xrightarrow{d} M_{n-1} \xrightarrow{d} M_{n-2} \xrightarrow{d} \cdots \xrightarrow{d} M_0 \to 0$$

and

$$M': \ 0 \to M'_n \xrightarrow{d'} M'_{n-1} \xrightarrow{d'} M'_{n-2} \xrightarrow{d'} c \ldots \xrightarrow{d'} M'_0 \to 0$$

are the Morse complexes corresponding to gradient vector fields V and V' on M which differ by a single arrow, then there is a map $\Phi : M_i \to M'_i$ which induces an isomorphism on homology. Try to construct the map Φ as explicitly as possible.

Together (a) and (b) prove the desired result.

(4) In the exercises to Lecture 1 we proved that every triangulated surface has a perfect gradient vector field. Consider the Morse complex corresponding to such a vector field. Prove that all of the differentials vanish (that is, each differential is the zero map). Can you understand this directly from the definition of the differential – that is by counting gradient paths?

LECTURE 3
Discrete Morse Theory and Evasiveness

1. The Main Results

So far, we have indicated some applications of discrete Morse theory to combinatorics and topology. We now present an application to computer science. The reader should see the reference [**36**] for a more complete treatment of the content of this section.

There is a wide variety of situations in which one has the ability to quickly ask a series of yes/no questions, with the goal of answering a more difficult question. For example, when one goes to the doctor with an illness, the doctor usually asks a series of yes/no questions, such as "Do you have a headache?", "Do you have a fever?", etc., using the information from the previous questions to decide what to ask next, with the goal of answering the more difficult question "What illness does my patient have?". When one takes a malfunctioning car to the mechanic, the mechanic often attempts to analyze the problem by testing the individual components one at a time, using the appropriate tools to ask "Is this component working properly?", with the goal of answering the more difficult question "What is wrong with this car?". The mathematical study of such questions began with the following sort of problem. Suppose that one has a network G of phones, or computers, or... which we think of as a collection of nodes, some of which are connected by arcs. We assume that the nework G is connected. That is, one can get from any node to any other node by a series of arcs. Now we suppose that there is an electrical storm, or a terrorist attack, or..., and some of the arcs are disabled. At that time, our first concern may not be "What is the precise network that remains?" but rather, we may primarily concerned with questions such as "Is the remaining network still connected?". This is the difficult question we wish to answer. We suppose further that we have the capability of testing each arc in the original network, in order to answer the question "Is this arc still working?". Of course, we can answer any question about the network if we simply test every arc in the original network, and determine precisely which of these arcs are still working, as that completely determines the remaining network. The precise question we want to analyse here is "Can one do any better?". That is, is there any strategy for testing the arcs such that we are guaranteed that we can answer the desired question before having tested each of the original arcs.

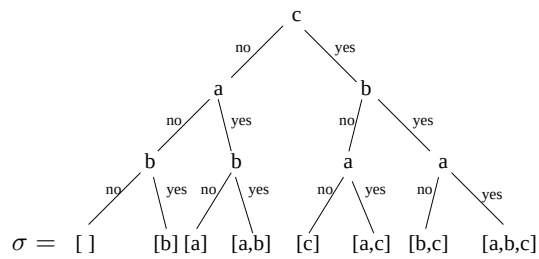


Figure 17. A search algorithm.

Let us begin with a simple example of the sort of thing we wish to study. Suppose there are three yes/no questions that we can easily ask. We label these questions $\{a, b, c\}$.

Assumption 1: We suppose that these questions have the property that their answers are independent of the order in which they are asked. (We will make this assumption for the rest of these notes.)

Then there are eight possible outcomes resulting from asking these three questions. We label these outcomes by listing the questions that yield the answer "yes" for that outcome. The possibilities are: $[\,]$, $[a]$, $[b]$, $[c]$, $[ab]$, $[ac]$, $[bc]$, $[abc]$.

Assumption 2: We assume that every set of answers is possible.

That is, one can easily imagine a set of questions with the property that questions b and c can not both be answered "yes", but we will not consider this possibility in these lectures. We make this assumption only for reasons of simplicity. The general situation is considered in [**41**]. Suppose that the following four outcomes are good: $[a]$,$[b]$,$[c]$,$[ab]$, and the remaining outcomes are bad. By asking these three questions, our goal is to determine whether the outcome is good or bad. We can, of course, accomplish this goal by asking all three questions. We are considered to have won this game if we achieve the goal before we ask the third question. A winning strategy, then, is one which guarantees that no matter what the outcome is, we can determine whether or not it is good or bad before asking the third question.

For example, consider the search algorithm shown in Figure 17, in which case we have listed the question to be asked next, given the answers to the previous questions. For example, we ask question c first, and if we get the answer "yes" we ask question b, but if we get the answer "no", we ask the question a. We observe that, asking questions in the indicated order, if the outcome is in the set $\{[\,], [b], [c], [ac]\}$, then we must ask the third question. Outcomes which require us to ask the third question are called *evaders* of the search algorithm, so the algorithm has 4 evaders. In fact, this is the best one can do. The following proposition is fairly easy to check by straightforward means.

Proposition 35. *Every search algorithm for the problem of determining membership in the set of good outcomes $\{[a], [b], [c], [ab]\}$ has at least 4 evaders. The number of evaders which are good equals the number of evaders which are bad, and hence there must be at least two of each.*

If we assume that each outcome is equally likely, then this proposition implies that no matter which search algorithm we choose, we will have to ask the third question at least half of the time. Note that this theorem does not say that every search algorithm has exactly 4 evaders, and it is rather easy to find search algorithms

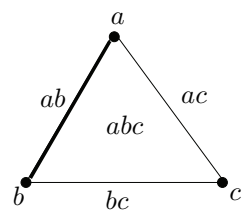


Figure 18. A topological approach to the problem.

with more than 4 evaders. If every search algorithm has some evaders, so that we have no winning strategy, then we say that the problem is *evasive*.

It is probably not at all clear to the reader what this topic is doing in a series of lectures on discrete Morse theory, but we will show that in fact these topics are intimately related. In particular, we will show that algebraic topology gives a way of understanding why some problems of this form are easy, and others are hard. First we observe that the problem can easily be stated in a more topological way. Consider the 2-dimensional simplex S with vertices labeled $\{a, b, c\}$. Then the faces of S can be indentified with the subsets of $\{a, b, c\}$, and hence with the 8 possible outcomes (see Figure 18). Then the good and bad outcomes partition the faces of S into 2 sets. In this setting we are given a partition of the set of faces, the outcome is a face σ of the simplex, and our goal is to determine which block of the partition contains σ. We are permitted to ask questions of the form "Is vertex v in σ?".

In this way, we can convert binary search problems (which satisfy Assumption 1) into the language of simplices. If we also require Assumption 2, then the sort of search problems we are considering lead to problems of the following form. Let S be an n-dimensional simplex, with vertices $v_0, v_1, \ldots, v_n$, F the set of faces of S, and

$$P: \ F = P_1 \sqcup P_2 \sqcup \ldots \sqcup P_k$$

a partition of F, which is known to you. Let σ be a face of S which is not known to you. Your goal is to determine which block of the partition P contains σ. In particular, you need not determine the face σ. You are permitted to ask questions of the form "Is v_i in σ?". You may use the answers to the questions you have already asked in determining which vertex to ask about next. Of course, you can determine which block contains σ by asking $n+1$ questions, since by asking about all $n+1$ vertices you can completely determine σ. You win this game if you answer the given question after asking fewer than $n+1$ questions.

Say that P is nonevasive if there is a winning strategy for this game, i.e there is a search algorithm that determines which block contains σ in fewer than $n+1$ questions, no matter what σ is. Say P is evasive otherwise.

One of the main issues we will have to deal with is that a block P_i of the partition need not be a subcomplex or have any other nice structure. Hence, the notion of the homology of such a set is problematic. Let P be any set of faces of a simplex S, and let $\mathbb{F}$ be a field. One of the main contributions of this and the following sections is a definition of the $\mathbb{F}$-Betti numbers of P. More precisely, for each $i = -1, 0, 1, \ldots$, we will define $B_i(P, \mathbb{F})$, the ith Betti number of P with respect to the field $\mathbb{F}$. We will also define the even and odd Betti numbers, denoted $B_e(P, \mathbb{F})$ and $B_o(P, \mathbb{F})$, respectively, and the total Betti number $B(P, \mathbb{F})$. For ease of notation, we will assume that the field $\mathbb{F}$ is fixed, and refer to $B_i(P), B_e(P), B_o(P)$

and $B(P)$. We will present the precise definition of these numbers in the next section. The basic idea is that the Betti number $B_i(P)$ is defined by restricting the chain complex of S over the field $\mathbb{F}$ to the faces in P. The result need not be a complex, since the composition of its consecutive differentials might be nonzero, but the dimension of its ith "homology" can still be defined as the dimension of its ith kernel minus the dimension of its $(i+1)^{st}$ image (if that is nonnegative, and 0 otherwise); see Proposition 47. At this point, we will state the main properties of these Betti numbers.

Theorem 36. *For any set of faces P*

(i)
$$B_e(P) \geq \sum_{i \text{ even}} B_i(P),$$
$$B_o(P) \geq \sum_{i \text{ odd}} B_i(P),$$
$$B(P) = B_e(P) + B_o(P).$$

(ii) $B_i(P) = 0$ *for i larger than the dimension of P.*

(iii) $B_e(P) - B_o(P) = \chi(P) = \sum_i (-1)^i \#\{i\text{-simplices in } P\}$

Our notion of a Betti number is equal to a standard notion in a number of settings.

Theorem 37.

(1) If P is a subcomplex of S, and the empty set (considered as a face of S) is an element of P, then for each i
$$B_i(P) = \dim \widetilde{H}_i(P, \mathbb{F}).$$
where the tilde denotes reduced homology. Moreover,
$$B_e(P) = \dim \widetilde{H}_{even}(P, \mathbb{F})$$
$$B_o(P) = \dim \widetilde{H}_{odd}(P, \mathbb{F})$$
$$B(P) = \dim \widetilde{H}_*(P, \mathbb{F}).$$

(2) If P is a subcomplex of S, except that the empty set is not element of P, then for each i
$$B_i(P) = \dim H_i(P, \mathbb{F}).$$
Moreover,
$$B_e(P) = \dim H_{even}(P, \mathbb{F})$$
$$B_o(P) = \dim H_{odd}(P, \mathbb{F})$$
$$B(P) = \dim H_*(P, \mathbb{F}).$$

(3) Let $\overline{P}$ denote the closure of P (i.e. the set consisting of the faces of P along with all of their faces), and let $\dot{P} = \overline{P} - P$. If $\dot{P}$ is a subcomplex of S (which contains the empty set) then for each i
$$B_i(P) = \dim H_i(\overline{P}, P, \mathbb{F}).$$
Moreover
$$B_e(P) = \dim H_{even}(\overline{P}, P, \mathbb{F})$$
$$B_o(P) = \dim H_{odd}(\overline{P}, P, \mathbb{F})$$
$$B(P) = \dim H_*(\overline{P}, P, \mathbb{F}).$$

Assuming these results for now, as well as the still undefined notion of Betti number, we present the main theorem of this section.

Theorem 38. *With all notation as above, for any search algorithm A the number of evaders of A which lie in any block P_j of the partition P is at least $B(P_j)$. Hence the total number of evaders is at least $\sum_{j=1}^{k} B(P_j)$.*

In fact, we can make this statement much more precise. Define the dimension of an evader to be the dimension of the face of S to which it corresponds. That is, if σ is any possible outcome, $\dim(\sigma)$ is

$$\text{(the number of questions answered “yes” if the outcome is } \sigma) - 1.$$

Theorem 39. *With all notation as above, for any search algorithm A the number of evaders of A of dimension i which lie in any block P_j of the partition P is at least $B_i(P_j)$. The number of even-dimensional evaders which lie in block P_j is at least $B_e(P_j)$, and the number of odd-dimensional evaders which lie in block P_j is at least $B_o(P_j)$.*

Before discussing the proof of this result, we would like to point out that Kahn, Saks and Sturtevant [**55**] first observed the relationship between evasiveness and algebraic topology. In their setting, the partition consists of precisely two blocks, $P : S = P_1 \sqcup P_2$, in which P_1 is a subcomplex. They proved the following theorem.

Theorem 40. *If $\tilde{H}_*(P_1) \neq 0$, where $\tilde{H}_*(P_1)$ denotes the reduced homology of P_1, then P is evasive.*

In fact, they proved something stronger, and we will come back to this point later. In [**39**] we used discrete Morse theory to make some of their results more quantitative along the lines of Theorems 38 and 39. The generalization in this section to more than two blocks is relatively minor. The extension to more general sets of faces is the major value of this newer work.

We illustrate the previous theorems by returning to the example introduced at the beginning of this section. Let P_1 denote the set $\{[a], [b], [c], [ab]\}$ of good outcomes, and let P_2 denote the complement, the set of bad outcomes. We observe that P_1 is a simplicial complex which does not contain the empty face. Hence by Theorem 37, $B(P_1)$ is equal to the dimension of the (unreduced) homology of P_1, which is 2. By Theorem 38, we learn that for any search algorithm, the number of evaders which lie in P_1 is at least 2. We observe that P_2 does not satisfy any of the hypotheses presented in Theorem 37, so one can not deduce its Betti numbers from that result. However, as the reader can check (after we present the definition of Betti numbers in the next section), its total Betti number is also 2.

The link between evasiveness and algebraic topology is provided by discrete Morse theory. Morse theory comes to the fore when one observes that a search algorithm induces a discrete vector field on S. For example, the search algorithm shown in Figure 17 induces the vector field

$$V = \{\, \{[\,] < [b]\},\ \{[a] < [a, b]\},\ \{[c] < [a, c]\},\ \{[b, c] < [a, b, c]\} \,\}$$

That is, V consists of those pairs of faces of S which are not distinguished by the search algorithm until the last question. There is slight subtlety here in that a search algorithm pairs a vertex with the empty face [ ], while in our original definition, it was not permitted to pair a simplex with [ ]. Thus, to get a true discrete vector field, we must remove this pair from V. (It is precisely this subtle point that results in the reduced homology of K being the relevant measure of topological complexity

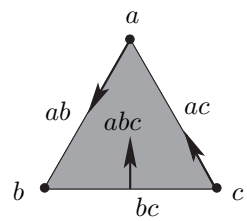


Figure 19. The induced vector field V.

in Theorem 37(1), rather than the unreduced homology.) However, for simplicity, from now on we will simply ignore this technical point.

Theorem 41. *For any search algorithm, let V denote the vector field consisting of pairs of nonempty faces of S which are not distinguished by the search algorithm until the last question. Then V is a gradient vector field.*

We will postpone the proof of this result until the end of this section.

For now, suppose that block P_j of the partition P is a subcomplex (containing the empty face). We will complete the proof in this setting before discussing the general case. Now restrict V to P_j by taking only those pairs in V such that both simplices are in P_j, and denote the resulting vector field by V_j. In our simple example, this results in the vector field

$$V_1 = \{\{[a] < [a,b]\}\}.$$

From the previous theorem, V has no closed orbits. Any discrete vector field consisting of a subset of the pairs of V has fewer paths, and hence also has no closed orbits. Therefore, V_j is a gradient vector field on P_j. Note that V pairs every face of S with another face, and hence there are no critical simplices except for the vertex which is paired with the empty set. Thus, ignoring that special vertex for the moment, the critical simplices of V_j are precisely the simplices of P_j which are paired in V with a face of S which is not in P_j. These are precisely the simplices of P_j which are the evaders of the search algorithm.

The Morse inequalities of Theorem 13 (i) immediately imply the following result.

Corollary 42. *If the block P_j of the partition P is a subcomplex (containing the empty face of S) then the number of evaders in P_j is at least* $\dim \widetilde{H}_*(P_j)$.

(We must use reduced homology here because of the minor issue surrounding the vertex paired with the empty set.) This yields Theorem 37 (in the case of a simplicial complex containing the empty face).

Suppose that P is nonevasive. Then there is some search algorithm which has no evaders. From our above discussion we have seen that this implies that P_j has a gradient vector field with no critical simplices. Actually, this is not quite true. The gradient vector field must have a critical vertex – the vertex that is paired with the empty face. These ideas lead to the following strengthening of Theorem 40.

Theorem 43. *If P is nonevasive, and if the block P_j of the partition is a subcomplex, then P_j collapses to a vertex.*

This theorem appears in [**55**], the paper that first established, and used to very good effect, a close relationship between evasiveness and topology. The interested

reader can consult [**36**] for some additional refinements of this theorem. This topic has been the subject of much study, and the reader can find more information about the connection between evasiveness and topology in the references [**11**], [**56**], [**76**], [**77**], and [**94**] .

We now present a proof of Theorem 41. Let S denote an n-simplex, and fix a search algorithm. Associate to each p-simplex α of S the sequence of integers

$$n(\alpha) = n_0(\alpha) < n_1(\alpha) < \cdots < n_p(\alpha)$$

where, for each i, question number $n_i(\alpha)$ is answered "yes" if $\sigma = \alpha$, and these are the only questions answered "yes".

Let V be the vector field induced by the search algorithm and

$$\alpha_1, \alpha_2$$

be a V-path. Then either (i) α_1 is a face of α_2 and $\{\alpha_1 < \alpha_2\} \in V$, or (ii) α_2 is a face of α_1 and $\{\alpha_1 < \alpha_2\} \notin V$. Let us consider case (ii) first. In this case, α_2 has one fewer vertex than α_1, and the vertex is not the subject of the $(n+1)^{st}$ question. Suppose the the vertex is the subject of the $n_i(\alpha_1)^{st}$ question. Then this question is answered "yes" for α_1, but "no" for α_2. This implies that

$$n(\alpha_2) = n_0(\alpha_1) < n_1(\alpha_1) < \cdots < n_{i-1}(\alpha_1) < n_i(\alpha_2) < \cdots$$

for some $i < n+1$, and such that $n_i(\alpha_2) > n_i(\alpha_1)$. Thus $n(\alpha_2) > n(\alpha_1)$ in the lexicographic order.

We now consider case (i), in which $\{\alpha_1 < \alpha_2\} \in V$, and continue the V-path one more step to $\alpha_1, \alpha_2, \alpha_3$. Then α_1 and α_2 are not distinguished until the $(n+1)^{st}$ question. Thus,

$$n(\alpha_2) = n_0(\alpha_1) < n_1(\alpha_1) < \cdots < n_p(\alpha_1) < n+1.$$

We now observe that the vertices of α_3 are a subset of the vertices of α_2. Suppose the vertex of α_2 which is not in α_3 is the vertex tested in question $n_i(\alpha_2)$. Then we must have $i \neq n+1$. This demonstrates that

$$n(\alpha_3) = n_0(\alpha_1) < n_1(\alpha_1) < \cdots < n_{i-1}(\alpha_1) < n_i(\alpha_3) < \cdots$$

for some $i < n+1$, and such that $n_i(\alpha_3) > n_i(\alpha_1)$. Thus $n(\alpha_3) > n(\alpha_1)$ in the lexicographic order, which is sufficient to prove that there are no closed orbits.

In [**53**], Jonsson investigates further the question of which gradient vector fields arise from decision trees. Anyone interested in this topic should also consult [**84**].

2. Betti Numbers for General Sets of Faces

In this section we examine how to extend the results of the previous section to sets of faces of a simplex that do not form a simplicial complex or have any other special structure. A more extensive treatment of these ideas can be found in [**41**]. Let $\mathbb{F}$ be any field. Let

$$\mathcal{V} : 0 \longrightarrow V_n \xrightarrow{\partial_n} V_{n-1} \xrightarrow{\partial_{n-1}} \cdots \xrightarrow{\partial_1} V_0 \xrightarrow{\partial_0} V_{-1} \longrightarrow 0$$

be a complex of finite dimensional vector spaces over the field $\mathbb{F}$. The ∂_i's are assumed to be linear maps, but we are not assuming that $\partial_d \circ \partial_{d+1} = 0$. Our goal is to define the "homology" of this complex.

If $\partial_p \circ \partial_{p+1} = 0$ for each p, then we say that $\mathcal{V}$ is a *differential complex* and that the ∂_i's form a *differential*. We recall that one defines the homology of a differential complex by the formula

$$H_p(\mathcal{V}) := \frac{\operatorname{Ker} \partial_p}{\operatorname{Im} \partial_{p+1}}. \tag{5}$$

For each p, choose a subspace X_p which is mapped isomorphically onto $\operatorname{Im} \partial_p$. Then we have that

$$V_p = X_p \oplus \operatorname{Ker} \partial_p.$$

In the case of a differential complex, $\operatorname{Im} \partial_{p+1} \subset \operatorname{Ker} \partial_p$ we can find a $Z_p \subset V_p$ so that

$$\operatorname{Ker} \partial_p = \operatorname{Im} \partial_{p+1} \oplus Z_p,$$

which implies that

$$V_p = X_p \oplus \operatorname{Im} \partial_{p+1} \oplus Z_p,$$

and the reader can easily check that $Z_p \cong H_p(\mathcal{V})$.

We now return to the general case of a nondifferential complex. That is, we no longer assume that $\partial_p \circ \partial_{p+1} = 0$. We will use the construction of the previous paragraph to define the homology of such a complex.

Definition 44. A homological decomposition D of the complex S is a decomposition

$$V_i = X_i \oplus Y_i \oplus Z_i,$$

for each i, with the property that for each i, ∂_i maps X_i isomorphically onto Y_{i-1}.

By the notation $V_i = X_i \oplus Y_i \oplus Z_i$ we mean that X_i, Y_i and Z_i are linear subspaces of V_i, such that their pairwise intersections are $\{0\}$, and they sum to give all of V_i. Homological decompositions always exist, since one can take $X_i = 0, Y_i = 0$, and $Z_i = V_i$, for each i.

For any homological decomposition D of $\mathcal{V}$, and any i, let $B_i(\mathcal{V}, D)$ denote the dimension of Z_i. We also define the even Betti number of D

$$B_e(\mathcal{V}, \mathcal{D}) := \sum_{i \text{ even}} B_i(\mathcal{V}, \mathcal{D}),$$

the odd Betti number of D

$$B_o(\mathcal{V}, \mathcal{D}) = \sum_{i \text{ odd}} B_i(\mathcal{V}, \mathcal{D}),$$

and the total Betti number of D

$$B(\mathcal{V}, \mathcal{D}) = B_e(\mathcal{V}, \mathcal{D}) + B_o(\mathcal{V}, \mathcal{D}) = \sum_i B_i(\mathcal{V}, \mathcal{D}).$$

We now define the Betti numbers of S by

$$B_i(\mathcal{V}) := \min_{\mathcal{D}} B_i(\mathcal{V}, \mathcal{D}),$$
$$B_e(\mathcal{V}) := \min_{\mathcal{D}} B_e(\mathcal{V}, \mathcal{D}),$$
$$B_o(\mathcal{V}) := \min_{\mathcal{D}} B_o(\mathcal{V}, \mathcal{D}),$$

and

$$B(\mathcal{V}) := \min_{\mathcal{D}} B(\mathcal{V}, \mathcal{D}).$$

We observe the following facts.

Proposition 45. *Let $\mathcal{V}$ be any finite complex of finite dimensional vector spaces.*
(i) $B_e(S) \geq \sum_{i\ even} B_i(S)$ *and* $B_o(S) \geq \sum_{i\ odd} B_i(S)$.
(ii) $B(S) = B_e(S) + B_o(S)$.

Example 46. A simple example will serve to show that the inequalities in part (i) of the proposition can be strict when $\mathcal{V}$ is not a differential complex. Consider the complex $\mathcal{V}$ with $V_0 = V_1 = V_2 = \mathbb{F}$, and $V_i = 0$ for $i = -1$ and $i > 2$. Suppose that ∂_1 and ∂_0 are both the identity map.

$$\mathcal{V} : 0 \longrightarrow \mathbb{F} \xrightarrow{\partial_1} \mathbb{F} \xrightarrow{\partial_0} \mathbb{F} \longrightarrow 0$$

Let D_1 denote the homological decomposition

$$0 \longrightarrow \mathbb{F} \oplus 0 \oplus 0 \xrightarrow{\partial_1} 0 \oplus \mathbb{F} \oplus 0 \xrightarrow{\partial_0} 0 \oplus 0 \oplus \mathbb{F} \longrightarrow 0.$$

We have that $B_1(\mathcal{V}, D_1) = B_2(\mathcal{V}, D_1) = 0$, while $B_0(\mathcal{V}, D_1) = 1$, which implies that $B_1(\mathcal{V}) = B_2(\mathcal{V}) = 0$, and $B_0(\mathcal{V}) \leq 1$.

Let D_2 denote the homological decomposition

$$0 \longrightarrow 0 \oplus 0 \oplus \mathbb{F} \xrightarrow{\partial_1} \mathbb{F} \oplus 0 \oplus 0 \xrightarrow{\partial_0} 0 \oplus \mathbb{F} \oplus 0 \longrightarrow 0.$$

In this case we see that $B_0(\mathcal{V}, D_2) = B_1(\mathcal{V}, D_2) = 0$, while $B_2(\mathcal{V}, D_2) = 1$, which implies that $B_0(\mathcal{V}) = B_1(\mathcal{V}) = 0$, and $B_2(\mathcal{V}) \leq 1$.

Thus we learn that $B_i(\mathcal{V}) = 0$ for every i. On the other hand $B_o(\mathcal{V}, D_1) = B_o(\mathcal{V}, D_2) = 0$, which implies that $B_o(\mathcal{V}) = 0$. We note that $B_e(\mathcal{V}, D_1) = B_e(\mathcal{V}, D_2) = 1$, and, in fact, once can easily see that $B_e(\mathcal{V}) = 1$.

In the case that $\mathcal{V}$ is a differential complex, we have

$$\dim H_i(\mathcal{V}) = \dim(\operatorname{Ker} \partial_i) - \dim(\operatorname{Im} \partial_{i+1}).$$

A remnant of this equation holds for general complexes.

Proposition 47. *For any complex S, whether a differential complex or not,*

$$B_i(\mathcal{V}) = \max\{\dim(\operatorname{Ker} \partial_i) - \dim(\operatorname{Im} \partial_{i+1}), 0\}$$

Since the right hand side is easily algorithmically computable by standard methods, this theorem implies that the generalized Betti numbers are also readily computable. Moreover, we note that these definitions do, in fact, generalize the standard definition those for a differential complex.

Theorem 48. *Suppose that $\mathcal{V}$ is a differential complex, and let $H_i(\mathcal{V})$ denote the homology of S as defined by the standard formula (5). Then for each i,*

$$B_i(\mathcal{V}) = \dim H_i(\mathcal{V}).$$

Moreover,

$$B_e(\mathcal{V}) = \sum_{i\ even} B_i(\mathcal{V})$$

$$B_o(\mathcal{V}) = \sum_{i\ odd} B_i(\mathcal{V}),$$

and

$$B(\mathcal{V}) = \sum_i B_i(\mathcal{V}).$$

Now let S be a simplex. A *simplex space* is defined to be a set of faces of S (we consider the empty set to be a face of S). Let K be a simplex space. Fix a field $\mathbb{F}$ and let $C_d(K,\mathbb{F})$ denote the oriented d-chains in K. This is defined in the usual way, even if K is not a subcomplex, that is, for all $d \geq 2$, let $S_d(K)$ denote the set of pairs (σ,ϵ), where σ is a d-simplex in K, and ϵ is an orientation on σ (i.e. an ordering of the vertices in σ modulo even permutations – so that each σ has two distinct orientations). We let $\widetilde{C}_d(K,\mathbb{F})$ denote the vector space of all formal linear combinations of elements in $S_d(K)$, and $C_d(K,\mathbb{F})$ the result of quotienting out $\tilde{C}_d(K,\mathbb{F})$ by the relation $-1(\sigma,\epsilon) = (\sigma,-\epsilon)$, where if ϵ denotes an orientation on a d-simplex σ, then $-\epsilon$ denotes the alternate orientation. If $d = -1, 0$, then each d-simplex in K has a unique orientation, and we let $C_d(K,\mathbb{F})$ denote the vector space of all formal linear combinations of d-simplices. Note that the only possible -1-dimensional simplex is the empty simplex. Thus, if K contains the empty simplex $C_{-1}(K,\mathbb{F}) \cong \mathbb{F}$ and otherwise $C_{-1}(K,\mathbb{F}) \cong 0$.

We now define a boundary map ∂_d from $C_d(K,\mathbb{F})$ to $C_{d-1}(K,\mathbb{F})$, by setting, for any oriented d-simplex $[x_0, x_1, \ldots, x_d]$,

$$\partial_d[x_0, x_1, \ldots, x_d] = \sum_i{}' (-1)^i [x_0, x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d].$$

where the notation $\sum'$ means that the sum is over all $i \subset \{0, 1, 2, \ldots, d\}$ such that the simplex $\{x_0, x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d\}$ is in K. As usual, the notation $[x_0, x_1, \ldots, x_d]$ denotes the simplex $\{x_0, x_1, \ldots, x_d\}$ along with the orientation induced by ordering the vertices in the manner in which they are listed.

We let $C(K,\mathbb{F})$ denote the complex

$$0 \longrightarrow C_n(K,\mathbb{F}) \xrightarrow{\partial_n} \ldots \xrightarrow{\partial_0} C_{-1}(K,\mathbb{F}) \longrightarrow 0.$$

Using the notation of the previous section, we denote by

$$B_i(K,\mathbb{F}), \quad B_e(K,\mathbb{F}), \quad B_o(K,\mathbb{F}), \quad B(K,\mathbb{F}),$$

the numbers

$$B_i(\mathcal{C}(K,\mathbb{F})), \quad B_e(\mathcal{C}(K,\mathbb{F})), \quad B_o(\mathcal{C}(K,\mathbb{F})), \quad B(\mathcal{C}(K,\mathbb{F})).$$

We observe that if K is a simplicial complex, and contains the empty simplex, then $C(K,\mathbb{F})$ is the usual reduced chain complex. If K is a simplicial complex except that K does not contain the empty simplex, then $C(K,\mathbb{F})$ is the usual (nonreduced) chain complex. If $\dot{K} = \bar{K} - K$ is a nonempty subcomplex (where $\bar{K}$ denotes the closure of K, i.e. K along with all of the faces of the elements in K, and $\dot{K}$ denotes the elements in $\bar{K}$ which are not in K), then $C(K,\mathbb{F})$ is isomorphic to the relative chain complex $C(\bar{K},\dot{K},\mathbb{F})$. In particular, in these cases, $C(K,\mathbb{F})$ is a differential complex, so, by Theorem 48, the Betti numbers we have defined are equal to the standard Betti numbers. This implies Theorem 37.

Example 49. Let S denote the two-dimensional simplex with vertices labeled a, b, c. Let K denote the simplex space consisting of the faces $[c]$, $[b,c]$ and $[a,b,c]$ (Figure 20). The chain complex for K can easily seen to be isomorphic to the complex examined in Example 46. From the calculations done there, we see that $B_i(K) = 0$ for each $i = -1, 0, 2, \ldots$, and $B_o(K) = 0$, while $B(K) = B_e(K) = 1$.

The next goal is to indicate that these Betti numbers allow us to apply the basic notions of discrete Morse theory to general simplex spaces. Let K be a simplex

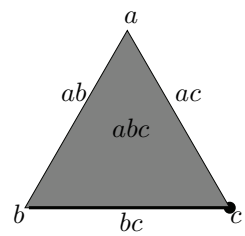


Figure 20. The complex K of example 49.

space. We define the basic combinatorial notions just as for a simplicial complex. A face a of S is said to be a *maximal element* of K if a is in K, and a is not a proper subset of any element in K. If a is a maximal element of K, we say that b is a *free face* of a in K if: b is in K, b is a codimension-one face of a, and a is the only element of K which properly contains b. Let K be a simplex space, a a maximal face of K, and b is a free face of a in K. The act of replacing K by $K - \{a, b\}$ is called a *simplicial collapse*. Say that K is *collapsible* if one can transform K into the empty simplex space by a sequence of simplical collapses.

Let K be a simplex space, and a a maximal element of K. We will call the act of replacing K by $K - a$ a *simplical removal*.

We will use the term *elementary simplicial reduction* to refer to either a simplical collapse or a simplicial removal. A *complete reduction* of K is any sequence of elementary reductions that transforms K into the empty simplex space. In particular, K is collapsible if and only if there is a complete reduction consisting solely of simplicial collapses.

Lemma 50. *Let K be a simplex space.*

(i) If $K' = K - \alpha$ for some maximal d-simplex α, then

$$B(K') \geq B(K) - 1.$$

(ii) If $K' = K - (Int(\alpha) \cup Int(\beta))$ is the result of a simplicial collapse, where α is a maximal d-simplex and β is a free face of α, then

$$B(K') \geq B(K).$$

Together, parts (i) and (ii) imply the following theorem.

Theorem 51. *Let K be a simplex space. In any complete reduction of K, the number of simplices which are taken out by a simplicial removal is at least $B(K)$.*

Corollary 52. *Let K be any simplex space, and V any gradient vector field on K. Then the number of critical cells of V is at least $B(K)$.*

Theorem 50 can be made more precise to include an understanding of how the individual $B_d(K)$ can change under simplicial collapse and simplicial removal that is sufficient to imply Theorem 39.

Example 53. We end this section with an example to illustrate that, unlike in the case of a simplicial complex, a simplicial collapse can increase the Betti numbers of a simplex space. Let S denote the two-dimensional simplex with vertices label a, b, c. Let K denote the simplex space consisting of the four faces $[a], [a, b], [b, c], [a, b, c]$. Then K is collapsible, since one can remove $[b, c]$ and $[a, b, c]$ by one simplicial

collapse, and the remaining two faces with a second simplicial collapse. Thus, all Betti numbers of K are zero. One the other hand, beginning with K, one can also remove the faces $[a, b]$ and $[a, b, c]$ by a simplicial collapse, resulting in the simplex space K' consisting of the faces $[a]$ and $[b, c]$. One can easily check that $B_0(K') = B_1(K') = B_e(K') = B_o(K') = 1$.

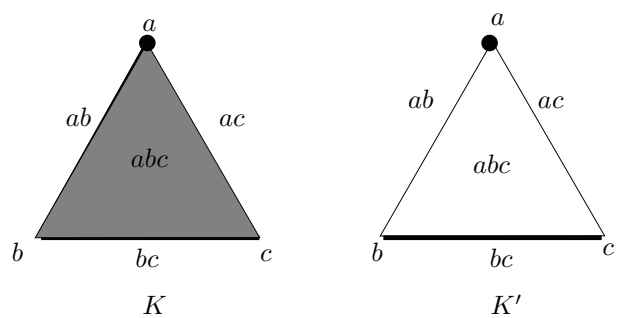


Figure 21. The simplex space K with vanishing Betti numbers collapses to K', which has nonzero Betti numbers.

Different algebraic extensions of discrete Morse theory appear in [**9**],[**52**] and [**81**]. These appoaches are quite similar in spirit to each other, and share some ideas with the work in this section.

3. Exercises for Lecture 3

(1) In Lecture 2 we constructed a perfect gradient field (i.e. one for which the Morse inequalities are equalities) on Δ_n^2, the complex of disconnected graphs on n vertices. Show that there is a decision tree which induces such a gradient vector field. (This observation is due to Jonsson, who found large classes of complexes which have perfect gradient vector fields which are induced by decision trees.)

(2) Consider the 2-dimensional simplicial complex on 5 vertices whose maximal simplices are [012], [023], [034], [045], [051], [123], [234], [345], [451], and [512]. Show that (i) this complex is collapsible and (ii) this complex is not nonevasive. (This example is due to Björner.)

(3) Returning to the first example in this lecture, prove directly from the definitions that the total Betti number of the set of bad outcomes is 2.

LECTURE 4
The Charney-Davis Conjectures

1. Introduction

These notes are intended to be an introduction to the Charney-Davis conjectures and some of their combinatorial implications. My aim is to provide a stimulating advertisement for a circle of ideas that is the subject of some fascinating recent work, most of which creates more questions than answers, and which has shed new light on some of the central questions in geometric combinatorics. The subject is a beautiful one, borrowing techniques and ideas from geometry, topology, analysis, algebra, algebraic geometry and combinatorics. My goal in these lectures is to present the topological and geometric context of these conjectures (as presented e.g. in [**15**],[**22**]), along with the most recent combinatorial understanding of them (as in Gal [**43**] and Brändén[**12**].) These notes will have been successful if some readers are inspired to consult these original sources, and to begin thinking about these conjectures.

The Charney-Davis conjectures, concerned with the relationship between geometry and topology, find their origins, as do most such questions, in the Gauss-Bonnet Theorem. Recall that the Gauss-Bonnet theorem states that if M is a compact surface with a Riemannian metric, then

$$\chi(M) = \frac{1}{2\pi}\int_M K \, d\text{area}$$

where K denotes the Gauss curvature of M. It follows that if $K \leq 0$ everywhere, then $\chi(M) \leq 0$.

Hopf conjectured the following generalization.

Conjecture 54. *If M is a compact Riemannian manifold of dimension $2n$ and the sectional curvature of M is ≤ 0 then $(-1)^n\chi(M) \geq 0$.*

[Recall that if M is an odd-dimensional manifold, then $\chi(M) = 0$.] This is not a suitable place for a primer in differential geometry, so we hope it will suffice to say that the condition that the sectional curvature is nonpositive means that every two-dimensional "orthogonal slice" of M is a surface of nonpositive Gauss curvature. This conjecture may seem a bit surprising, and perhaps unintuitive, at first glance. However, some general considerations point in this direction. Most notably, one has the following observations.

Proposition 55.

(1) *Let M_1 and M_2 be compact manifolds, then*

$$\chi(M_1 \times M_2) = \chi(M_1)\chi(M_2).$$

(2) *If M_1 and M_2 are Riemannian manifolds with nonpositive sectional curvature, and $M_1 \times M_2$ is endowed with the product Riemannian metric, then $M_1 \times M_2$ has nonpositive sectional curvature.*

Thus, if M_1 and M_2 are nonpositively curved Riemannian manifolds for which the conclusion of Hopf's conjecture holds, then the same is true for $M_1 \times M_2$. In particular, Hopf's conjecture holds for any product of arbitrarily many nonpositively curved surfaces.

Allendoerfer, Fenchel and Weil ([**1**],[**29**], [**2**]), and later Chern ([**17**]), proved a higher dimensional version of the Gauss-Bonnet theorem, which, for a compact Riemannian manifold, has the general form

$$\chi(M) = \int_M R \, dvol,$$

where R is a function of the curvature of M and is usually called the Chern-Gauss-Bonnet integrand. Chern [**18**] gives a proof (attributed to Milnor) that in dimension 4, if the sectional curvature is ≤ 0 everywhere, then $R \geq 0$. In particular:

Corollary 56. *If M is a compact Riemannian 4-manifold with sectional curvature ≤ 0, then $\chi(M) \geq 0$.*

However, Geroch [**44**] proved that this approach is insufficient to settle Hopf's conjecture in higher dimensions.

Theorem 57. *In even dimensions ≥ 6, there exist Riemannian metrics with sectional curvature ≤ 0 such that the Chern-Gauss-Bonnet integrand achieves both signs.*

So, in higher dimensions another approach is necessary. Before discussing alternate approaches, and partial results, we will take a detour to discuss some generalizations and extensions of Hopf's conjecture. From now on, when we say that a Riemannian manifold M has *nonpositive curvature,* we mean that all sectional curvatures are ≤ 0.

It is a theorem of Cartan and Hadamard that if M^n has nonpositive curvature, then $\widetilde{M}$, the universal cover of M, is diffeomorphic to $\mathbb{R}^n$. A manifold is said to be *aspherical* if its universal cover is contractible. Thurston generalized Hopf's conjecture to the following

Conjecture 58. *Let M^{2n} be a smooth, compact, aspherical manifold. Then*

$$(-1)^n \chi(M) \geq 0.$$

This is quite interesting, as the hypothesis has changed from a geometric condition to one that is purely topological. Our interests, however, lie in a different direction. Riemannian curvature is expressed in terms of 2nd derivatives of the metric. Thus, Hopf's conjecture, as it is usually understood, is a statement about manifolds which are at least twice differentiable. However, Alexandrov showed how one could speak of nonpositive curvature for continuous, but nonsmooth, manifolds. Let M be a complete Riemannian manifold. The Hopf-Rinow theorem states that for any two points p and q in M, there exists a minimal geodesic from p to q (i.e. a curve γ

from p to q satisfying length(γ) = distance(p,q)). Let x, y and z be three points in M, and $\widetilde{x}$, $\widetilde{y}$ and $\widetilde{z}$ be three points in $\mathbb{R}^2$ such that

$$d_M(x,y) = d_{\mathbb{R}^2}(\widetilde{x},\widetilde{y}), \quad d_M(x,z) = d_{\mathbb{R}^2}(\widetilde{x},\widetilde{z}), \quad \text{and} \quad d_M(y,z) = d_{\mathbb{R}^2}(\widetilde{y},\widetilde{z}).$$

(Such $\widetilde{x}$, $\widetilde{y}$ and $\widetilde{z}$ always exist.) Let γ be a minimal geodesic from y to z, and $\widetilde{\gamma}$ the straight line from $\widetilde{y}$ to $\widetilde{z}$. Since $|\gamma| = |\widetilde{\gamma}|$ there is a natural identification between points in γ and points in $\widetilde{\gamma}$. In [**3**] Alexandrov proves the following result.

Theorem 59.

(1) If M is simply-connected and has nonpositive curvature, then for any point p in γ, if $\tilde{p}$ is the corresponding point in $\tilde{\gamma}$, we have

$$d_M(x,p) \leq d_{\mathbb{R}^2}(\tilde{x},\tilde{p}).$$

(2) If M is not simply-connected then the above inequality is still true if one restricts to triples x, y and z which are sufficiently close to one another.

(3) The converse is also true: If the above inequality holds for all sufficiently close x, y and z then M has nonpositive curvature.

Alexandrov's theorem shows that the property of nonpositive curvature is equivalent to a property which can be expressed in terms of the distance function without reference to any derivatives. Hence we can use this theorem to make sense of the notion of nonpositive curvature for spaces without a smooth structure. By replacing $\mathbb{R}^2$ with other constant curvature surfaces, we can also make sense of the notion of having curvature bounded above by any real number. While a smooth structure is not necessary, this approach does require the existence of geodesics, so we must restrict attention to spaces for which this notion makes sense.

Definition 60. Let M be a metric space.

(1) Let x and y be two points in M and $d = dist(x,y)$. Then a *(minimal) geodesic* between x and y is an isometry $\gamma : [0,d] \to M$ with $\gamma(0) = x$ and $\gamma(d) = y$.

(2) M is said to be a *length space* (or a *geodesic space*) if every pair of points can be joined by a geodesic.

Motivated by Alexandrov's theorem, a length space M is said to be CAT(0) (C for Comparison or Cartan, A for Alexandrov, and T for Toponogov, who proved related comparison theorems) if the following condition holds: For any three points x, y and z in M, let $\widetilde{x}$, $\widetilde{y}$ and $\widetilde{z}$ be three points in $\mathbb{R}^2$ such that

$$d_M(x,y) = d_{\mathbb{R}^2}(\widetilde{x},\widetilde{y}), \quad d_M(x,z) = d_{\mathbb{R}^2}(\widetilde{x},\widetilde{z}), \quad \text{and} \quad d_M(y,z) = d_{\mathbb{R}^2}(\widetilde{y},\widetilde{z}).$$

Let γ be any minimal geodesic from y to z, and γ' the straight line from $\widetilde{y}$ to $\widetilde{z}$. Let p be any point in γ, and $\widetilde{p}$ the corresponding point in $\widetilde{\gamma}$. Then we require, for all choices of x, y, z, γ and p that

$$d_M(x,p) \leq d_{\mathbb{R}^2}(\widetilde{x},\widetilde{p}).$$

Some basic facts about CAT(0) spaces are left as exercises (see the end of this section).

Parts (2) and (3) of Theorem 59 lead to the following definition (which first appeared, along with many far-reaching implications and applications of this idea, in [**48**]).

Definition 61. Say a length space M is nonpositively curved (NP) if the CAT(0) inequality is true for all sufficiently close triples x, p, q.

The basic relationship between these notions is the following.

Theorem 62. *M is CAT (0) if and only if M is simply connected and nonpositively curved.*

We now specialize to a subclass of length spaces, namely polyhedra. A *polyhedron* is defined by a set $\{p_1, p_2, p_3, \dots\}$ of convex polytopes together with a collection of isometric identifications of some faces of the polytopes. Quotienting out by these identifications yields a topological space M which has a given cell decomposition. We require that in the resulting space, if two polytopes meet, they do so along a face of each. In these notes we also require that there be a uniform upper bound on the diameter of the polytopes, and that the resulting cell complex be locally finite.

It is useful now to introduce the notion of the link of a vertex. Let M be a polyhedron, and v a vertex of M. For any ϵ, let $S(v, \epsilon)$ denote the points in M which lie in a polytope incident to v and whose distance from v is precisely ϵ. For ϵ small enough, $S(v, \epsilon)$ is a topological space with an induced cell decomposition which is, up to isomorphism, independent of ϵ. This space, along with its cell decomposition, is defined to be the *link of* v, and is denoted by $link(v)$. If M has the property that the link of every vertex is a combinatorial sphere (i.e. it has a subdivision isomorphic to a subdivision of the boundary complex of a simplex), then M is a manifold. In this case we call M a *piecewise Euclidean manifold* (or PE manifold).

It will be important later that the link of each vertex comes also with a natural geometry. If ϵ is the radius of S, we can normalize the metric on $S \cap M$ by multiplying by $1/\epsilon$, so that the cells of $S \cap M$ are now convex cells from a sphere of radius 1. When we speak of distances and lengths on $link(v)$, it is always with respect to this spherical geometry.

For any polyhedron M there is a natural notion of arc-length for curves in M, induced by the Euclidean structure in each polytope. We can then define the distance between two points in M to be the infimum of the length of the curves connecting the two points.

Theorem 63. *Any polyhedron with arclength and distance defined as above is a length space.*

Thus, using Definition 61, we can speak of a polyhedron having nonpositive curvature. We can now state the first conjecture of Charney and Davis, which is a direct analogue of Hopf's conjecture.

Conjecture 64. *If M is a nonpositively curved compact PE manifold of dimension $2n$, then $(-1)^n \chi(M) \geq 0$.*

(This conjecture, and all of the other conjectures presented in this section first appeared in [**15**], and the reader should certainly consult that paper for a more complete discussion, as well as some initial evidence for the conjectures.) Since the first positive steps towards the Hopf conjecture were proved using the generalized Gauss-Bonnet theorem, it seems reasonable to begin our examination of the Charney-Davis conjecture with a similar approach. Let us begin with the case of a PE surface. Let M be a PE surface, and v a vertex of M. Let

$$k(v) = 1 - \sum_{f > v} angle(f, v), \tag{6}$$

where the sum is over all faces f which contain v, and $angle(f, v) \in [0, 1]$ denotes the normalized interior angle of f at v, i.e. the usual angle (in radians) divided by 2π. Then one can check in a straightforward manner the following very classical formula

$$\chi(M) = \sum_v k(v). \tag{7}$$

The relationship between the previous discussion and the current topic is provided by the following lemma.

Lemma 65. *A PE surface M is nonpositively curved if and only if $k(v) \leq 0$ for each vertex v.*

The Charney-Davis conjecture, in the case of PE surfaces, follows immediately.

This discussion was generalized to higher dimensions by Banchoff [**8**]. Let M be a polyhedron and v a vertex of M. Define

$$k(v) = \sum_{i=0}^{n} (-1)^i \sum_{\{\alpha^{(i)} > v\}} [v, \alpha] \tag{8}$$

where $\{\alpha^{(i)} > v\}$ denotes the set of i-dimensional cells of M which contain v, and $[v, \alpha]$ denotes the normalized exterior angle of α at v. That is, $[v, \alpha]$ is the fraction of the sphere consisting of outward pointing normals to supporting hyperplanes of α at v. Equivalently, $[v, \alpha]$ is the fraction of linear functions on α which achieve their maximum at v. Banchoff proved the following generalization of (7).

Theorem 66. *If M is a polyhedron, then*

$$\chi(M) = \sum_v k(v). \tag{9}$$

Recall that the local approach that was sufficient to prove Hopf's theorem in dimensions 2 and 4 is not sufficient in higher directions. Charney and Davis, perhaps somewhat surprisingly, conjecture that the corresponding local approach to their conjecture works in all dimensions.

Conjecture 67. *Let M be a PE manifold of dimension $2n$. If M is nonpositively curved, then for every vertex of M*

$$(-1)^n k(v) \geq 0.$$

The function $k(v)$ can, in a straightforward way, be written in terms of the link of v with its natural geometry as a complex of spherical cells. Let $\widetilde{k}$ denote this function, so that

$$k(v) = \widetilde{k}(\text{link}(v)).$$

The next step is to determine which simplicial complexes can arise as links of vertices in nonpositively curved PE manifolds. Roughly speaking, a Riemannian manifold has nonpositive curvature if and only if the boundary of each small metric ball is larger, in some sense, than the corresponding Euclidean sphere. Something similar is true for PE manifolds. That is, a PE manifold is nonpositively curved if the link is larger, in a precise sense, than a standard sphere of radius 1. More precisely, say that a complex L of spherical cells is *large* if for every pair of points x and y in L, with $dist(x, y) < \pi$, there is a unique geodesic connecting them. The following is a theorem of Gromov [**48**].

Theorem 68. *A PE manifold M is nonpositively curved if and only if the link of every vertex is large.*

The strongest version (from this point of view) of the Conjecture 67 can now be stated.

Conjecture 69. *Let L be a spherical complex (i.e. a cell complex in which each cell has the geometry of a convex cell in a sphere of radius 1) which is homeomorphic to a sphere of dimension $2n-1$. If L is large, then*

$$(-1)^n \widetilde{k}(L) \geq 0.$$

Note that this is a more general conjecture that what is needed, as our original setting required consideration only of combinatorial spheres (a more restrictive class than all topological spheres). However, it does not seem constructive at this point to quibble over such distinctions, since it is not at all clear what the right context is for this conjecture. Moreover, in just a moment we will weaken this hypothesis even further.

For the remainder of these notes, we will restrict attention to the case in which M is a cubical complex (i.e. all of the polytopal cells in M are geometric cubes). In this case, the links have a very special structure, namely every edge of each of the spherical simplices has length $\pi/2$. One can easily see that for each $\alpha^{(i)} > v$, we have $[v, \alpha] = (1/2)^i$. Since each such α corresponds to an $(i-1)$ simplex in the link of v, we have the formula

$$\widetilde{k}(L) = 1 + \sum_{i=0}^{\dim(M)-1} \left(\frac{-1}{2}\right)^{i+1} f_i(L), \tag{10}$$

where $f_i(L)$ denotes the number of i-simplices in L. Gromov showed that there is a simple combinatorial test for whether such a link is large.

Definition 70. Say that a simplicial complex L is *flag* if every clique spans a simplex. That is, is $v_1, v_2, \ldots, v_k$ are vertices in L, and they are all pairwise adjacent, then they span a simplex.

Theorem 71. *A cubical PE manifold is nonpositively curved if and only if the link of every vertex is flag.*

Thus, in this case, Conjecture 69 implies the following statement.

Conjecture 72. *Let L be a simplicial complex which is homeomorphic to a sphere of dimension $2n-1$. If L is flag, then $(-1)^n \widetilde{k}(L) \geq 0$, where $\widetilde{k}(L)$ is given by the formula (10*

This conjecture is very combinatorial in nature, but still has one topological, noncombinatorial, ingredient, namely the hypothesis that L be homeomorphic to a sphere. There is a natural generalization of triangulated spheres which has a more combinatorial flavor. A *Gorenstein** complex (or *a generalized homology sphere*) is a simplicial complex with the property that, for every $p \geq 0$, the link of every p-simplex has the homology of an $(n-p-1)$-sphere. If L is a simplicial complex which is homeomorphic to an n-sphere, or, more generally, any homology n-sphere, then L is Gorenstein*. Thus, to place these ideas in a more combinatorial setting, it is natural to consider the following generalization of Conjecture 72.

Conjecture 73. *Let L be a $(2n-1)$-dimensional Gorenstein* complex. If, in addition, L is flag, then $(-1)^n\widetilde{k}(L) \geq 0$, where $\widetilde{k}(L)$ is given by the formula (10).*

It is not at all surprising that this conjecture, which is stated completely in combinatorial terms, has received the most attention from the combinatorics community. We will discuss the connections with other combinatorial notions in Section 3. For now we note that this conjecture, or Conjecture 72, implies Conjecture 64 for cubical complexes. It is a simple, but still rather surprising, fact that the converse is true. The following result is due to Babson-Billera-Chan [**5**] and Bridson-Haefliger [**13**].

Theorem 74. *Let L be any simplicial complex. Then there is a finite cubical polyhedron M with the property that the link of every vertex of M is isomorphic to L.*

Proof. Let V denote the vertex set of L. Consider the cube $C = [0,1]^V \subset \mathbb{R}^V$ endowed with its standard cubical cell decomposition. We will find M as a subcomplex of C. For every simplex σ of L, let $\mathbb{R}^\sigma$ denote the linear subspace of $\mathbb{R}^V$ traced out by varying the coordinates corresponding to vertices in σ. Let M be the union of all faces of C which are parallel to some $\mathbb{R}^\sigma$, for some simplex σ of L. Then the vertices of M are the vertices of C, and the link of every vertex is isomorphic to L. □

Applying this result to the case when L is a combinatorial sphere yields a cubical PE manifold M. If, in addition, L is flag, it follows from Theorem 71 that M is nonpositively curved.

Corollary 75. *The Charney-Davis Conjecture 64 is true for cubical nonpositively curved PE manifolds if and only if Conjecture 72 is true for combinatorial spheres.*

2. Exercises for Lecture 4

(1) Prove that if M is a CAT(0) space then
   (a) For any points p and $q \in M$, there is a unique minimal geodesic from p to q.
   (b) M is contractible.
(2) Prove the formula (7) for any triangulated surface M.
(3) Prove the formula (9) for any finite polyhedron M.
(4) Show that the formula (8) specializes to (6) in the case of a triangulated surface.
(5) (a) Show that the barycentric subdivision of any polyhedron (in fact any CW complex) is flag.
   (b) Show that the join of any two flag simplicial complexes is flag.
   (c) Show that if L is a flag simplicial complex and v is any vertex of L, then $\mathrm{link}(v)$ and $\mathrm{star}(v)$ are both flag.
(6) Let L_1 and L_2 be simplicial spheres of dimension $2n_1 - 1$ and $2n_2 - 1$, respectively, which satisfy the conjectured inequality $(-1)^{n_i}\widetilde{k}(L_i) \geq 0$ and are flag. Show that $L_1 * L_2$, the join of L_1 and L_2, also satisfies the conjectured inequality.

LECTURE 5
From Analysis to Combinatorics

1. Hodge Theory and the Hopf-Charney-Davis Conjectures

In this section we present an overview of some of the beautiful analytical ideas that have been used to study the Hopf-Charney-Davis conjectures. Hodge theory is one of the standard ways of investigating the homological implications of geometric information, so it should not be too surprising that it has played a central role in this subject. To date, the most substantial partial results towards the Hopf and Charney-Davis conjectures have been proved using some variation of the Hodge theoretic approach we present here. To avoid some technical details, we will present the ideas in the combinatorial category. However, everything in this section can, with suitable care, be applied in the Riemannian setting.

Let X be a finite CW complex, and let $C^p(X)$ denote the space of real-valued p-cochains on X. Let

$$d_p : C^p(X) \to C^{p+1}(X)$$

denote the usual coboundary operator. Then $d^2 = 0$, and the singular cohomology of X, $H^*(X, \mathbb{R})$, is isomorphic to the cohomology of the cochain complex

$$C^*(X) : 0 \longrightarrow C^0(X) \xrightarrow{d_0} C^1(X) \xrightarrow{d_1} C^2(X) \longrightarrow \cdots$$

That is

$$H^p(X, \mathbb{R}) \cong \frac{\operatorname{Ker} d_p}{\operatorname{Im} d_{p-1}}.$$

Now endow each $C^p(X)$ with a (positive definite) inner product by declaring the canonical basis to be orthonormal. More explicitly, for each p, let $S_p(X)$ denote the set of p-cells in X. Choose an orientation for each element in $S_p(X)$. Then for α and β in $C^p(X)$ set

$$\langle \alpha, \beta \rangle = \sum_{y \in S_p(x)} \alpha(y)\beta(y). \tag{11}$$

Note that this inner product is independent of the chosen orientations.

Let d_p^* denote the adjoint of the operator d_p. That is,

$$d_p^* : C^{p+1}(X) \longrightarrow C^p(X)$$

is the unique map satisfying

$$\langle d_p\alpha, \beta\rangle = \langle \alpha, d_p^*\beta\rangle$$

for every p-cochain α and $(p+1)$-cochain β.

Define, for each p, the (combinatorial) Laplace operator

$$\Box^p := d_{p-1}d_{p-1}^* + d_p^*d_p : C^p(X) \longrightarrow C^p(X).$$

There is much that can be said about this fascinating operator but for us, the main point is the following.

Theorem 76. *For each p*

$$\mathrm{Ker}(\Box^p) \cong H^p(X).$$

Proof. Using basic linear algebra, and the fact that $\mathrm{Ker}(d_p) \supset \mathrm{Im}(d_{p-1})$, we have that

$$\begin{aligned} H^p(X) &\cong \frac{\mathrm{Ker}(d_p)}{\mathrm{Im}(d_{p-1})} \\ &\cong \mathrm{Ker}\, d_p \cap (\mathrm{Im}\, d_{p-1})^\perp \\ &\cong \mathrm{Ker}\, d_p \cap \mathrm{Ker}\, d_{p-1}^* \\ &= \mathrm{Ker}(\Box^p). \end{aligned}$$

The last equality follows from the observation that if $\alpha \in \mathrm{Ker}(\Box^p)$, then

$$0 = \langle \Box^p\alpha, \alpha\rangle = |d_{p-1}^*\alpha|^2 + |d_p\alpha|^2,$$

so that

$$\alpha \in \mathrm{Ker}\, d_p \cap \mathrm{Ker}\, d_{p-1}^*.$$

□

Cochains in the kernel of $\Box^p$ are called *harmonic*. We will denote the space of harmonic p-cochains in X by $\mathcal{H}^p(X)$.

So far, in this section, we have been considering the case of a finite CW complex. How do things change in the case of an infinite complex? Of particular interest to us is the case of the universal cover of a finite complex. With that in mind, let Y denote an infinite CW complex that is the covering space of some finite complex. Let us take a look at Hodge theory on Y. Let

$$C^*(Y) : 0 \longrightarrow C^0(Y) \xrightarrow{d_0} C^1(Y) \xrightarrow{d_1} C^2(Y) \longrightarrow \cdots$$

denote the cochain complex of Y. Hodge theory requires inner products. We quickly realize that the standard formula (11) does not yield a well-defined inner product in the infinite setting. There are various possible ways to proceed. However, if one desires to work with Hilbert spaces, there is a natural choice. Let $C_2^p(Y)$ denote the L^2 p -cochains on Y. That is, if $S_p(Y)$ denotes the set of p-cells in Y, each endowed with an orientation, then

$$C_2^p(Y) = \{\alpha \in C^p(Y) \text{ s.t. } \sum_{y\in S_p(Y)} (\alpha(y))^2 < \infty\}.$$

Then $C_2^p(Y)$ is a Hilbert space with respect to the inner product

$$\langle \alpha, \beta\rangle = \sum_{y\in S_p(y)} \alpha(y)\beta(y).$$

The next step is to replace the standard cochain complex on Y by the complex of L^2 cochains. To do this, one requires the following lemma.

Lemma 77. *$d_p(C_2^p(Y)) \subset C_2^{p+1}(Y)$.*

The proof is left as an exercise. (See Exercise 1 at the end of this lecture.)

Now consider the L^2 cochain complex

$$C_2^*(Y) : 0 \longrightarrow C_2^0(Y) \xrightarrow{d_0} C_2^1(Y) \xrightarrow{d_1} C_2^2(Y) \longrightarrow \cdots .$$

One might wish to proceed by defining the L^2-cohomology of Y by the usual formula

$$\frac{\mathrm{Ker}(d_p : C_2^p(Y) \to C_2^{p+1}(Y))}{\mathrm{Im}(d_{p-1} : C_2^{p-1}(Y) \to C_2^p(Y))}$$

This is certainly possible (this is called the unreduced L^2 cohomology). However, it does lead to certain difficulties, since $\mathrm{Im}\, d_{p-1}$ need not be a closed subspace of $\mathrm{Ker}\, d_p$, and hence the quotient need not inherit the structure of a Hilbert space. With that in mind, we define the L^2 cohomology of Y to be

$$H_2^p(Y) := \frac{\mathrm{Ker}(d_p)}{\overline{\mathrm{Im}}(d_{p-1})}$$

where $\overline{\mathrm{Im}}(d_{p-1})$ indicates that we take the closure of $\mathrm{Im}(d_{p-1})$ in $\mathrm{Ker}(d_p)$. (This quotient is sometimes called the *reduced* L^2-cohomology.)

Following the same procedure as before, we can construct the adjoint operator $d^* C_2^p(Y) \to C_2^{p-1}(Y)$ (see exercise 1 at the end of this section), and the Laplace operator

$$\Box_2^p : C_2^p(Y) \longrightarrow C_2^p(Y).$$

Let $\mathcal{H}_2^p(Y)$ denote the kernel of the operator $\Box_2^p$. The proof of Theorem 76, applied in this setting, yields the following result.

Theorem 78. *$\mathcal{H}_2^p(Y) \cong H_2^p(Y)$.*

Now let X be a finite CW complex. We define the pth Betti number of X to be the dimension of $H^p(X)$, and denote this number by $b^p(X)$. From Theorem 76 we know that

$$b^p(X) = \dim \mathcal{H}^p(X).$$

Let $\pi^p : C^p(x) \to \mathcal{H}^p(x)$ denote the orthogonal projection. Then we can also write

$$b^p(X) = \mathrm{trace}(\pi^p).$$

A useful way to calculate the trace of an operator is to express the operator as a matrix with respect to some basis, and then to take the sum of the diagonal elements of the matrix. Let us carry out this procedure here, and represent π^p as a matrix with respect to the standard basis for the cochain space. Let $\alpha_1, \ldots, \alpha_k$ denote an orthogonal basis for $\mathcal{H}^p(X)$ (so that $k = b^p(X)$). Then we can write

$$\pi^p = \sum_{i=1}^{k} \alpha_i \otimes \alpha_i^*$$

where $\alpha_i^* : C^p(X) \to \mathbb{R}$ is the map that takes β to $\langle \beta, \alpha_i \rangle$. The function

$$K^p : S_p(X) \times S_p(X) \longrightarrow \mathbb{R}$$

given by

$$K^p(x,y) = \sum_{i=1}^{k} \alpha_i(x)\alpha_i(y)$$

is a matrix for the operator π^p, in the sense that for any $\beta \in C^p(X)$, and any $x \in S_p(X)$,

$$\begin{aligned}[\pi^p(\beta)](x) &= \sum_{i=1}^{k} \alpha_i(x)\langle \beta, \alpha_i\rangle = \sum_{i=1}^{k} \alpha_i(x) \left(\sum_{y\in S_p(X)} \alpha_i(y)\beta(y) \right) \\ &= \sum_{y} \left(\sum_{i=1}^{k} \alpha_i(x)\alpha_i(y) \right) \beta(y) = \sum_{y} K^p(x,y)\beta(y).\end{aligned}$$

It follows that

$$b^p(X) = \operatorname{trace}(\pi^p) = \sum_{x\in S_p(X)} K^p(x,x)$$

(This identity can easily be proved directly, without the preceeding discussion.) The function

$$K^p(x,x) = \sum_{i=1}^{k} \alpha_i^2(x)$$

is sometimes called the pth *local Betti number* (as its integral gives the pth Betti number). We chose an orthonormal basis for the space of harmonic cochains on order to define this function, but one can easily check that it is independent of the choices.

Now let us consider again the case of an infinite CW complex Y which has an action by a group G, such that Y/G is finite. In this case, if $\mathcal{H}_2^p(Y) \neq 0$, then it is necessarily infinite-dimensional. Still, much of the previous discussion makes sense in this setting. That is, one can define the orthogonal projection

$$\pi^p : C_2^p(Y) \to \mathcal{H}_2^p(Y).$$

While the trace of this operator is not defined, we can still construct a kernel $K^p(x,y)$, $x,y \in S_p(Y)$, given by

$$K^p(x,y) = \sum_{i} \alpha_i(x)\alpha_i(y)$$

where $\{\alpha_i\}$ is an orthonormal basis for $\mathcal{H}^p(Y)$. Just as for the finite complex, we can restrict this operator to the diagonal and consider the pth local Betti number $K^p(x,x) = \sum_i \alpha_i^2(x)$, $x \in S_p(Y)$. Summing these entries, however, yields an infinite result. At this point, we use the extra information we have, namely the fact that everything is invariant under the action of the group G. Let $S_p^*(Y) \subset S_p(Y)$ denote a set of p-cells of Y containing exactly one p-cell from each G-orbit in $S_p(Y)$. Then $S_p^*(Y)$ is a finite set, and the values $K^p(x,x)$ for $x \in S_p^*(Y)$ determine $K^p(x,x)$ for all x.

With that in mind, define the *G-trace* of π^p, denoted by $\tau_G(\pi^p)$ to be the result of summing $K^p(x,x)$ over $x \in S_p^*(Y)$. That is

$$\tau_G(\pi^p) = \sum_{x\in S_p^*(Y)} K^p(x,x) \in [0,\infty).$$

(The G-trace, denoted by τ_G, of any G-equivariant operator on $C_2^p(Y)$ can be defined by the same procedure.) We will denote the number $\tau_G(\pi^p)$ by $b_G^p(Y)$, and call it the pth *L^2-Betti number of* Y. One simple, but essential observation, is that $b_G^p(Y) = 0$ if and only if $\mathcal{H}_2^p(Y) = 0$. The definition of $b_G^p(Y)$ may seem a bit *ad hoc*. However, the reader can consult [**4**] to see how this definition results from natural notions in the study of operator algebras. In that context, $b_G^p(Y)$ is the von Neumann trace of the operator $\pi^p : C_2^p(Y) \to \mathcal{H}_2^p(Y)$.

Now let us restrict attention to the case in which $\widetilde{X}$ is the universal cover of a finite CW complex X, and we take the group G to be the fundamental group of X, acting freely on $\widetilde{X}$ in the usual way. In this setting, Dodziuk proved that the L^2-Betti numbers of $\widetilde{X}$ computed combinatorially from a cell decomposition are equal to those calculated from the Riemannian Laplacian, and that these numbers are homotopy invariants of X [**23**]. For our purposes, the main property of the L^2-Betti numbers of $\widetilde{X}$ is the following result of Atiyah [**4**].

Theorem 79. *Let $\widetilde{X}$ be the universal cover of a finite CW complex X, and take the group G to be the fundamental group of X. Then*

$$\chi(X) = \sum_i (-1)^i b_G^p(\widetilde{X}). \tag{12}$$

Thus, L^2-Betti numbers are another tool at our disposal for investigating the Euler characteristic. It may not be clear how one could use this new information to investigate the Hopf-Charney-Davis conjectures, but a link is provided by the following beautiful conjecture of Singer.

Conjecture 80. *Let X be a compact aspherical n-manifold, and $\widetilde{X}$ the universal cover of X. Then for all $p \neq n/2$, $\mathcal{H}_2^p(\widetilde{X}) = 0$.*

Applying Theorem 79, we see that Singer's conjecture immediately implies the Hopf conjecture (as generalized by Thurston, Conjecture 58). While the Hopf conjecture is trivial for odd dimensional manifolds, Singer's conjecture is not. Singer's conjecture can quite easily be shown to be true for surfaces (it follows from the fact that there are no L^2 harmonic functions on a complete Riemannian manifold of infinite volume). It has also been shown to hold for 3-manifolds for which Thurston's geometrization conjecture is true [**66**], locally symmetric spaces [**24**], negatively curved Kähler manifolds [**49**], manifolds with sufficiently pinched negative curvature [**26**], aspherical manifolds whose fundamental group contains an infinite amenable normal subgroup [**16**], and manifolds which fiber over S^1 [**67**].

It is quite natural to guess that Singer's conjecture also holds for suitable piecewise Euclidean manifolds. The following conjecture, along with a series of related conjectures, appears in Section 8 of [**22**].

Conjecture 81. *Let X be a compact nonpositively curved PE manifold of dimension n, and let $\widetilde{X}$ be the universal cover of X. Then for all $p \neq n/2$, $\mathcal{H}_2^p(\widetilde{X}) = 0$.*

This conjecture implies the Charney-Davis conjecture 64. In [**22**], Davis and Okun used this circle of ideas to establish Conjecture 72 for 3-dimensional flag simplicial spheres (and somewhat more generally).

Theorem 82. *Let L be any flag simplicial 3-sphere, then $\widetilde{k}(L) \geq 0$, where $\widetilde{k}$ is as in (10).*

Very roughly speaking, for any flag simplicial 3-sphere, a special nonpositively curved 4-dimensional cubical PE manifold M is constructed (using the structure of right-angled Coxeter groups) which has the properties that the link of every vertex is identified with L, and Singer's conjecture can be shown to hold for M. This is a wonderful result, requiring a lot of hard work, and Davis and Okun introduce some powerful new ideas into the subject. The reader is strongly encouraged to consult their paper. Following the discussion in Section 1, Theorem 82 implies the following general result.

Theorem 83. *If X is any finite nonpositively curved cubical PE manifold of dimension 4, then $\chi(X) \geq 0$.*

2. The Charney-Davis Conjecture and the h-vector

In this section we focus attention on Conjectures 72 and 73, and show that they reside quite naturally in the well-developed circle of ideas surrounding the investigation of f-vectors of simplicial complexes. In (10) the formula for the relevant function $\widetilde{k}(L)$ is given in terms of the f-numbers of L. In a number of settings, especially those related to commutative algebra and toric varieties, it has proved very useful to study certain special linear combinations of the f-numbers, called the h-numbers. In [**15**] Charney and Davis observe that Conjectures 72 and 73 can be restated in a very nice way in terms of the h-numbers. For any finite n-dimensional simplicial complex K, define the f-polynomial of K to be the generating function of the f-numbers. More explicitly, set $f(K,t) = \sum_{i=0}^{n+1} f_{i-1}(K)t^i$,where we define f_{-1} to be 1. Define the h-polynomial of K, $h(K,t) = \sum_{i=0}^{n+1} h_i(K)t^i$, by the formula

$$h(K,t) = (1-t)^{n+1} f(K, \frac{t}{t-1}). \tag{13}$$

(We will often write $h(t)$ for $h(K,t)$ if it will not cause any confusion.) It follows immediately from (10) and (13) that for any n-dimensional simplicial complex K,

$$h(K,-1) = 2^{n+1}\widetilde{k}(K).$$

Hence, we can now restate Conjecture 73 as

Conjecture 84. *If K is any simplicial Gorenstein* $(2n-1)$-complex which is flag, then $(-1)^n h(K,-1) \geq 0$.*

In this form, the conjecture can more easily be compared to other conjectures and results concerning the h-vectors of simplicial spheres and related spaces. One advantage of the h-polynomial is that it is quite easy to state the Dehn-Somerville relations. Say that K is *Eulerian* if the link of every i-simplex, $i \geq -1$, has the same Euler characteristic as a sphere of dimension $n-i-1$, so that, in particular, every Gorenstein* complex, and thus every triangulated sphere, is Eulerian. If K is Eulerian, then

$$h(t) = t^{n+1}h(t^{-1}), \tag{14}$$

(equivalently, $h_i = h_{n+1-i}$ for each i).

An n-dimensional simplicial complex is said to be *Cohen-Macaulay* if for every i the link of every i-simplex has nonzero reduced homology only in dimension $n-i-1$. So, for example, every Gorenstein* complex, and thus every triangulated sphere,

is Cohen-Macaulay. It can be shown using algebraic methods that if K is Cohen-Macaulay, then all of the coefficients of the h-polynomial are nonnegative (see Corollary II.3.2 of [**90**]).

To every rational polytope P that is simple (i.e. its boundary complex is dual to a simplicial complex – and every such polytope can be perturbed slightly to be rational), one can associate a toric variety X_P in a natural way (e.g. see [**42**] [**19**]). Danilov [**19**] showed that for each i, $b_{2i}(X_P)$, the $2i^{th}$ Betti number of X_P is equal to $h_i(P)$. This, and related indentities, has proved to be an immensely powerful source of information about the h-polynomial, as well as a new inspiration for the study of toric varieties.

For example, in [**61**] Reiner and Leung proved that if K is the boundary complex of a simple $2n$-polytope, then $h(K,-1)$ is equal to the signature of an associated toric variety. They were able to then show, using the Hirzebruch signature formula that if such a complex K satisfies a certain local convexity property (which is stronger than flag) then $(-1)^n h(K,-1) \geq 0$. Probably the most striking application of toric techniques is Stanley's proof of Theorem 87, stated below.

From another direction, special tools are available for simplicial complexes which arise as the order complex of a poset. For example, the following result is due to Babson.

Theorem 85. *If K is the the boundary complex of a simplicial $2n$-polytope, and is the order complex of a poset, then $(-1)^n h(K,-1) \geq 0$.*

Note that (i) the order complex of a poset is always flag and (ii) the barycentric subdivision of any simplicial complex is the order complex of a poset. Therefore this theorem implies the Charney-Davis conjecture for the barycentric subdivision of the boundary complex of any $2n$-dimensional simplicial polytope.

We can give an outline of the proof. For Eulerian n-dimensional order complexes, it has proved very useful to introduce a refinement of the h-vector, known as the cd-index, a homogeneous polynomial of degree n, $\Phi_K(c,d)$ in two noncommuting variables (where d is considered to have degree 2). For any $(2n-1)$-dimensional Eulerian complex K which is the order complex of a poset, Babson observed that we have the relationship

$$\Phi_K(0,-2) = h(K,-1),$$

and $\Phi_K(0,-2)$ is equal to $(-2)^n$ times the coefficient of $d^n$of a polytope, then all coefficients of the cd-index are nonnegative [**89**], so the result follows.

Stanley has made the following conjecture.

Conjecture 86. *If K is a Gorenstein* order complex, then every coefficient of the cd-index is nonnegative.*

By Babson's argument, this would imply Conjecture 73 for any Gorenstein* complex which is the order complex of a poset. We note that the barycentric subdivision of any Gorenstein* complex is Gorenstein* and is the order complex of a poset.[1]

The Charney-Davis conjecture is related to some of the central conjectures in the subject. For example, consider the Generalized Lower Bound Theorem (GLBT) of Stanley (proved with toric methods [**88**], and then later reproved by McMullen using only ingredients from convex geometry [**69**][**70**]) .

[1]After this lecture was delivered, the preprint by K. Karu, *The cd-index of fans and lattices*, math.AG/0410513, appeared in which Karu claims to prove this conjecture. See also his follow-up preprint *Lefschetz decomposition and the cd-index of fans*, math.AG/0509220.

Theorem 87. *If K is an n-dimensional simplicial complex which is the boundary of an $(n+1)$-polytope, then the h-numbers of K are unimodal. That is*

$$h_0(K) \leq h_1(K) \leq \cdots \leq h_{\lfloor \frac{n+1}{2} \rfloor}(K).$$

These are the most general linear inequalities satisfied by the f-vectors of simplicial polytopes. One of the central open problems in the study of f-vectors is to determine precisely for which simplicial complexes the conclusion of the GLBT holds. For example, does it hold for all Gorenstein* complexes (this is sometimes called the Generalized Lower Bound Conjecture), or all triangulated spheres, or all PL spheres? Independently, Kalai and Stanley have shown that the conclusion holds for the boundary complex of any triangulated $(n+1)$-ball which appears as a subcomplex of a simplicial $(n+1)$-polytope, but it is not clear which spheres arise in this way. This unimodality property has some relation to the Charney-Davis conjecture.

Theorem 88. *Let K be a Gorenstein* complex of dimension $2n-1$. Suppose that $h(K,t)$ only real roots. Then the following two conclusions hold.*

(1) the h-numbers of K are unimodal, i.e. the conclusion of the GLBT holds;
(2) $(-1)^n h(K,-1) \geq 0$.

While we have stated this result in terms of h-polynomials of simplicial complexes, this theorem is really just a statement about polynomials with real coefficients satisfying a symmetry relation as in (14). Part two of this theorem is due to Charney and Davis (see Lemma 7.5 of [**15**]). In fact, they prove the stronger statement that the conclusion holds as long as $h(K,t)$ has no nonreal roots of modulus 1. The first part of this theorem is due to Isaac Newton!

With this theorem in mind, it is natural to make the following *Real Root Conjecture* (apparently due originally to Januzkiewicz, see [**20**]).

Conjecture 89. *For any Gorenstein* complex K which is flag, $h(K,t)$ has only real roots.*

In [**75**], Reiner and Welker consider these questions for the order complex of a graded poset P. This special case of the real root conjecture was formulated earlier, and is known as the Neggers-Stanley conjecture. Without proving the Neggers-Stanley conjecture, they are able to prove the implications of this conjecture. More precisely, they construct a simplicial polytope with the same h-polynomial as the order complex of P, and thus the unimodality of the h numbers of the order complex follows from Theorem 87. By other means (using the results of [**61**]) they establish the Charney-Davis conjecture for K_P for graded posets of width 2.

More recently, Brändén [**12**] has proved the Charney-Davis conjecture for K_P, as well as the unimodality of the h-numbers, for *any* graded poset P. That is, he establishes both conclusions of Theorem 88 for such complexes. He does not do this by proving that the h-polynomial has real roots, however. Let us take a moment to discuss Brändén's approach, an approach that was also presented, independently, in the recent work of Gal [**43**]. We know from the Dehn-Sommerville relations (14) that the h-polynomial of any Eulerian complex of dimension $2n-1$ satisfies the symmetry

$$h(t) = t^{2n} h(t^{-1}).$$

The polynomials

$$p_i(t) = t^i(1+t)^{2n-2i}, \quad i = 0, 1, 2, \ldots, n$$

form a basis for the vector space of polynomials of degree $2n$ with this symmetry. Thus, for any Eulerian complex K of dimension $2n$ we can write

$$h(K,t) = \sum_{i=0}^{n} a_i(K) p_i(t). \tag{15}$$

and the $a_i(K)$'s are uniquely determined by this identity. We can make two simple observations. First, we see that

$$1 = h(K,0) = a_0(K).$$

Second, we see that

$$h(K,-1) = (-1)^n a_n(K). \tag{16}$$

Both Brändén and Gal make the following observation.

Theorem 90. *Let K be an Eulerian complex of dimension $2n-1$. Suppose that $a_i(K) \geq 0$ for each $i = 0, 1, \ldots, n$. Then*

(i) the h-numbers of K are unimodal (i.e. the conclusion of the GLBT holds), and

(ii) $(-1)^n h(K,-1) \geq 0$.

Note that (i) follows immediately from (15) and the observation that the coefficients of each of the p_i's is unimodal, and (ii) from (16).

In [**12**], Brändén establishes the nonnegativity of the $a_i(K)$'s for the order complex of any graded poset, and hence he established the unimodality of the h-polynomial and the Charney-Davis conjecture for such complexes. In [**43**] Gal filled in more of the picture. In particular, he proved

Theorem 91. *(i) The real root conjecture is true for spheres up to dimension 4. (This follows from the results of Davis and Okun.)*

(ii) The real root conjecture is false in all dimensions ≥ 5, and counterexamples are found among boundary complexes of simplicial polytopes.

This certainly seems to put an end to the idea of using Theorem 88 to prove both the Charney-Davis conjecture, and the Generalized Lower Bound Conjectures. In [**43**], Gal turns his attention to Theorem 90 and (in a slightly different notation) makes the following conjecture.

Conjecture 92. *If K is a Gorenstein* complex which is flag, then for all i, $a_i(K) \geq 0$.*

From Theorem 90 we see that this conjecture implies Conjecture 73. Moreover, Gal shows that his conjecture is weaker than the real root conjecture.

Theorem 93. *Let K be a Gorenstein* complex of dimension $2n-1$. Suppose that $h(K,t)$ has only real roots. Then for all i, $a_i(K) \geq 0$.*

Proof. Let $\gamma(K,t)$ denote the generating function of the a_i's. That is

$$\gamma(K,t) = \sum_{i=0}^{n} a_i(K)t^i.$$

One can easily check the following relation.

$$(1+t)^{2n}\gamma(K, \frac{t}{(1+t)^2}) = h(K,t). \tag{17}$$

Assume that all of the roots of $h(K,t)$ are real. Since the coefficients of $h(K,t)$ are all nonnegative and $h(K,0) = 1$, the roots must all be negative. It follows from (17) that the roots of $\gamma(K,t)$ occur either when $t = -1$ or at places when $t/(1+t)^2$ is real and negative, which implies that t is real and negative. Since the roots are real, we can apply Isaac Newton's result Theorem 88 (1) to deduce that the $a_i(K)$'s are unimodal. We also know that $\gamma(K,0) = a_0(K) = 1$. Since all of the roots are negative, it follows that the coefficient $a_n(K)$ of the highest order term is ≥ 0, and now, by the unimodality, it follows that all of the coefficients are ≥ 0. $\square$

As evidence for Conjecture 92, Gal shows that for a flag Gorenstein* complex, the coefficient of t in γ is nonnegative. This is equivalent to the statement that a flag complex simplicial sphere of dimension n has at least $2(n+1)$ vertices (with equality only for the boundary complex of a cross polytope). The other coefficients remain quite mysterious. Much work remains to be done, to discover the geometric/topological meaning behind these numbers, and to begin to assess the truth of this new conjecture. At present, based on the evidence presented here, there is every reason to believe that with Conjecture 92 we have found the right formulation of the problem.

3. Exercises for Lecture 5

Exercises for Section 1 of Lecture 5.

(1) Let Y be an infinite CW complex which is a covering space of a finite CW complex.
   (a) Show that $d(C_2^p(Y)) \subset C_2^{p+1}(Y)$, and, moreover, that d is a bounded operator. That is, there is a constant c so that for any $\alpha \in C_2^p(Y)$, $|d\alpha| \leq c|d\alpha|$.
   (b) Prove that one can define an adjoint operator $d^* : C_2^{p+1}(Y) \to C_2^p(Y)$, and that this operator is also bounded.

(2) Let X be a finite polyhedron. Let c_d denote the number of d-cells for each d. The goal in this problem is to prove that

$$\sum_p (-1)^p \dim \mathcal{H}^p(M) = \sum_p (-1)^p c_p.$$

Of course, using Theorem 76 this is just the standard formula for the Euler characteristic, but we have something else in mind. For $t \in [0,\infty)$, let

$$I(t) = \sum_p (-1)^p \mathrm{trace}(e^{-t\,\Box^p}),$$

where $\Box^p : C^p(X) \to C^p(X)$ is the Laplace operator, and $e^{-t\,\Box^p}$ is defined by a power series expansion.

[The operator $e^{-t\,\Box^p}$ is the unique solution to the differential equation on the space of operators on $C^p(X)$

$$\frac{\partial}{\partial t}L(t) = -\Box^p L(t), \;\; L(0) = I,$$

where I denotes the identity map on $C^p(X)$, and this characterization, can be used to give an alternate definition for $e^{-t\,\Box^p}$.]

(a) Show that $I(0) = \sum_p (-1)^p c_p$.
(b) Show that $\lim_{t\to\infty} I(t) = \sum_p (-1)^p \dim \mathcal{H}^p(M)$.
(c) Show that $d/dt\; I(t) = 0$ for all $t \in [0, \infty)$.

(3) Let us use the approach from exercise 2 to prove Theorem 79. In this case, let X be a finite polyhedron, Y the universal cover, and G the fundamental group of X. Define

$$I(t) = \sum_p (-1)^p \tau_G(e^{-t\,\Box^p}),$$

where $\Box^p : C_2^p(Y) \to C_2^p(Y)$ is the Laplace operator on Y. Show that

(a) $I(0)$ is the left hand side of (12) and $\lim_{t\to\infty} I(t)$ is the right hand side of (12).
(b) $d/dt\; I(t) = 0$ for all $t \in [0, \infty)$.

(4) Show that if Y is an infinite polyhedron, then $\mathcal{H}_2^0(\widetilde{Y}) = 0$.
(5) Let Y denote the real line given the structure of an infinite polyhedron by placing a vertex at each integer point. What is the (reduced) L^2-cohomology of Y? What is the unreduced L^2-cohomology of Y?

Exercises for Section 2 of Lecture 5.

(1) The best exercise is to calculate $f(t)$, $h(t)$, and $\gamma(t)$ for your favorite Eulerian complexes. Start with simple complexes, and then keep going.
(2) If you have never done this before: Prove the Dehn-Somerville relations (14) for any Eulerian complex.
(3) Prove identity (17).
(4) Find explicit formulas for the first few coefficients of $\gamma(K, t)$ in terms of the f-vector of K.
(5) Show that the coefficient of t in $\gamma(K, t)$ is always ≥ 0 for a Gorenstein* complex that is flag.

BIBLIOGRAPHY

1. C. Allendoerfer, *The Euler number of a Riemannian manifold*, Amer. J. Math., **62** (1940), pp. 243–248.
2. C. Allendoerfer and A. Weil, *The Gauss-Bonnet theorem for Riemannian polyhedra*, Trans. Amer. Math. Soc., **53** (1943), pp. 101–129.
3. A. Alexandrov, *A theorem on triangles in a metric space and some of its applications*, Trudy Math. Inst. Stekl., **38** (1951), pp. 5–23.
4. M. Atiyah, *Elliptic operators, discrete groups and von Neumann algebras*, Astérisque, **32-33** (1976), pp. 43–72.
5. E. Babson, L. Billera and C. Chan, *Neighborly cubical spheres and a cubical lower bound conjecture,* Israel J. Math., **102** (1997), pp. 297–315.
6. E. Babson, A. Björner, S. Linusson, J. Shareshian and V. Welker, *Complexes of not i-connected graphs*, Topology, **38** (1999), pp. 271–299.
7. E. Babson, P. Hersh, *Discrete Morse functions from lexicographic orders*, Trans. Amer. Math. Soc., **357** (2005), pp. 509–534.
8. T. Banchoff, *Critical points and curvature for embedded polyhedra,* J. Differential Geom., **1** (1967), pp. 245–356.
9. E. Batzies and V. Welker, *Discrete Morse theory for cellular resolutions*, J. Reine Angew. Math., **543** (2002), pp. 147–168.
10. L. Billera, S. Holmes and K. Vogtmann, *Geometry of the space of phylogenic trees*, Adv. in Appl. Math., **29** (2002), pp. 733–767.
11. A. Björner, *Topological methods*, in Handbook of combinatorics, vol. 2, R. Graham, M. Grötschel, and L. Lovász eds., North-Holland/Elsevier, Amsterdam, 1995, pp. 1819–1872.
12. P. Brändén, *Sign-graded posets, unimodality of W-polynomials and the Charney-Davis conjecture*, Electron. J. Combin., **110** (2005), pp. 127–145.
13. M. Bridson and A. Haefliger, *Metric spaces of nonpositive curvature*, Grundlehren der Math. Wiss. 319, Springer-Verlag, Berlin, 1999.
14. M. Chari, *On discrete Morse functions and combinatorial decompositions*, Discrete Math, **217** (2000), pp. 101–113.
15. R. Charney and M. Davis, *The Euler characteristic of a nonpositively curved, piecewise Euclidean manifold*, Pacific J. Math., **171** (1995), pp. 117–137.
16. J. Cheeger and M. Gromov, *L_2-cohomology and group cohomology*, Topology, **25** (1986), pp. 189–215.

17. S. Chern, *A simple intrinsic proof of the Gauss-Bonnet formula for closed Riemannian manifolds,* Ann. of Math. (2), **45**(1944), pp. 747–752.
18. S. Chern, *On curvature and characteristic classes of a Riemannian manifold*, Abh. Math. Sem. Univ. Hamburg, **20** (1956), pp. 117–126.
19. V. Danilov, *The geometry of toric varieties*, Russian Math. Surveys, **33** (1978), pp. 97–154.
20. M. Davis, J. Dyrnara, T. Januszkiewicz and B. Okun, *Decompositions of Hecke-von Neumann modules and the L^2-cohomology of buildings*, preprint (2004).
21. M. Davis, *Nonpositive curvature and reflection groups*, in Handbook of Geometric Topology, R. Daverman, R. Sher (eds.), Elsevier, Amsterdam, 2002, pp. 373–422.
22. M. Davis and B. Okun, *Vanishing theorems and conjectures for the ℓ^2-homology of right-angled Coxeter groups*, Geometry and Topology, **5** (2001), pp. 7–74.
23. J. Dodziuk, *de Rham-Hodge theory for L^2-cohomology of infinite covers*, Topology, **16** (1977), pp. 157–165.
24. J. Dodziuk, *L^2-harmonic forms on rotationally symmetric Riemannian manifolds*, Proc. Amer. Math. Soc., **77** (1979), pp. 395–400.
25. X. Dong, *Topology of bounded degree graph complexes*, J. Algebra, **262** (2003), pp. 287–312.
26. H. Donnelly and F. Xavier, *On the differential form spectrum of negatively curved Riemannian manifolds*, Amer. J. of Math., **106** (1984), pp. 169–185.
27. A. Duval, *A combinatorial decomposition of simplicial complexes*, Israel J. of Math., **87** (1994), pp. 77–87.
28. B. Eckmann, *Introduction to l^2-methods in topology: reduced l^2-homology, harmonic chains, l^2-Betti numbers*, (notes prepared by Guido Mislin), Israel J. Math., **117** (2000), pp.183–219.
29. W. Fenchel, *On total curvature of Riemannian manifolds: I,* J. London Math. Soc., **15** (1940), pp. 15–22.
30. S. Fomin and N. Reading, *Root systems and generalized associahedra*, this volume.
31. R. Forman, *A discrete Morse theory for cell complexes*, in Geometry, topology & physics, S.T. Yau (ed.), Conf. Proc. Lecture Notes Geom. Topology, International Press, Cambridge, MA, 1995, pp. 112–125.
32. R. Forman, *Morse theory for cell complexes*, Adv. in Math., **134** (1998), pp. 90–145.
33. R. Forman, *Witten-Morse theory for cell complexes*, Topology, **37** (1998), pp. 945–979.
34. R. Forman, *Combinatorial vector fields and dynamical systems*, Math. Zeit., **228** (1998), pp. 629–681.
35. R. Forman, *Combinatorial differential topology and geometry*, in New Perspectives in Algebraic Combinatorics, Math. Sci. Res. Inst. Publ. 38, Cambridge Univ. Press, Cambridge, 1999, pp. 177–206.
36. R. Forman, *Morse theory and evasiveness*, Combinatorica, **20** (2000), pp. 498–504.
37. R. Forman, *Novikov-Morse theory for cell complexes*, Int. J. of Math., **4** (2002), pp. 333–368.

38. R. Forman, *The cohomology ring and discrete Morse theory*, Trans. of the Amer. Math. Soc., **354** (2002), pp. 5063–5085.
39. R. Forman, *A user's guide to discrete Morse theory*, Seminaire Lotharingien de Combinatoire (electronic), **48** (2002), article B48c.
40. R. Forman, *Some applications of combinatorial differential topology*, Graphs and patterns in mathematics and theoretical physics, Proc. Sympos. Pure Math., 73, Amer. Math. Soc., Providence, RI, 2005, pp. 281–313.
41. R. Forman, *A topological approach to the game of "20 Questions"*, preprint.
42. W. Fulton, *Introduction to toric varieties*, Annals of Mathematics Studies 131, Princeton University Press, 1993.
43. S. Gal, *Real Root Conjecture fails for five and higher dimensional spheres*, Discrete Comput. Geom., **34** (2005), pp. 269–284.
44. R. Geroch, *Positive sectional curvature does not imply positive Gauss–Bonnet integrand*, Proc. Amer. Math. Soc., **54** (1976), pp. 267–270.
45. L. Glaser, *Geometrical combinatorial topology*, Vol. 1, Van Nostrand Reinhold Mathematical Studies # 27, 1970.
46. M. Goresky and R. MacPherson, *Stratified Morse theory*, in Singularities, Part I (Arcata, CA, 1981), Proc. Sympos. Pure Math., 40, Amer. Math. Soc., R.I., 1983, pp. 517–533.
47. M. Goresky and R. MacPherson, *Stratified Morse theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3), 14, Springer Verlag, Berlin-New York, 1988.
48. M. Gromov, *Hyperbolic groups*, in Essays in group theory, (S.M. Gersten (ed.)), Math. Sci. Res. Inst. Publ. 8, Springer-Verlag, New York, 1987, pp. 75–263.
49. M. Gromov, *Kähler hyperbolicity and L^2-Hodge theory*, J. Diff. Geom., **33** (1991), pp. 263–292.
50. P. Hersh, *On optimizing discrete Morse functions*, Adv. in Appl. Math., **35** (2005), pp. 294–322.
51. P. Hersh and V. Welker, *Gröbner basis degree bounds on $Tor^{k[\Lambda]}(k,k)$. and discrete Morse theory for posets*, in Integer points in polyhedra - geometry, number theory, algebra, optimization, Contemp. Math., 374, Amer. Math. Soc., Providence, RI, (2005), pp, 101–138.
52. J. Jonsson, *On the topology of simplicial complexes related to 3-connected and hamiltonian graphs*, J. Combin. Theory, Ser. A., **104** (2003), pp. 169–199.
53. J. Jonsson, *Optimal decision trees for simplicial complexes*, Electron. J. Combin., **12** (2005), pp. 211–239.
54. M. Joswig and M. Pfetsch, *Computing optimal discrete Morse functions*, in Workshop on Graphs and Combinatorial Optimization (electronic), Electron. Notes Discrete Math., 17. Elsevier, Amsterdam, (2004), pp. 191–195.
55. J. Kahn, M. Saks and D. Sturtevant, *A topological approach to evasiveness*, Combinatorica, **4** (1984), pp. 297–306.
56. D. Kleitman and D. Kwiatkowski, *Further results on the Aanderaa-Rosenberg conjecture*, J. Combin. Theory, Ser. B, **28** (1980), pp. 85–95.
57. D. Kozlov, *Collapsibility of $\Delta(\pi_n)/S_n$ and some related CW complexes*, Proc. of the Amer. Math. Soc., **128** (2000), pp. 2253–2259.
58. D. Kozlov, *Topology of spaces of hyperbolic polynomials and combinatorics of resonances*, Israel J. of Math., **132** (2002), 189–206.

59. W. Kühnel, *Triangulations of manifolds with few vertices*, in Advances in Differential Geometry and Topology, World Sci. Publishing, N.J., 1990, pp. 59–114.
60. C. Lee, *The associahedron and triangulations of the n-gon*, Europ. J. of Comb., **10** (1989), pp. 551–560.
61. N. Leung and V. Reiner, *The signature of a toric variety*, Duke Math. J., **111** (2002), pp. 253–286.
62. T. Lewiner, H. Lopes, G. Tavares, *Visualizing Forman's discrete vector field*, in Mathematical Visualization III, H.-C. Hege, K. Polthier (eds.), Springer-Verlag, Heidelberg, 2002, pp. 95–112.
63. T. Lewiner, H. Lopes, G. Tavares, *Optimal discrete Morse functions for 2-manifolds*, Comput. Geom., **26** (2003), pp. 221–233.
64. T. Lewiner, H. Lopes, G. Tavares, *Towards optimality in discrete Morse theory*, Experiment. Math., **12** (2003), pp. 271–285.
65. S. Linusson and J. Shareshian, *Complexes of t-colorable graphs*, SIAM J. Discrete Math., **16** (2003), pp. 371–389.
66. J. Lott and W. Lück, *L^2-topological invariants of 3-manifolds*, Invent. Math, **120** (1995), pp. 15–60.
67. W. Lück, *L^2-Betti numbers of mapping tori and groups*, Topology, **33** (1994), pp. 203–214.
68. A. Lundell and S. Weingram, *The topology of CW complexes*, Van Nostrand Reinhold Company, New York, 1969.
69. P. McMullen, *On simple polytopes,* Invent. Math, **113** (1993), pp. 419–444.
70. P. McMullen, *Weights on polytopes*, Discrete Comput. Geom., **15** (1996), pp. 363–388.
71. J. Milnor, *Morse theory*, Annals of Mathematics Study No. 51, Princeton University Press, 1962.
72. J. Milnor, *Lectures on the h-cobordism theorem*, Princeton Mathematical Notes, Princeton University Press, 1965.
73. M. Morse, *The calculus of variations in the large*, Amer. Math. Soc. Colloquium Pub. 18, Amer. Math. Soc., Providence R.I., 1934.
74. M. Morse, *Bowls of a non-degenerate function on a compact differentiable manifold*, in Differential and Combinatorial Topology (A Symposium in Honor of Marston Morse), Princeton University Press (1965), pp. 81–104.
75. V. Reiner and V. Welker, *On the Charney-Davis and the Neggers-Stanley conjectures*, J. Combin. Theory Ser. A, **109** (2005), pp. 247–280.
76. R. Rivest and J. Vuillemin, *A generalization and proof of the Aanderaa-Rosenberg conjecture*, 7th Annual ACM Symposium on Theory of Computing, Albuquerque, New Mexico, 1975, pp. 6–12.
77. R. Rivest and J. Vuillemin, *On recognizing graph properties from adjacency matrices*, Theor. Comp. Sci, **3** (1976), pp. 371–384.
78. M. Schwartz, *Morse homology*, Progress in Mathematics 111, Birkhäuser Verlag, Basel, 1993.
79. J. Shareshian, *Discrete Morse theory for complexes of 2-connected graphs*, Topology, **40** (2001), pp. 681–701.
80. V. Sharko, *Functions on manifolds, algebraic and topological aspects*, Translations of Math. Monographs 131, Amer. Math. Soc., Providence, R.I., 1993.

81. E. Sköldberg, *Combinatorial discrete Morse theory from an algebraic viewpoint*, Trans. Amer. Math. Soc., **358** (2006), pp. 115–129.
82. S. Smale, *On gradient dynamical systems*, Annals of Math., **74** (1961), pp. 199–206.
83. S. Smale, *The generalized Poincaré conjecture in dimensions greater than four*, Annals of Math., **74** (1961), pp. 391–406.
84. D. Soll, *Evasive Simpliziale Komplexe und Diskrete Morse Theorie*, Diploma Thesis, Philipps-Universität Marburg, 2002.
85. R. Stanley, *A combinatorial decomposition of acyclic simplicial complexes*, Discrete Math., **118** (1993), pp. 175–182.
86. J. Stasheff, *Homotopy associativity of H-spaces*, Trans. of the Amer. Math. Soc., **108** (1963), pp. 275–292.
87. J. Stasheff, *The prehistory of operads*, in Operads: proceedings of the Renaissance conferences (Hartford, CT/Luminy, 1995), Cotemp. Math. 2002, pp. 9–14.
88. R. Stanley, *The number of faces of a simplicial convex polytope,* Adv. in Math., **35** (1980), pp. 236–238.
89. R. Stanley, *Flag f-vectors and the cd-index*, Math. Zeitschrift, 216 (1994), pp. 483–499.
90. R. Stanley, *Combinatorics and commutative algebra*, 2nd ed., Birkäuser Verlag, Basel, 1996.
91. R. Stanley, *Positivity problems and conjectures in algebraic combinatorics*, in Mathematics: frontiers and perspectives (V. Arnold, M. Atiyah, P. Lax, and B. Mazur, eds.), AMS, Providence, RI, 2000, pp. 295–319.
92. V. Turchun, *Homology of complexes of biconnected graphs*, Uspekhi Mat. Nauk., **52** (1997), pp. 189–190.
93. V. Vassiliev, *Complexes of connected graphs*, in The Gelfand mathematical seminars, 1990–1992, Birkhäuser Verlag, Boston, 1993, pp. 223–235.
94. V. Welker, *Constructions preserving evasiveness and collapsibility*, Discrete Math., **207** (1999), pp. 243–355.
95. J. Whitehead, *Simplicial spaces, nuclei, and m-groups*, Proc. London Math. Soc., **45** (1939), pp. 243–327.

Geometry of q and q,t-Analogs in Combinatorial Enumeration

Mark Haiman and Alexander Woo

IAS/Park City Mathematics Series
Volume 14, 2004

Geometry of q and q, t-Analogs in Combinatorial Enumeration

Mark Haiman and Alexander Woo

Introduction

The aim of these lectures was to give an overview of some combinatorial, symmetric-function theoretic, and representation-theoretic developments during the last several years in the theory of Hall-Littlewood and Macdonald polynomials. The motivating problem for all these developments was Macdonald's 1988 positivity conjecture [**20, 21**]. The positivity conjecture asserts that certain polynomials have non-negative integer coefficients, and so it naturally raised the question of how to understand Macdonald polynomials combinatorially. This question remains open, even after the proof of the positivity conjecture in [**16**], using methods from algebraic geometry. The latest developments, which will be discussed at the end of these notes, for the first time promise progress on the combinatorial side of the problem.

The lectures start with basics and proceed towards a discussion of the most recent combinatorial advances. Along the way, I have taken as my central topic the q and q, t-analogs of classical combinatorial themes such as Catalan numbers, enumeration of trees and parking functions, and Lagrange inversion. The surprising connection between these themes and the theory of Macdonald polynomials was one of the most beautiful discoveries to emerge from work on the positivity conjecture. This topic also serves nicely to motivate the combinatorial conjectures discussed in the final lecture.

The subject as a whole has grown far beyond what can be covered in a series of introductory lectures. Omitted entirely are the algebraic geometrical aspects [**2, 14, 16, 17**]. Also omitted is a treatment of the full list of other quantities, not quite so immediately connected with classical enumeration, which are also expressed by formulas involving Macdonald polynomials, and are known or conjectured to be Schur-positive, for which combinatorial interpretations are still sought [**1**]. Yet another direction not touched on here is the link with representation theory of

[1]Dept. of Mathematics, University of California, Berkeley, CA.
E-mail address: `mhaiman@math.berkeley.edu, awoo@math.berkeley.edu`.
Work supported in part by NSF grant DMS-0301072 (M.H.).

©2007 American Mathematical Society

Cherednik algebras and their degenerations [**4, 5, 10, 11**]. A more advanced but less up-to-date survey of some of these topics can be found in [**18**].

My heartfelt thanks go to Alexander Woo, who conducted discussion and exercise sessions associated with the lectures and did most of the work in preparing these notes. In the process he greatly improved the exposition, worked out missing details, and took pains to clarify those points which proved most troublesome for students in the discussion sections. Credit for whatever good qualities the following notes may possess is mostly due to him.

–M.H.

LECTURE 1
Kostka Numbers and q-Analogs

1.1. Definition of Kostka Numbers

Let n be a nonnegative integer. A **partition** of n is a non-decreasing sequence of nonnegative integers $\lambda = (\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_l)$ such that $n = \lambda_1 + \lambda_2 + \cdots + \lambda_l$. The number n is known as the **size** of λ and denoted $|\lambda|$. Assuming we have written λ so that $\lambda_l \neq 0$, the number l is the **length** of λ and denoted $l(\lambda)$.

We can associate to any partition a pictorial representation called the **Young diagram**, or sometimes the **Ferrers diagram**. It consists of boxes (i, j) in the first quadrant such that $j < \lambda_{i+1}$. For example, the Young diagram for the partition $\lambda = (5, 4, 2)$ is in Figure 1. Note the standard convention in the literature, which we follow, is that boxes are labelled (row, column) as in upside-down matrix coordinates.

To keep notation simple, we will frequently write λ to indicate its diagram when there is no possibility of confusion.

A **semistandard Young Tableaux** (abbreviated SSYT) of **shape** λ is a filling of the boxes of the diagram of λ by positive integers, that is, a function $T : \text{diagram}(\lambda) \to \mathbb{Z}_{>0}$, such that rows are non-decreasing as one moves to the right, and columns are strictly increasing as one moves up. For example, Figure 2 is a SSYT of shape $(5, 4, 2)$.

The **content** μ of a tableau is the sequence $\mu_1, \ldots, \mu_k$ with $\mu_i = \#(T^{-1}(i))$. It is obviously a composition of n (that is, a sequence $\mu_1, \ldots, \mu_k$ of nonnegative integers such that their sum is n). The SSYT in Figure 2 has content $\mu = (3, 3, 2, 2, 1)$.

The **Kostka number** $K_{\lambda\mu}$ is then the number of SSYT of shape λ and content μ. Kostka numbers (and, by extension, Young tableaux) have significance in the

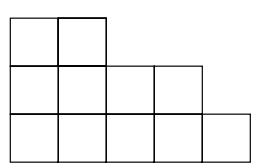


Figure 1. Young diagram for $\lambda = (542)$

3	5			
2	2	4	4	
1	1	1	2	3

Figure 2. SSYT of shape $\lambda = (542)$ and content $\mu = (33221)$

theory of symmetric functions, and in the representation theory of S_n and of GL_n. We will visit these interpretations of $K_{\lambda\mu}$ in order.

1.2. $K_{\lambda\mu}$ in Symmetric Functions

We have the following lemma whose proof consists of finding a simple bijection and is left as an exercise:

Lemma 1. *$K_{\lambda\mu}$ is independent of the order of the parts of μ.*

This states that, for example, $K_{(41),(221)} = K_{(41),(212)} = K_{(41),(122)}$. Therefore, we can, and by convention usually will, consider $K_{\lambda\mu}$ only in the case where μ is a partition.

To an SSYT T, we can associate the monomial $x^T := \prod_{c\in\lambda} x_{T(c)}$ in $\mathbb{Z}[x_1, x_2, \ldots]$ or $\mathbb{C}[x_1, x_2, \ldots]$. In this product we have written $c \in \lambda$ to indicate that c is a cell in the diagram of λ. Now we can associate to each partition λ the **Schur function**

$$s_\lambda = \sum_{\mathrm{SSYT}(\lambda)} x^T.$$

By our definition, this is a "polynomial" in infinitely many variables, and, by Lemma 1, it is symmetric. The Schur functions form a basis for the ring of symmetric functions. Although they may seem unmotivated at first, in light of what follows, the Schur functions should probably be considered the most natural basis for the ring of symmetric functions.

Perhaps the most obvious basis for the ring of symmetric functions consists of the **monomial symmetric functions**, defined by

$$m_\mu = x_1^{\mu_1} x_2^{\mu_2} \cdots x_k^{\mu_k} + \text{all symmetric terms.}$$

By our definition of s_λ, we have

$$s_\lambda = \sum_\mu K_{\lambda\mu} m_\mu,$$

where the sum is taken over all partitions μ, or equivalently all partitions μ of size $|\lambda|$.

1.3. S_n Representations

Let V be a finite dimensional S_n representation, that is, a finite dimensional $\mathbb{C}$-vector space with a linear action by S_n. For any partition μ of n, there is the **Young subgroup** $S_\mu = S_{\mu_1} \times S_{\mu_2} \times \cdots \times S_{\mu_k} \subseteq S_n$, where the S_{μ_1} factor permutes the first μ_1 letters, the S_{μ_2} factor permutes the $\mu_1 + 1$-th through $\mu_1 + \mu_2$-th letters, and so on. Now let V^{S_μ} denote the subspace of V fixed by every element of S_μ. Then there

is a symmetric function associated to V, called the **Frobenius characteristic** of V, defined by

$$F_V(\mathbf{x}) = \sum_{|\mu|=n} (\mathrm{dim} V^{S_\mu}) m_\mu(\mathbf{x}).$$

(This is not quite the usual definition of F_V, as for example in [**21, 22**], but it can easily be seen to be equivalent to the usual one.)

A representation is said to be **irreducible** if it has no proper sub-representations. By a classical theorem of Maschke, any representation of a finite group (over $\mathbb{C}$) splits as the direct sum of irreducible representations. Therefore, it suffices to study the irreducible representations. For S_n, we have the following theorem of Frobenius.

Theorem 1. *The irreducible representations V_λ of S_n are determined up to isomorphism by their Frobenius characteristics, and $F_{V_\lambda}(\mathbf{x}) = s_\lambda(\mathbf{x})$.*

Note that Frobenius characteristic is additive on direct sums, so this theorem essentially describes the representation theory (over $\mathbb{C}$) of S_n completely.

As examples, we have the two one-dimensional irreducible representations of S_n.

Example 1. *Let $V = \mathbb{C}$, with S_n acting trivially. (In other words, every element of S_n acts as the identity.) Then $\mathrm{dim} V^{S_\mu} = 1$ for every S_μ, so $F_V(\mathbf{x}) = \sum_{|\mu|=n} m_\mu(\mathbf{x})$. This representation is clearly irreducible, so $F_V(\mathbf{x}) = s_\lambda(\mathbf{x})$ for some partition of n. Since the partition (n) has the property that, for any μ of size n, there is exactly one SSYT of shape (n) and content μ, we have $F_V(\mathbf{x}) = s_{(n)}(\mathbf{x})$.*

Example 2. *Now let $V = \mathbb{C}$, but with S_n acting by sign. That is, let $w \in S_n$ act by the identity if w is an even permutation but by -1 if w is odd. Except for $\mu = (1, 1, \ldots, 1) = (1^n)$, every S_μ has an odd permutation, so $\mathrm{dim} V^{S_\mu} = 0$ except when $\mu = (1^n)$. Therefore, $F_V(\mathbf{x}) = m_{(1^n)}(\mathbf{x})$. The unique partition λ which admits only one symmetry class of SSYT's of the given shape is $\lambda = (1^n)$, so $F_V(\mathbf{x}) = s_{(1^n)}(\mathbf{x})$.*

The symmetric functions associated with these examples have special importance and therefore have their own names. The symmetric function $s_{(n)}$ is known as the **complete homogeneous symmetric function** (of degree n) and is denoted h_n. The symmetric function $s_{(1^n)}$ is known as the **elementary symmetric function** and is denoted e_n.

There is another interpretation of $K_{\lambda\mu}$ in terms of the representation theory of S_n, which we will only sketch briefly. Let W_μ be the set of words of content μ, that is, words with μ_1 1's, μ_2 2's, and so on, and let S_n act on these words by permuting the positions of their letters. Extending by linearity gives an S_n representation on $\mathbb{C} \cdot W_\mu$. Then we have that

Proposition 1.

$$F_{\mathbb{C}\cdot W_\mu}(\mathbf{x}) = \sum_\lambda K_{\lambda\mu} s_\lambda(\mathbf{x}),$$

or, equivalently,

$$W_\mu \cong \bigoplus_\lambda V_\lambda^{\oplus K_{\lambda\mu}}.$$

This can be proven by identifying $\mathbb{C} \cdot W_\mu$ with the induced representation $\mathbb{C} \uparrow_{S_\mu}^{S_n}$ and using Frobenius reciprocity.

1.4. GL_n Representations

Now we consider finite dimensional GL_n representations. Here we restrict ourselves to **rational** representations; that is, a representation V determines a map $\rho : \mathrm{GL}_n \to \mathrm{GL}(V)$, and we require that we can find polynomials f_{ij} (in n^2 variables) such that for a matrix g, $\rho(g)$ is the matrix $1/\det(g)^N\,[f_{ij}(g_{11},\cdots,g_{nn})]$ for some nonnegative integer N. If such polynomials exist with $N \le 0$, then V is a **polynomial** representation.

The 1-dimensional trivial representation is polynomial, with $\rho(g) = [1]$, and the n-dimensional defining representation is also polynomial, since $\rho(g) = g$. For a rational (resp. polynomial) representation V, there are naturally defined rational (resp. polynomial) representations on $V^{\otimes k}$, $\bigwedge^k V$ and $\mathrm{Sym}^k V$.

Both $\bigwedge^k$ and Sym^k can be considered as operations which construct new representations from existing ones. They have generalizations, one for each partition λ, called **Schur functors**, and denoted $\mathcal{S}^\lambda$. Given a representation V, $\mathcal{S}^\lambda V$ is defined as follows.

For any l, there is the natural inclusion $\bigwedge^l V \to V^{\otimes l}$ given by $v_1 \wedge \cdots \wedge v_l \mapsto \sum_{\sigma\in S_l} \mathrm{sgn}(\sigma) v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(l)}$, and the natural surjection $V^{\otimes l} \to \mathrm{Sym}^l V$ given by $v_1 \otimes \cdots \otimes v_n \mapsto v_1 \cdots v_n$. Note that these maps respect the GL_n action, so they are maps not only of vector spaces but also of GL_n representations. Let $T = \bigotimes_{(i,j)\in\lambda} V^{(i,j)}$, where each $V^{(i,j)}$ is an isomorphic copy of V, so that $T \cong V^{\otimes|\lambda|}$.

Now we define the map $i : \bigwedge^{\lambda'_1} V \otimes \cdots \otimes \bigwedge^{\lambda'_{\lambda_1}} V \to T$ by using the natural inclusion given above to map the tensor factor $\bigwedge^{\lambda'_k}$ to $\bigotimes_{j=1}^{\lambda'_k} V^{(j,k)}$. Then define the map $\pi : T \to \mathrm{Sym}^{\lambda_1} V \otimes \cdots \otimes \mathrm{Sym}^{\lambda_{l(\lambda)}} V$ by using the natural surjection to map $\bigotimes_{i=1}^{\lambda_k} V^{(k,i)}$ to $\mathrm{Sym}^{\lambda_k} V$, and let $\phi = \pi \circ i$. Finally, $\mathcal{S}^\lambda V$ is defined to be $\mathrm{im}\,\phi$. In particular, for $\lambda = (k)$, $\mathcal{S}^{(k)} = \mathrm{Sym}^k$, and for $\lambda = (1^k)$, $\mathcal{S}^{(1^k)} = \bigwedge^k$. Since, assuming V is a rational (resp. polynomial) representation, ϕ is a map of rational (resp. polynomial) GL_n representations, $\mathcal{S}^\lambda V$ is also a rational (resp. polynomial) representation.

For the remainder of this section let V be the n-dimensional defining representation, and let $V^\lambda := \mathcal{S}^\lambda V$.

Theorem 2. (1) *The representation V^λ is irreducible, and every irreducible polynomial representation of GL_n is V^λ for some λ.*

(2) *Let $T \subseteq \mathrm{GL}_n$ denote the subgroup of (invertible) diagonal matrices, and let* $g(\mathbf{x}) := \begin{bmatrix} x_1 & & 0 \\ & \ddots & \\ 0 & & x_n \end{bmatrix} \in T$. *Then $\mathrm{tr}(V^\lambda, g(\mathbf{x})) = s_\lambda(\mathbf{x})$. Equivalently, there is a decomposition of V^λ, considered as a representation of T, into $V^\lambda = \bigoplus_\mu (V^\lambda)_\mu$, where $g(\mathbf{x})$ acts on $(V^\lambda)_\mu$ by multiplication by $\mathbf{x}^\mu$, and $\dim(V^\lambda)_\mu = K_{\lambda\mu}$.*

The most basic examples are $\lambda = (k)$, in which case $\mathrm{tr}(\mathrm{Sym}^k V, g(\mathbf{x})) = h_k(\mathbf{x})$, and $\lambda = (1^k)$ for $k \le n$, for which $\mathrm{tr}(\bigwedge^k V, g(\mathbf{x})) = e_k(\mathbf{x})$.

1.5. The q-Analog $K_{\lambda\mu}(q)$

The aim of this section is to describe a q-analog of $K_{\lambda\mu}$ known as $\widetilde{K}_{\lambda\mu}(q)$ and make some brief remarks about its properties. Here (algebraic) geometry makes its appearance. For each partition μ, we will define a variety Y_μ whose cohomology ring will have a natural action of S_n. Then we will define $\widetilde{K}_{\lambda\mu}(q)$ as the graded Frobenius characteristic of this cohomology ring.

Let $\mathcal{N}$ be the set of nilpotent $n\times n$ matrices. This set can be given the structure of an algebraic variety; nilpotent matrices are precisely the matrices X for which $X^n = 0$, and the entries of X^n are polynomials in the entries of X, so $\mathcal{N}$ is defined as an affine variety in $\mathbb{C}^{n\times n}$ by the vanishing of these polynomials. The variety $\mathcal{N}$ is singular, so we would like to understand it better by studying a smooth variety similar to it. More precisely, we would like a **resolution of singularities** for $\mathcal{N}$, that is, a variety $\mathcal{Z}$ along with a map $\pi : \mathcal{Z} \to \mathcal{N}$, with the properties that $\mathcal{Z}$ is smooth, and π is both projective and birational. (Birational means that π is an isomorphism on an open dense set, and projective means that π can be factored as some inclusion $i : \mathcal{Z} \to \mathcal{N} \times \mathbf{P}^k$ (for some k) followed by the usual projection $\mathcal{N} \times \mathbf{P}^k \to \mathcal{N}$.)

To construct $\mathcal{Z}$, we need the **flag variety**. A **flag** is a sequence of vector subspaces of $\mathbb{C}^n$ denoted $F_\bullet = 0 \subseteq F_1 \subseteq F_2 \subseteq \cdots \subseteq F_{n-1} \subseteq \mathbb{C}^n$, satisfying $\dim F_i = i$. The flag variety contains, as a set, all flags in $\mathbb{C}^n$; as a variety or manifold it is the quotient G/B where $G = \mathrm{GL}_n$ and B consists of the upper triangular matrices. Now we can let

$$\mathcal{Z} = \{(X, F_\bullet) \in \mathcal{N} \times G/B : XF_i \subseteq F_{i-1} \text{ for all } i\},$$

with the map π being the projection onto the first factor.

Now we show that $\mathcal{Z}$ is smooth. Let $\psi : \mathcal{Z} \to G/B$ be the projection onto the second factor, and let $E_\bullet$ be the standard flag, that is, the flag with $E_i = \mathbb{C}\cdot\{e_1, \ldots, e_i\}$, where $\{e_1, \ldots, e_n\}$ is the standard basis of $\mathbb{C}^n$. The fiber $\psi^{-1}(E_\bullet)$ is given by $\psi^{-1}(E_\bullet) = \{(X, E_\bullet) : X \text{ is upper triangular}\}$, so $\psi^{-1}(E_\bullet)$ is a $\binom{n}{2}$-dimensional vector space. Moreover, for any flag $F_\bullet$, $F_\bullet = gE_\bullet$ for some $g \in G$, and $\psi^{-1}(F_\bullet) = \big\{(gXg^{-1}, F_\bullet) : X \text{ is upper triangular}\big\}$, also a vector space of dimension $\binom{n}{2}$. This makes $\mathcal{Z}$ into a vector bundle over G/B; since G/B is smooth, $\mathcal{Z}$ must also be smooth. (Technically we also need to check that $\mathcal{Z}$ is locally trivial over G/B, but this is also easy to check using the group action.)

The map π is projective because G/B is a projective variety. Also, for any X whose Jordan form has only one Jordan block, $\pi^{-1}(X)$ consists of a single flag, so, as these matrices X form an open dense subset of $\mathcal{N}$, π is birational.

Now let G act on $\mathcal{N}$ by conjugation; that is, $g \cdot X := gXg^{-1}$ for $g \in G$ and $X \in \mathcal{N}$. Let μ be a partition. Let M_μ be the nilpotent Jordan matrix with Jordan blocks of size $\mu_1, \mu_2, \cdots, \mu_k$, and $O_\mu = \mathrm{GL}_n \cdot M_\mu$. These orbits cover all of $\mathcal{N}$, since every matrix has a Jordan form and we can conjugate by permutation matrices to rearrange the Jordan blocks so that their sizes are in non-increasing order. We have a corresponding action on $\mathcal{Z}$ by $g \cdot (X, F_\bullet) := (gXg^{-1}, gF_\bullet)$, so the fibers of π over points in the same G orbit are isomorphic. Let $Y_\mu = \pi^{-1}(P)$ for some point $P \in O_\mu$. (We will only be interested in isomorphism invariants of Y_μ, so the choice of point is irrelevant.)

For example, for $\mu = (n)$, $Y_{(n)}$ is a single point, as already stated above. At the other extreme, when X is the zero matrix, $(X, F_\bullet) \in \mathcal{Z}$ for every $F_\bullet$, so for $\mu = (1^n)$, $Y_{(1^n)} \cong G/B$.

The following theorem allows what we will consider the definition of $K_{\lambda\mu}(q)$.

Theorem 3.

(1) *The natural map $H^*(G/B, \mathbb{C}) \to H^*(Y_\mu, \mathbb{C})$ is surjective.*
(2) *There are geometrically defined S_n actions on $H^*(G/B, \mathbb{C})$ and $H^*(Y_\mu, \mathbb{C})$ such that the above map is S_n-equivariant.*
(3) *$H^*(Y_\mu, \mathbb{C}) \cong \mathbb{C} \cdot W_\mu \cong \bigoplus_\lambda V_\lambda^{\oplus K_{\lambda\mu}}$ as S_n-representations.*

Now we define $\widetilde{K}_{\lambda\mu}(q)$ by

$$\widetilde{K}_{\lambda\mu}(q) = \sum_i K_{\lambda\mu}^{(i)} q^i,$$

where $K_{\lambda\mu}^{(i)}$ is defined by

$$H^{2i}(Y_\mu, \mathbb{C}) \cong_{S_n} \bigoplus V_\lambda^{\oplus K_{\lambda\mu}^{(i)}}.$$

(The original definition of $K_{\lambda\mu}(q)$ is related to $\widetilde{K}_{\lambda\mu}(q)$ by $K_{\lambda\mu}(q) = q^N \widetilde{K}_{\lambda\mu}(q^{-1})$, where N is a positive integer depending on μ. The $\widetilde{K}_{\lambda\mu}(q)$ appear to be somewhat more natural so we will be using this form throughout the lectures.)

From the definition and part 3 of the theorem, it is clear that $\widetilde{K}_{\lambda\mu}(1) = K_{\lambda\mu}$, and that $\widetilde{K}_{\lambda\mu}(q)$ is a polynomial with positive integer coefficients, but it is not so clear how to compute $\widetilde{K}_{\lambda\mu}(q)$. We will see later a formula of Shoji and Lusztig for $\widetilde{K}_{\lambda\mu}(q)$, but it will be a rational expression from which it is not obvious that $\widetilde{K}_{\lambda\mu}$ is a polynomial, much less one with positive coefficients.

However, there is a combinatorial definition due to Lascoux and Schützenberger which gives $\widetilde{K}_{\lambda\mu}(q) = \sum_{T \in \mathrm{SSYT}(\lambda,\mu)} q^{\mathrm{cc}(T)}$, where the **co-charge** $\mathrm{cc}(T)$ is a certain rather subtle combinatorial statistic on tableaux. Somewhat unsatisfactorily, the proofs that the two definitions are equivalent rely on showing that they both satisfy initial conditions and recurrence relations which are sufficient to determine $\widetilde{K}_{\lambda\mu}(q)$. A better proof would explain this equivalence by explicitly finding a basis of $H^*(Y_\mu, \mathbb{C})$ indexed by tableaux whose co-charge is equal to the cohomology degree of the basis element, with the S_n action on the cohomology ring closely related to the S_n action on the corresponding tableaux. However, no such conceptually satisfactory proof is yet known.

1.6. Exercises

(1) Prove Lemma 1.
(2) Define $h_\mu(\mathbf{x}) := h_{\mu_1}(\mathbf{x}) h_{\mu_2}(\mathbf{x}) \cdots h_{\mu_{l(\mu)}}(\mathbf{x})$. Show that $F_{\mathbb{C}\cdot W_\mu}(\mathbf{x}) = h_\mu(\mathbf{x})$. Deduce that $h_\mu(\mathbf{x}) = \sum_\lambda K_{\lambda\mu} s_\lambda$.
(3) Find a basis and weight space decomposition of V^λ (the GL_n representation) for $\lambda = (2, 1^{k-2})$.
(4) Let $V = \mathbb{C}^n = \mathbb{C} \cdot \{e_1, \ldots, e_n\}$ be the defining representation of S_n, that is, with the action $w \cdot e_i = e_{w(i)}$. Decompose V into irreducibles and $F_V(\mathbf{x})$ into Schur functions, corresponding to your decomposition of V.

LECTURE 2
Catalan Numbers, Trees, Lagrange Inversion, and their q-Analogs

2.1. Catalan Numbers

The Catalan numbers C_n are known to count many different combinatorial objects, but for the sake of brevity we will only mention a small number which will be important for these lectures.

Let w be a string consisting of n left parentheses "(" and n right parentheses ")". The string w is **proper** if it makes sense as a parenthesization of something, that is, if, reading from left to right, one has encountered at every point at least as many left parentheses as right parentheses.

To every proper parentheses string we can associate a **Dyck path**, that is, a lattice path from $(0, n)$ to $(n, 0)$ (using Cartesian coordinates) which always remains below (or on) the line defined by $x + y = n$. We do this by starting at $(0, n)$ and, as we read a string w from left to right, moving down by $(0, -1)$ every time we encounter a "(" and moving to the right by $(1, 0)$ every time we encounter a ")". By considering the Dyck path as the boundary of the diagram of a partition, the set of Dyck paths is also equivalent to the set of partitions $\lambda \subseteq \delta(n)$, where $\mu \subseteq \nu$ for partitions μ and ν means that the diagram of μ fits inside the diagram of ν (that is, $\mu_i \leq \nu_i$ for all i), and $\delta(n)$ is the partition $(n-1, n-2, \ldots, 1, 0)$.

For example, the above bijections associate the word "()(()())" with the Dyck path in Figure 1 and the partition $(2, 1, 1)$.

The **Catalan numbers** C_n can then be defined as the number of proper parentheses strings (of n left and n right parentheses), or equivalently the number of Dyck

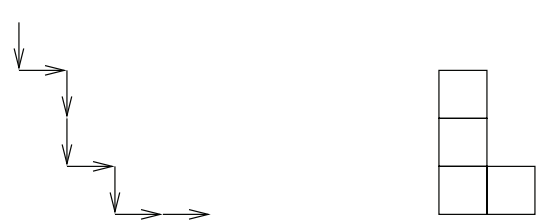


Figure 1. Dyck path and partition corresponding to ()(()())

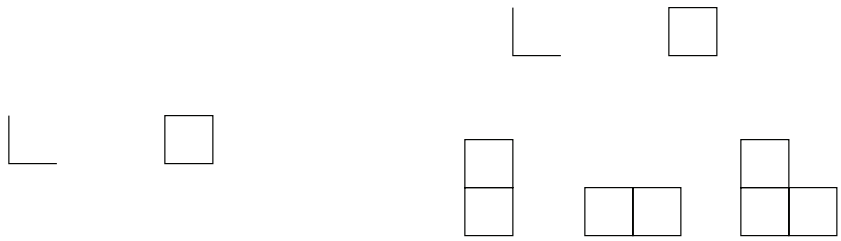


Figure 2. Partitions counted by C_2 and C_3

paths from $(0,n)$ to $(n,0)$, or equivalently the number of partitions inside $\delta(n)$. We have $C_0 = C_1 = 1$, $C_2 = 2$, and $C_3 = 5$, as demonstrated by the Figure 2.

As is frequently useful in combinatorics, we can try to calculate or get a formula for C_n by using a **generating function**. In this case, this means a power series $C(x)$ defined by $C(x) := \sum_n C_n x^n$.

Given a proper parentheses string, the initial "(" matches with some ")", and between those parentheses is a proper parenthesization of some length k, while after the specified ")" is another proper parentheses string of length $n-1-k$. In other words, a non-empty proper parentheses string looks like $(A)B$, where A and B are respectively parentheses strings of length k and $n-1-k$. Therefore, the Catalan numbers satisfy the recurrence $C_n = \sum_{k=0}^{n-1} C_k C_{n-1-k}$. In terms of the generating function, we have

$$C(x) = 1 + xC(x)^2.$$

We can get a formula for C_n by solving for $C(x)$ and using the binomial theorem, but we will instead get one by using Lagrange inversion later in this lecture. For now, note that our equation can be rewritten as

$$xC(x)(1 - xC(x)) = x,$$

or equivalently $F_1(x) \circ (xC(x)) = x$, where $\circ$ denotes functional composition and $F_1(x) = x(1-x)$. In other words, $F_1(x)$ and $xC(x)$ are **compositional inverses**.

2.2. Rooted Trees

A **tree** is a connected graph with no cycles, and a **labelled tree** is a tree whose vertices are assigned distinct labels. A **rooted tree** is a tree in which one vertex is distinguished and designated as the root. Let t_n be the number of labelled rooted trees with vertex set $\{1, 2, \ldots, n\}$, with $t_0 = 0$ by convention. A **rooted forest** is a graph with labelled vertices and no cycles where each connected component has a vertex designated as the root. Note that the number of rooted forests on n vertices is the same as the number of unrooted trees with vertex set $\{0, \ldots, n\}$, which is $t_{n+1}/(n+1)$; this is because, as illustrated in Figure 3, for any rooted forest we can construct a tree by creating a new vertex labelled 0 and adding an edge between 0 and each root, and conversely given a tree with vertex set $\{0, \ldots, n\}$ we can construct a labelled rooted forest by removing the vertex 0 and calling each vertex formerly attached to 0 the root of its connected component.

As with Catalan numbers, we can form a generating series, but in this case it will be more convenient to form the **exponential generating series** $T(x) = \sum_n t_n x^n/n!$. This allows us to use the Exponential Formula [**22**, Cor. 5.1.6] to

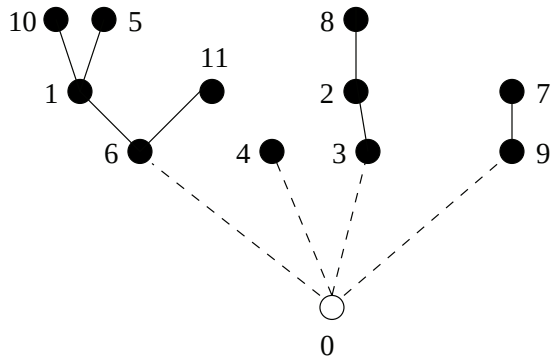


Figure 3. Construction of a rooted tree from a rooted forest

relate the generating series for the number of rooted trees and the number of rooted forests, so that, if $h_n = t_{n+1}/(n+1)$ is the number of rooted labelled forests and $H(x) = \sum_n h_n x^n/n!$, we have $H(x) = e^{T(x)}$. Therefore,

$$e^{T(x)} = \sum_n \frac{t_{n+1}}{n+1}\frac{x^n}{n!} = \sum_n t_{n+1}\frac{x^n}{(n+1)!} = \frac{T(x)}{x}.$$

Then $T(x)e^{-T(x)} = x$ so $F_2(x) = xe^{-x}$ is the compositional inverse of $T(x)$.

2.3. The Lagrange Inversion Formula

Given these examples, it would be nice to have a formula which, given a power series, calculates the coefficients of its compositional inverse. The Lagrange inversion formula exactly fulfills this need.

Theorem 4. *Let $E(x) = \sum_n e_n x^n$ and $K(x) = \sum_n k_n x^n$ be power series, with $e_0 = k_0 = 1$. Then*

$$\frac{x}{E(x)} \circ (xK(x)) = x$$

if and only if

$$k_n = \frac{1}{n+1}\,[x^n]\,E(x)^{n+1}.$$

Here and below the symbol $[x^n]$ denotes the coefficient of x^n in the quantity that follows. We will later prove Theorem 4 as a special case of a q-Lagrange inversion theorem. A direct proof can be found, for example, in [**22**, Thm 5.4.2].

Let us use this theorem to calculate explicit formulas for C_n and t_n.

To solve for C_n, let $E(x) = 1/(1-x)$. Then $x/E(x) = F_1(x)$, so

$$\frac{x}{E(x)} \circ (xC(x)) = x.$$

Applying the Lagrange inversion theorem,

$$C_n = \frac{1}{n-1}\,[x^n]\,\frac{1}{(1-x)^{n+1}} = \frac{1}{n-1}\left(\binom{n+1}{n}\right) = \frac{1}{n-1}\binom{2n}{n}.$$

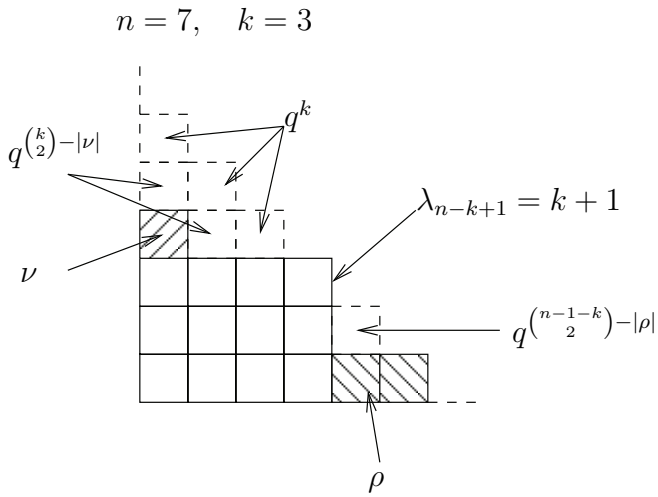


Figure 4. The q-Catalan recurrence illustrated for $\lambda = (6, 4, 4, 1)$.

It turns out to be slightly easier to solve for h_n, the number of rooted forests. If we let $E(x) = e^x$, then $x/E(x) = F_2(x)$, and

$$\frac{x}{E(x)} \circ (xH(x)) = x.$$

Once again applying Lagrange inversion,

$$\frac{h_n}{n!} = \frac{1}{n+1}\,[x^n]\,e^{(n+1)x} = \frac{1}{n+1}\frac{(n+1)^n}{n!},$$

so $h_n = (n+1)^{n-1}$, and $t_n = n^{n-1}$.

2.4. q-Analogs

The Catalan numbers have two q-analogs, but we will only be concerned with the one originally defined by Carlitz and Riordan [**3**], namely $C_n(q) = \sum_{\lambda \subseteq \delta(n)} q^{\binom{n}{2}-|\lambda|}$. This q-analog satisfies a recurrence as follows. We can separate all partitions $\lambda \subseteq \delta(n)$ into classes $\mathcal{C}_n^{(k)}$ for $0 \le k < n-1$ by putting λ in $\mathcal{C}_n^{(k)}$ if k is the smallest number such that $\lambda_{n-k-1} = k+1 = \delta(n)_{n-k-1}$, and $k = n-1$ if no such number exists. For example, as illustrated in Figure 4, the partition $(6, 4, 4, 1) \subseteq \delta(7)$ belongs in $\mathcal{C}_7^{(3)}$. Now, for $\lambda \in \mathcal{C}_n^{(k)}$, let ν be the partition defined by $\nu_i = \lambda_{n-k-1+i}$, and let ρ be the partition defined by $\rho_i = \lambda_i - k - 1$, for i, $1 \le i \le n-k-1$, as illustrated in Figure 4. Notice that $\nu \subseteq \delta_k$, and $\rho \subseteq \delta_{n-1-k}$. Furthermore,

$$\binom{n}{2} - |\lambda| = (k + \binom{k}{2} - |\nu|) + (\binom{n-1-k}{2} - |\rho|),$$

so we have the recurrence

$$C_n(q) = \sum_{k=0}^{n-1} q^k C_k(q) C_{n-1-k}(q).$$

Now we turn to a q-analog of $h_n = (n+1)^{n-1}$. It is possible to give a statistic on rooted labelled forests that produces this q-analog, but it will be more convenient

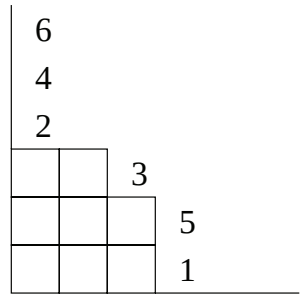


Figure 5. Tableau associated with the parking function $f(2) = f(4) = f(6) = 1, f(3) = 3, f(1) = f(5) = 4$

for us to define this q-analog later using parking functions. A **parking function** is a function $f : \{1, \ldots, n\} \to \{1, \ldots, n\}$ such that $\#f^{-1}(\{1, \ldots, k\}) \geq k$ for all $k \in \{1, \ldots, n\}$. (The name comes from the following description. Suppose we have n parking spaces on a one way street, labelled in order, and n cars. The cars arrive at the street in order, and each car k immediately goes to its preferred parking space $f(k)$. If it is already filled by a previous car, then it keeps going and parks in the first empty space. The condition above is then satisfied if and only if every car successfully finds a parking space without having to enter the street a second time.) Denote the set of parking functions for n cars by $\mathrm{PF}(n)$ The symmetric group S_n acts on $\mathrm{PF}(n)$ by $w \cdot f = f \circ w^{-1}$ for $w \in S_n$ and $f \in \mathrm{PF}(n)$.

We can represent a parking function by a tableaux of **skew shape** $(\lambda + (1^n))/\lambda$ for some partition λ, that is a filling of the boxes in $\lambda + (1^n)$ but not in λ, strictly increasing in columns and weakly increasing in rows (although in this case there are no relevant rows) as usual. Let $\widetilde{f}$ be f sorted into non-increasing order; in other words, we want $\widetilde{f} = w \cdot f$ for some w such that $\widetilde{f}(i) \geq \widetilde{f}(i+1)$ for all i, $1 \leq i \leq n-1$. Now we specify λ by requiring $\lambda_i = \widetilde{f}(i) - 1$. Note that the requirement that f (or $\widetilde{f}$) be a parking function is equivalent to requiring that $\lambda \subseteq \delta(n)$. Now the j-th column in $(\lambda + (1^n))/\lambda$ will have $f^{-1}(j)$ many open boxes to fill, and we fill them with the elements of $f^{-1}(j)$ in increasing order. Figure 5 shows the tableau associated with the parking function $f(2) = f(4) = f(6) = 1, f(3) = 3, f(1) = f(5) = 4$. The content of this tableaux is always (1^n).

Note that $\binom{n}{2} - |\lambda| = \sum_{i=1}^{n} i - \sum_{i=1}^{n} f(i)$, and we will denote this quantity as $\mathrm{wt}(f)$. (This quantity is sometimes known as the "frustration factor" of the parking function since it counts the sum total of how far drivers park from their preferred space.) Let $P_n(q) := \sum_{f \in \mathrm{PF}(n)} q^{\mathrm{wt}(f)}$. Counting parking functions according to the partition representing them, we get that

$$P_n(q) = \sum_{\lambda \subseteq \delta(n)} q^{\binom{n}{2} - |\lambda|} \binom{n}{\alpha_0, \alpha_1, \cdots, \alpha_{n-1}},$$

where α_i is the number of parts of λ of size i (adding parts of 0 if necessary so that λ has exactly n parts; that is, $\alpha_0 = n - \sum_{i=1}^{n-1} \alpha_i$).

2.5. q-Lagrange Inversion

To understand the above q-analogs better, we will give a q-analog of Lagrange inversion. Of course, for q-Lagrange inversion to make sense, we first have to define a q-analog of functional composition. The relevant definition is due to Garsia and Gessel independently [**8, 9**].

Definition 1. *Let $F(x) = \sum_n f_n x^n$ and $G(x) = \sum_n g_n x^n$ be power series with $g_0 = 0$. Then the q-**composition** of $F(x)$ and $G(x)$, denoted $F(x) \circ_q G(x)$, is defined to be $\sum_n f_n G(q^{n-1}x)G(q^{n-2}x)\cdots G(qx)G(x)$.*

Note that, if we substitute $q = 1$, we have $F(x) \circ G(x) = \sum_n f_n G(x)^n$, which is just ordinary composition of functions. Now we have a theorem, due to Garsia [**8**], which states that this setting gives a good q-analog of compositional inverses.

Theorem 5. *We have*

$$F(x) \circ_q G(x) = x \text{ if and only if } G(x) \circ_{q^{-1}} F(x) = x.$$

Furthermore, when the above holds,

$$(\Psi(x) \circ_q G(x)) \circ_{q^{-1}} F(x) = (\Psi(x) \circ_{q^{-1}} F(x)) \circ_q G(x) = \Psi(x) \text{ for all } \Psi(x).$$

Proof. Suppose that $F(x) \circ_q G(x) = x$. We will show that for any power series $\Psi(x)$, $(\Psi(x) \circ_{q^{-1}} F(x)) \circ_q G(x) = \Psi(x)$. In other words, we will show that, if we define maps $\pi, \phi : \mathbb{C}(q)[[x]] \to \mathbb{C}(q)[[x]]$ by $\pi : \Psi \mapsto \Psi \circ_{q^{-1}} F$ and $\phi : \Psi \mapsto \Psi \circ_q G$, $\phi \circ \pi$ is the identity map. Now, two power series are equal if and only if they are equal modulo x^N for every N, so we can view π and ψ as countable sequences of maps of finite-dimensional vector spaces. Therefore, if $\phi \circ \pi$ is the identity, $\pi \circ \phi$ is the identity, so, once we show that $(\Psi(x) \circ_{q^{-1}} F(x)) \circ_q G(x) = \Psi(x)$, we will have proven that $(\Psi(x) \circ_q G(x)) \circ_{q^{-1}} F(x) = \Psi(x)$. Then the forward direction of the first statement follows by $\Psi(x) = x$, which shows $G(x) \circ_{q^{-1}} F(x) = x$, and the reverse direction follows by substituting q^{-1} for q.

Now we show that $(\Psi(x) \circ_{q^{-1}} F(x)) \circ_q G(x) = \Psi(x)$. First we need the following lemma whose proof is straightforward from the definition of q-composition and left as an exercise.

Lemma 2.

$$(xF(x)) \circ_q G(x) = G(x)(F(x) \circ_q G(qx)) = G(x)(F(x) \circ_q G(x))_{x \mapsto qx}.$$

Now, if $F(x) \circ_q G(x) = x$, by applying the lemma we have $(xF(x)) \circ_q G(x) = G(x)qx$, by applying the lemma again we have $(x^2F(x)) \circ_q G(x) = G(x)G(qx)q^2x$, and by applying the lemma repeatedly we have

$$(x^rF(x)) \circ_q G(x) = G(x)G(qx)\cdots G(q^{r-1}x)q^rx.$$

Therefore, for all power series $\Phi(x) = \sum_n \phi_n x^n$,

$$\begin{aligned}(\Phi(x)F(x)) \circ_q G(x) &= \sum_n (\phi_n x^n F(x)) \circ_q G(x) \\ &= \sum_n \phi_n x q^n G(q^{n-1}x)G(q^{n-2}x)\cdots G(qx)G(x) \\ &= (\Phi(qx) \circ_q G(x))x.\end{aligned}$$

Apply the above equation with $\Phi(x) = F(q^{-1}x)$ to get

$$(F(q^{-1}x)F(x)) \circ_q G(x) = x^2.$$

Then apply the equation with $\Phi(x) = F(q^{-2}x)F(q^{-1}x)$, and the last equation, to get

$$(F(q^{-2}x)F(q^{-1}x)F(x)) \circ_q G(x) = x^3.$$

By induction, we have

$$(F(q^{-(n-1)}x) \cdots F(q^{-1}x)F(x)) \circ_q G(x) = x^n.$$

Therefore, for any power series $\Psi(x) = \sum_n \psi_n x^n$,

$$\begin{aligned}(\Psi(x) \circ_{q^{-1}} F(x)) \circ_q G(x) &= \sum_n \psi_n (F(q^{-(n-1)}x) \cdots F(qx)F(x)) \circ_q G(x) \\ &= \sum_n \psi_n x^n \\ &= \Psi(x),\end{aligned}$$

as desired. □

For usual functional composition, it turned out that it was easier to get the explicit Lagrange inversion formula for the modified form

$$\frac{x}{E(x)} \circ xK(x) = x,$$

or equivalently,

$$K(x) = E(x) \circ xK(x),$$

was easier to solve for the coefficients. (The equivalence is obvious once one stops using the $\circ$ notation.) Similarly, for q-composition, it is easier to state the q-Lagrange inversion formula for the following forms, whose equivalence is left as a (not so trivial) exercise.

Proposition 2.

$$\frac{x}{E(x)} \circ_q xK(qx) = x$$

if and only if

$$K(x) = E(x) \circ_q xK(x).$$

Now we are ready state the q-Lagrange inversion formula. It will not have a simple algebraic form, but will instead be a combinatorial sum that relates to the q-analogs described in section 2.4.

Theorem 6. *Let $E(x) = \sum_n e_n x^n$ and $K(x) = \sum_n k_n(q)x^n$ be power series, with $e_0 = k_0(q) = 1$. Then*

$$\frac{x}{E(x)} \circ_q (xK(qx)) = x$$

if and only if

$$k_n(q) = \sum_{\lambda \in \delta(n)} q^{\binom{n}{2} - |\lambda|} e_{\alpha_0(\lambda)} e_{\alpha_1(\lambda)} \cdots e_{\alpha_{n-1}(\lambda)},$$

where $\alpha_i(\lambda)$ is the number of parts of λ having size i, and $\alpha_0 = n - \sum_{i=1}^{n-1} \alpha_i$. (For example, if $n = 4$ and $\lambda = (3, 1, 1)$, then $\alpha_1 = 2$, $\alpha_0 = \alpha_3 = 1$, and $\alpha_2 = 0$.)

Proof. By Proposition 2,

$$\frac{x}{E(x)} \circ_q xK(qx) = x$$

if and only if

$$K(x) = E(x) \circ_q xK(x),$$

and, expanding the second equation, we have the recurrence

$$\begin{aligned}
k_n(q) &= [x^n]\left(1 + \sum_{r>0} e_r q^{r-1} x K(q^{r-1}x) \cdots qxK(qx)xK(x)\right) \\
&= [x^n] \sum_{r>0} e_r q^{\binom{r}{2}} x^r K(q^{r-1}x) \cdots K(qx)K(x) \\
&= \sum_{r>0} e_r q^{\binom{r}{2}} \left[x^{n-r}\right] K(q^{r-1}x) \cdots K(qx)K(x) \\
&= \sum_{r>0} e_r q^{\binom{r}{2}} \sum_{m_1+\cdots+m_r=n-r} \prod_i [x_i^m]\, K(q^{r-i}x) \\
&= \sum_{r>0} e_r q^{\binom{r}{2}} \sum_{m_1+\cdots+m_r=n-r} \prod_i q^{(r-i)m_i} k_{m_i}(q) \\
&= \sum_{r>0} e_r \sum_{m_1+\cdots+m_r=n-r} q^{\sum_i (m_i+1)(r-i)} \prod_i k_{m_i}(q)
\end{aligned}$$

It is clear that this recurrence has a unique solution (given the initial condition $k_0(q) = 1$), so we need to show that

$$k_n(q) = \sum_{\lambda \in \delta(n)} q^{\binom{n}{2} - |\lambda|} e_{\alpha_0(\lambda)} e_{\alpha_1(\lambda)} \cdots e_{\alpha_{n-1}(\lambda)}$$

satisfies this recurrence.

As with the recurrence for the q-Catalan numbers, we will show this recurrence holds by dividing the set of partitions $\lambda \in \delta(n)$ into classes. First put λ into the class $\mathcal{K}^{(r)}$ where $r = n - l(\lambda)$. Now we further subdivide each class $\mathcal{K}^{(r)}$ into classes $\mathcal{K}^{(r)}_{(m_1,\ldots,m_r)}$, one for each composition $m_1, \ldots, m_r$ of $n-r$. For each $\lambda \in \mathcal{K}^{(r)}$ and each i with $0 \le i \le r-1$, let $n_i(\lambda)$ be the largest number less than or equal to $n-r$ such that $\lambda_{n_i} > n - n_i - r + i$. (Recall that the n_i-th part of $\delta(n)$ has size $n - n_i$, so n_i is the highest row, not including the top r rows, with fewer than $r-i$ entries in $\delta(n) - \lambda$.) Notice that, by definition, $\lambda_{n-r} > 0 = n - (n-r) - r + 0$, so $n_0 = n - r$, and we set $n_r = 0$ by convention. Now let $m_i(\lambda) = n_{i-1}(\lambda) - n_i(\lambda)$, and place λ into $\mathcal{K}^{(r)}_{(m_1,\ldots,m_r)}$ accordingly; it is clear that $m_1, \ldots, m_r$ will be a composition of $n-r$. Figure 6 shows that $(13, 10, 7, 7, 6, 2, 2, 1)$ is in $\mathcal{K}^{(6)}_{(3,0,3,1,0,1)}$ (for $n = 14$).

For each partition λ in $\mathcal{K}^{(r)}_{(m_1,\ldots,m_r)}$, and each i with $1 \le i \le r$, we define partitions $\nu^{(i)}(\lambda)$ by letting $\nu^{(i)}_j(\lambda) = \lambda_{n_i(\lambda)+j} - \lambda_{n_{i-1}(\lambda)}$ for j such that $1 \le j \le m_i$.

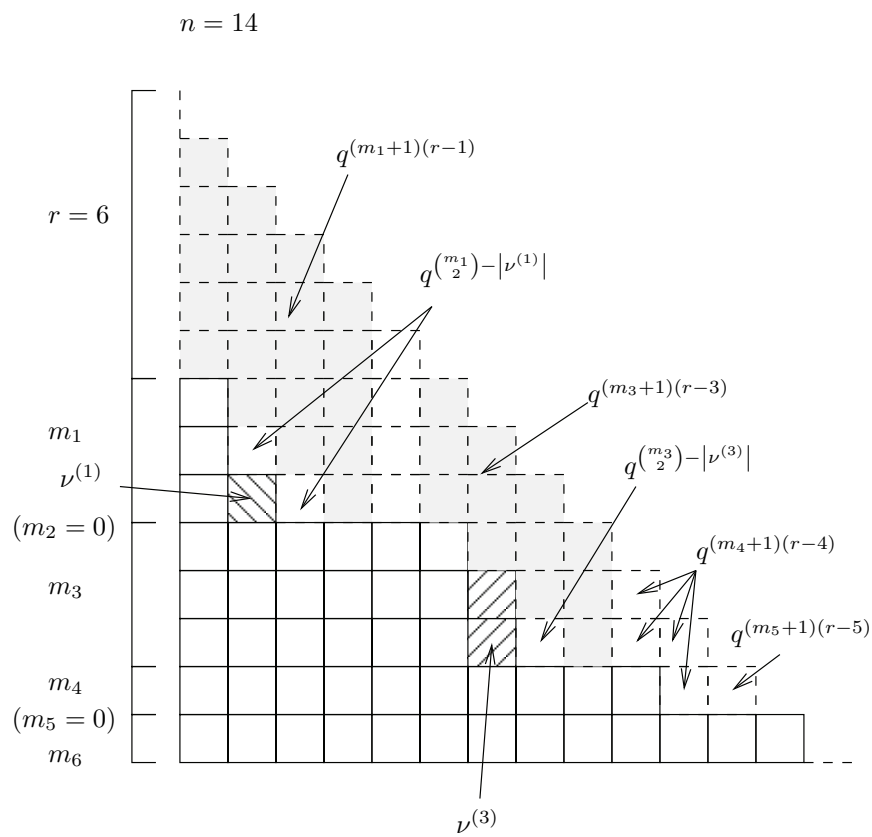


Figure 6. The q-Lagrange inversion recurrence illustrated for $\lambda = (13, 10, 7, 7, 6, 2, 2, 1)$.

Now Figure 6 shows that

$$\begin{aligned} q^{\binom{n}{2}-|\lambda|} & e_{\alpha_0(\lambda)} e_{\alpha_1(\lambda)} \cdots e_{\alpha_{n-1}(\lambda)} \\ &= q^{\sum_i (m_i+1)(r-i) + \sum_i \binom{m_i}{2} - |\nu^{(i)}(\lambda)|} e_{\alpha_0(\lambda)} \prod_{i=1}^{r} \prod_{j=\lambda_{n_i-1}}^{\lambda_{n_i-1}} e_{\alpha_j(\lambda)} \\ &= e_r q^{\sum_i (m_i+1)(r-i)} \prod_i q^{\binom{m_i}{2}-|\nu^{(i)}|} e_{\alpha_0(\nu^{(i)})} \cdots e_{\alpha_{m_i-1}(\nu^{(i)})}. \end{aligned}$$

Therefore, abbreviating $e_{\alpha_0(\nu^{(i)})} \cdots e_{\alpha_{m_i-1}(\nu^{(i)})}$ to $e_{\boldsymbol{\alpha}(\nu)}$,

$$\begin{aligned} k_n(q) &= \sum_{\lambda \in \delta(n)} q^{\binom{n}{2}-|\lambda|} e_{\alpha_0(\lambda)} e_{\alpha_1(\lambda)} \cdots e_{\alpha_{n-1}(\lambda)} \\ &= \sum_r e_r \sum_{m_1+\cdots+m_r=n-r} q^{\sum_i (m_i+1)(r-i)} \sum_{\lambda \in \mathcal{K}^{(r)}_{(m_1,\ldots,m_r)}} \prod_i q^{\binom{m_i}{2}-|\nu^{(i)}|} e_{\boldsymbol{\alpha}(\nu)} \\ &= \sum_{r>0} e_r \sum_{m_1+\cdots+m_r=n-r} q^{\sum_i (m_i+1)(r-i)} \prod_i k_{m_i}(q), \end{aligned}$$

which is the desired recurrence. $\square$

We now relate q-Lagrange inversion to q-Catalan numbers and to parking functions counted by frustration factor. Let $E(x) = 1/(1-x)$; then $e_n = 1$ for all n. We see that, by the above theorem,

$$C(x;q) := \sum_n C_n(q)x^n = \sum_n \left(\sum_{\lambda \in \delta(n)} q^{\binom{n}{2}-|\lambda|} \right) x^n$$

is the specified solution to the q-Lagrange inversion problem $\frac{x}{E(x)} \circ_q xK(qx) = x$, so we have

$$x(1-x) \circ_q xC(qx;q) = x.$$

As for parking functions, let $E(x) = e^x$; then $e_n = \frac{1}{n!}$. Now,

$$n!k_n(q) = \sum_{\lambda \in \delta(n)} q^{\binom{n}{2}-|\lambda|} \binom{n}{\alpha_0, \alpha_1, \ldots, \alpha_n} = P_n(q),$$

so the exponential generating function $P(x;q) = \sum_n P_n(q)x^n/n!$ solves $xe^{-x} \circ_q xP(qx;q) = x$. In particular, this shows, by setting $q = 1$, that $P_n(1) = (n+1)^{n-1}$, or that the number of parking functions for n cars is the same as the number of rooted forests on n vertices.

2.6. Exercises

(1) Prove Lemma 2.
(2) Prove Proposition 2.
(3) Use Theorem 6 to prove Theorem 4 by setting $q = 1$. (Hint: First show that, if $(\alpha_0, \ldots, \alpha_n) \in \mathbb{N}$ satisfy $\alpha_0 + \cdots + \alpha_n = n$, the sequence $(\alpha_0, \ldots, \alpha_n)$ has a unique rotation $(\beta_0, \ldots, \beta_n)$ such that there is a partition $\lambda \subseteq \delta(n)$ with $\alpha_i(\lambda) = \beta_i$ for all i.)
(4) Prove directly that there are $(n+1)^{n-1}$ parking functions on $\{1, \ldots, n\}$.
(5) Let S_n act on $\mathrm{PF}(n)$ as previously stated, and view $\mathbb{C} \cdot \mathrm{PF}(n)$ as an S_n representation graded by $\mathrm{wt}(f)$. Show that $\mathbb{C} \cdot \mathrm{PF}(n)$ is a direct sum of induced representations $\mathbb{C} \uparrow_{S_\mu}^{S_n}$ (which are respectively isomorphic to the representations $\mathbb{C} \cdot W_\mu$ introduced in Lecture 1) in which the generating function for the multiplicity of $\mathbb{C} \uparrow_{S_\mu}^{S_n}$ in the graded degrees is equal to the coefficient of $e_{\mu_1} \cdots e_{\mu_k}$ in $k_n(q)$.

LECTURE 3
Macdonald Polynomials

The Macdonald polynomials are a basis for the ring of symmetric functions over the base field $\mathbb{Q}(q,t)$. This basis has a number of useful and interesting properties, but, unfortunately, the polynomials are difficult to write out explicitly; indeed we will only have space to give an abstract definition and a number of their important properties, mostly without proof. These statements will require some notation and machinery, as well as motivation, from the general theory of symmetric functions, which we will now proceed to explain in the first part of this lecture.

Throughout this lecture, Λ_k denotes the ring of symmetric functions over the base field (or occasionally base ring) k.

3.1. Symmetric Function Bases and the Involution ω

In the first lecture, we saw two bases for the ring $\Lambda_{\mathbb{Q}}$, namely the monomial symmetric functions and the Schur functions. We now proceed to define three more bases.

The **complete homogeneous symmetric function** h_n is defined by $h_n := \sum_{|u|=n} m_\mu = s_{(n)}$. In other words, h_n is the sum of all the monomials of degree n. We can then define h_μ for all partitions μ by $h_\mu := h_{\mu_1} h_{\mu_2} \cdots h_{\mu_k}$. By Exercise 1.6(2), we have that $h_\mu = \sum_\lambda K_{\lambda\mu} s_\lambda$.

The **elementary symmetric function** e_n is defined by $e_n := m_{(1^n)} = s_{(1^n)}$; it is the sum of all square-free monomials of degree n. We define e_μ similarly by $e_\mu := e_{\mu_1} \cdots e_{\mu_k}$.

Finally, the **power sum symmetric function** p_n is defined by $p_n := m_{(n)} = \sum_i x_i^n$, with p_μ defined by $p_\mu := p_{\mu_1} \cdots p_{\mu_k}$.

As μ ranges over all partitions, the sets $\{h_\mu\}$, $\{e_\mu\}$, and $\{p_\mu\}$ are all bases of $\Lambda_{\mathbb{Q}}$.

Let $\omega\colon \Lambda_{\mathbb{Q}} \to \Lambda_{\mathbb{Q}}$ be the ring homomorphism sending e_n to h_n; since the e_n are algebraically independent and generate $\Lambda_{\mathbb{Q}}$, this map ω exists and is unique. Let λ' denote the partition **conjugate** to λ, that is, the partition whose diagram is the transpose of the diagram of λ. We will see later that ω is in fact an involution, that $\omega s_\lambda = s_{\lambda'}$, and that $\omega p_k = (-1)^k p_k$. In terms of the representation theory of S_n, ω corresponds to tensoring by the sign representation.

3.2. Plethystic Substitution

Let R be a ring, and designate some (possibly infinite) set $\{a_1, a_2, \ldots\}$ of elements of R as **indeterminates** with the property that, for each $k \in \mathbb{Z}_{\geq 0}$, there exists a ring homomorphism $\phi_k : R \to R$ with $\phi_k(a_i) = a_i^k$. Given $A \in R$ and some symmetric function f, we define $f[A]$, the **plethystic substitution** of A into f, as follows. First define $p_k[A] := \phi_k(A)$. Then let $p_\mu[A] := p_{\mu_1}[A]p_{\mu_2}[A]\cdots p_{\mu_l}[A]$. Finally, since the power sum symmetric functions are a basis for the ring of symmetric functions, we extend linearly to all symmetric functions.

The most trivial example is as follows. Let $R = \mathbb{C}[x_1, x_2, \ldots]$, and all x_i be indeterminates. Then for any symmetric function f, if $X = \sum_i x_i$, $f[X] = f(\mathbf{x})$. Less trivially, note that $p_k[-X] = -p_k(\mathbf{x}) = (-1)^k \omega p_k(\mathbf{x})$. Therefore, for any symmetric function f homogeneous of degree d, $f[-X] = (-1)^d \omega f(\mathbf{x})$.

It is customary to neglect to specify R and the set of indeterminates, and allow the set of indeterminates to be all the variables appearing in the expression. The ring R then will be the polynomial ring in the indeterminates, or the field of rational functions in the indeterminates, or some other similar object such as the formal power series ring or Laurent series ring in the indeterminates, subject to any relations we have imposed. If we do impose any relations, we must be careful not to impose relations which make some ϕ_k no longer well-defined; that is, if we impose some relation $x(a_1, a_2, \ldots) = y(a_1, a_2, \ldots)$, we must take care that $\phi_k(x(a_1, a_2, \ldots)) = \phi_k(y(a_1, a_2, \ldots))$ for every k.

For example, if t is an indeterminate, X is any expression in R, and f a symmetric function homogeneous of degree d, $f[tX] = t^d f[X]$. However, $f[-X] = (-1)^d \omega f[X] \neq (-1)^d f[X]$. This is because $t = -1$ is not an allowable relation for an indeterminate t, since $\phi_k(-1) = -1$ but $\phi_k(t) = t^d \neq -1$ for k even.

However, if we take $X = x_1 + x_2 + \cdots$ for an infinite set of variables, then the $p_k[X]$ are algebraically independent, so any plethystic equation which holds for X holds identically for any expression in place of X. The same is true if we have several independent infinite sets of variables X, Y, Z, and so on.

3.3. The Cauchy Kernel and Hall Inner Product

Let Ω denote the **Cauchy kernel**, which is the symmetric power series $\Omega := \sum_{n\geq 0} h_n$, so for $X = x_1 + x_2 + \cdots$, $\Omega[X] = \sum_{n\geq 0} h_n[X] = \prod_i 1/(1 - x_i)$. Notice that, if we have $Y = y_1 + y_2 + \cdots$ as well, $\Omega[X + Y] = \Omega[X]\Omega[Y]$, and since this identity holds with X and Y both sums of infinite sets of variables, it holds for any X and Y. In particular, $\Omega[X]\Omega[-X] = \Omega[0] = 1$, so $\Omega[-X] = \prod_i (1 - x_i) = \sum_n (-1)^n e_n(\mathbf{x})$. Taking the degree n piece of this identity, we have $h_n[-X] = (-1)^n e_n[X]$, which shows that, for f a homogeneous symmetric function of degree d, $\omega f[X] = (-1)^d f[-X]$.

Now we define the **Hall inner product** $\langle \cdot, \cdot \rangle$ on symmetric functions by declaring that $\langle h_\lambda, m_\mu \rangle = 0$ if $\lambda \neq \mu$, and $\langle h_\mu, m_\mu \rangle = 1$. This inner product has the following interpretation in terms of Ω.

Proposition 3. *Two bases $\{u_\lambda\}$ and $\{v_\lambda\}$ are dual (with respect to the Hall inner product) if and only if $\Omega[XY] = \sum_\lambda u_\lambda[X] v_\lambda[Y]$.*

Proof. First note that $\Omega[XY]=\prod_{i,j} 1/(1-x_iy_j)=\prod_i \Omega[x_iY]=\prod_i \sum_n x_i^n h_n[Y]=\sum_\lambda m_\lambda[X]h_\lambda[Y]$, and, by symmetry, $\Omega[XY]=\sum_\lambda h_\lambda[X]m_\lambda[Y]$. (This is known as the first **Cauchy formula**.)

Let $\langle\cdot,\cdot\rangle_{\mathbf{x}}$ denote the Hall inner product with respect to the x variables only. Then $\langle m_\lambda[X],\Omega[XY]\rangle_{\mathbf{x}} = \langle m_\lambda[X],h_\lambda[X]m_\lambda[Y]\rangle_{\mathbf{x}} = m_\lambda[Y]$, so linearity implies $\langle f[X],\Omega[XY]\rangle_{\mathbf{x}} = f[Y]$ for all f. If $\Omega[XY]=\sum_\lambda u_\lambda[X]v_\lambda[Y]$, then $\{u_\lambda\}$ and $\{v_\lambda\}$ are dual bases, because $\langle v_\lambda[X],\sum_\lambda u_\lambda[X]v_\lambda[Y]\rangle_{\mathbf{x}} = v_\lambda[Y]$. Since the Hall inner product is non-degenerate, the only way to have $\langle v_\lambda[X],g[XY]\rangle_{\mathbf{x}} = v_\lambda[X]$ for all λ is to have $g=\Omega$, which proves the reverse direction. □

It can be shown, for example by using the Robinson-Schensted-Knuth correspondence, that $\Omega[XY]=\sum_\lambda s_\lambda[X]s_\lambda[Y]$, so the Schur functions are an orthonormal basis for $\Lambda_{\mathbb{Q}}$ under this inner product. Therefore, in terms of the representation theory of S_n, we therefore have that $\dim(\mathrm{Hom}^{S_n}(V,W)) = \langle F_V, F_W\rangle$ for any two representations V and W.

Finally, note that ω is an isometry with respect to the Hall inner product. In other words, $\langle \omega f,\omega g\rangle = \langle f,g\rangle$ for any symmetric functions f and g.

3.4. Dominance Ordering

The final definition we need is a partial order on partitions known as **dominance order**. Being the only order relation we will use on partitions, we will simply denote it by $\leq$. We say $\lambda\leq\mu$ if $\lambda_1+\cdots+\lambda_k \leq \mu_1+\cdots+\mu_k$ for every positive integer k.

Proposition 4. *If $\mu\not\leq\lambda$, then $K_{\lambda\mu}=0$.*

Proof. Let T be a SSYT of shape λ. Note that, for any k, any box of T with a label $i<k$ must occur in one of the first k rows. Therefore, if μ is the content of T, $\mu_1+\cdots+\mu_k \leq \lambda_1+\cdots+\lambda_k$. Since this holds for any k, we have that, if $K_{\lambda\mu}\neq 0$, then $\mu\leq\lambda$, proving the proposition. □

We will also need the following proposition, whose (not entirely trivial) proof is left as an exercise.

Proposition 5. *$\lambda\leq\mu$ iff $\lambda'\leq\mu'$.*

3.5. Definition of Macdonald Polynomials

Now we are ready to define the Macdonald polynomials and their predecessors, the Hall-Littlewood polynomials.

Theorem-Definition 1. *The ring $\Lambda_{\mathbb{Q}(t)}$ has a unique basis $\widetilde{H}_\mu(\mathbf{x};t)$ characterized by*

(1) $\widetilde{H}_\mu(\mathbf{x};t)\in\mathbf{Q}(t)\{s_\lambda|\lambda\geq\mu\}$
(2) $\widetilde{H}_\mu[X(1-t);t]\in\mathbf{Q}(t)\{s_\lambda|\lambda\geq\mu'\}$
(3) $\widetilde{H}_\mu[1;t]=1$.

These polynomials are known as the **Hall-Littlewood polynomials**. *Moreover, $\widetilde{H}_\mu(\mathbf{x};t)=\sum_\lambda \widetilde{K}_{\lambda\mu}(t)s_\lambda(\mathbf{x})$, where the $\widetilde{K}_{\lambda\mu}(t)$ are as in Lecture 1.*

Theorem-Definition 2. *The ring $\Lambda_{\mathbb{Q}(q,t)}$ has a unique basis $\widetilde{H}_\mu(\mathbf{x};q,t)$ characterized by*

(1) $\widetilde{H}_\mu[X(1-q);q,t] \in \mathbf{Q}(q,t)\,\{s_\lambda | \lambda \geq \mu\}$
(2) $\widetilde{H}_\mu[X(1-t);q,t] \in \mathbf{Q}(q,t)\,\{s_\lambda | \lambda \geq \mu'\}$
(3) $\widetilde{H}_\mu[1;q,t] = 1$.

These polynomials are known as the **Macdonald polynomials**.

Now we can define a two variable q,t-analog of the Kostka numbers.

Definition 2. *Since the Schur functions are a basis for* $\Lambda_{\mathbb{Q}(q,t)}$,

$$\widetilde{H}_\mu(\mathbf{x};q,t) = \sum_\lambda \widetilde{K}_{\lambda\mu}(q,t) s_\lambda(\mathbf{x})$$

for some rational functions $\widetilde{K}_{\lambda\mu}(q,t)$. *These are the* q,t**-Kostka numbers**.

We do not have time to give a complete proof of these theorems. We will however prove uniqueness and give an indication of the main ingredients in the existence proof. Details can be found in [**15, 21**].

Pick an arbitrary degree n. Suppose $\left\{\widetilde{H}_\mu(\mathbf{x};q,t)\right\}_{|\mu|=n}$ and $\left\{\widetilde{H}'_\mu(\mathbf{x};q,t)\right\}_{|\mu|=n}$ are two bases of the degree n part of $\Lambda_{\mathbb{Q}(q,t)}$ characterized by the three given conditions. Order these bases by choosing the same refinement of dominance order for both. Then there will be a transition matrix, which we denote T, which tells us how to write elements of one basis with respect to the other. Our goal is to show that T must be the identity matrix.

Define the operator $\Pi_{(1-q)}$ on $\Lambda_{\mathbb{Q}(q,t)}$ by $\Pi_{(1-q)}f = f[X(1-q)]$. Also, define $\Pi_{1/(1-q)}$ by $\Pi_{1/(1-q)}f = f[X/(1-q)]$, and similarly define $\Pi_{(1-t)}$ and $\Pi_{1/(1-t)}$. By checking for $f = p_n$, we see that $\Pi_{(1-q)}$ and $\Pi_{1/(1-q)}$ are clearly inverse to each other, as are $\Pi_{(1-t)}$ and $\Pi_{1/(1-t)}$.

Let P and P' be transition matrices respectively expressing $\left\{\widetilde{H}_\mu(\mathbf{x};q,t)\right\}_{|\mu|=n}$ and $\left\{\widetilde{H}'_\mu(\mathbf{x};q,t)\right\}_{|\mu|=n}$ in terms of the basis $\{s_\lambda[X/(1-q);q,t]\}_{|\lambda|=n}$. By applying $\Pi_{1/(1-q)}$ to condition (1), we see that both P and P' are upper triangular. Therefore, $T = P^{-1}P'$ is upper triangular. Rewriting condition (2) as $\widetilde{H}_\mu[X(1-q);q,t] \in \mathbf{Q}(q,t)\,\{s_{\lambda'} | \lambda \leq \mu\}$ and applying $\Pi_{1/(1-t)}$, we have that the transition matrices expressing $\left\{\widetilde{H}_\mu(\mathbf{x};q,t)\right\}_{|\mu|=n}$ and $\left\{\widetilde{H}'_\mu(\mathbf{x};q,t)\right\}_{|\mu|=n}$ in terms of the basis $\{s_{\lambda'}[X/(1-t);q,t]\}_{|\lambda|=n}$ are lower triangular, so T is also lower triangular, and therefore diagonal. Condition (3) then forces T to be the identity.

Now we outline the ideas behind the existence proof. Let $X = x_1 + x_2 + \cdots$ as usual. Define a linear operator Δ_0 on $\Lambda_{\mathbb{Q}(q,t)}$ by

$$\Delta_0 f = [u^0](f[X + (1-q)(1-t)u^{-1}]\Omega[-uX]),$$

and define the linear operator Δ by

$$\Delta f = \frac{f - \Delta_0 f}{(1-q)(1-t)}.$$

The operator Δ is known as the **Macdonald operator**, and the existence of Macdonald polynomials is proved by showing that $\Pi_{(1-q)}\Delta\Pi_{1/(1-q)}$ is upper triangular with respect to $\{s_\lambda\}_{|\lambda|=n}$, with diagonal entries $B_\lambda(q,t) := \sum_{(i,j)\in\lambda} t^i q^j = \sum_i t^{i-1}(1-q^{\lambda_i})/(1-q)$. (Note the convention is that powers of q increase as one

moves to the right and powers of t increase as one moves up.) Therefore, eigenfunctions for Δ must satisfy (1). Furthermore, Δ and B_λ are symmetric with respect to simultaneously exchanging q and t and exchanging μ and μ', so these eigenfunctions must satisfy (2). Condition (3) is just a scalar normalization factor. Note that, in particular,

$$\Delta \widetilde{H}_\mu(\mathbf{x}; q, t) = B_\mu(q, t)\widetilde{H}_\mu(\mathbf{x}; q, t).$$

Some properties of Macdonald polynomials are easy to see from the definition and theorem. First, $\widetilde{H}_\mu(\mathbf{x}; 0, t) = \widetilde{H}_\mu(\mathbf{x}; t)$, and $\widetilde{K}_{\lambda\mu}(0, t) = \widetilde{K}_{\lambda\mu}(t)$. In other words, setting $q = 0$ in a Macdonald polynomial recovers the corresponding Hall-Littlewood polynomial. Also, the definition looks the same when we both swap q and t and swap μ and μ', so by uniqueness, $\widetilde{H}_\mu(\mathbf{x}; q, t) = \widetilde{H}_{\mu'}(\mathbf{x}; t, q)$. In particular, if $\mu = \mu'$, then $\widetilde{H}_\mu(\mathbf{x}; q, t)$ is symmetric under switching q and t.

From the definition it is possible to compute $\widetilde{H}_{(n)}(\mathbf{x}; q, t)$. Every partition dominates $(1^n) = (n)'$, so the second condition is vacuous. The first condition states that $\widetilde{H}_{(n)}[X(1-q); q, t] = fh_n(\mathbf{x})$ for some $f \in \mathbb{Q}(q, t)$, or, equivalently, that $\widetilde{H}_{(n)}(\mathbf{x}; q, t) = fh_n[X/(1-q); q, t]$ for some f. Now we use the third condition to solve for f; namely $f = \widetilde{H}_{(n)}[1; q, t]/h_n[1/(1-q)]$. Note that

$$h_n[1/(1-q)] = h_n(1, q, q^2, \ldots) = \sum_{l(\lambda)\le n} q^{|\lambda|} = \sum_{\lambda_1 \le n} q^{|\lambda|} = \frac{1}{(1-q)(1-q^2)\cdots(1-q^n)}.$$

Therefore,

$$\widetilde{H}_{(n)}(\mathbf{x}; q, t) = (1-q)\cdots(1-q^n)h_n\left[\frac{X}{1-q}\right].$$

Next we compute $\widetilde{H}_\mu(\mathbf{x}; q, 1)$ for all μ. First, note that $\Delta \mid_{t=1}$ is a derivation on $\Lambda_{\mathbb{Q}(q)}$; that is, for any $f, g \in \Lambda_{\mathbb{Q}(q)}$, $\Delta(fg) \mid_{t=1} = f(\Delta(g) \mid_{t=1}) + (\Delta(f) \mid_{t=1})g$. Since $\Delta \mid_{t=1}$ is linear on $\Lambda_{\mathbb{Q}(q)}$, this statement can be proven by showing that it holds when $f = p_\mu$ and $g = p_\nu$, and this is left as an exercise. Now note that $\widetilde{H}_{(n)}(\mathbf{x}; q, t) = \widetilde{H}_{(n)}(\mathbf{x}; q, 1)$, so we have that

$$\Delta \widetilde{H}_{(n)}(\mathbf{x}; q, 1) = B_{(n)}(q, 1)\widetilde{H}_{(n)}(\mathbf{x}; q, 1) = (1-q^n)/(1-q)\widetilde{H}_{(n)}(\mathbf{x}; q, 1).$$

Using that $\Delta \mid_{t=1}$ is a derivation on $\Lambda_{\mathbb{Q}(q)}$,

$$\Delta \mid_{t=1} (\widetilde{H}_{\mu_1}(\mathbf{x}; q, 1)\cdots\widetilde{H}_{\mu_k}(\mathbf{x}; q, 1)) = B_\mu(q, 1)(\widetilde{H}_{\mu_1}(\mathbf{x}; q, 1)\cdots\widetilde{H}_{\mu_k}(\mathbf{x}; q, 1)).$$

Now the uniqueness of Macdonald polynomials tells us that

$$\widetilde{H}_\mu(\mathbf{x}; q, 1) = \prod_i \widetilde{H}_{\mu_i}(\mathbf{x}; q, 1).$$

Finally, we compute $\widetilde{H}_\mu(\mathbf{x}; 1, 1)$ for all μ. First,

$$\widetilde{H}_{(1)}(\mathbf{x}; q, 1) = (1-q)h_1[\frac{X}{1-q}; q, 1] = h_1(\mathbf{x}),$$

so

$$\widetilde{H}_{(1)}(\mathbf{x}; 1, 1) = h_1(\mathbf{x}).$$

In particular, it follows from the previous paragraph, specialized at $q = 1$, that

$$\widetilde{H}_{(1^n)}(\mathbf{x}; 1, 1) = \widetilde{H}_{(1)}(\mathbf{x}; 1, 1)^n = h_{(1^n)}(\mathbf{x}).$$

Now, since

$$\widetilde{H}_{(n)}(\mathbf{x}; q, 1) = \widetilde{H}_{(1^n)}(\mathbf{x}; 1, q),$$

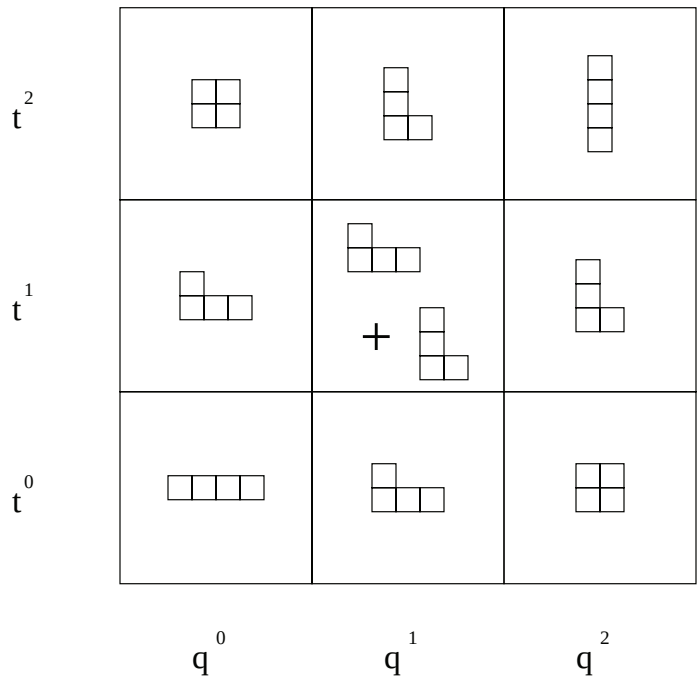


Figure 1. $\widetilde{H}_{(22)}(\mathbf{x};q,t)$

we get that

$$\widetilde{H}_\mu(\mathbf{x};1,1)=\prod_i \widetilde{H}_{(\mu_i)}(\mathbf{x};1,1)=h_{(1^{|\mu|})}(\mathbf{x}).$$

Note in particular this does not depend on the partition μ as long as $|\mu|=n$.

3.6. More Properties of Macdonald Polynomials

The theory of Macdonald polynomials is a large subject fit for another series of lectures, so we will not be able to cover most of it. Instead we merely state here, without proof, a few facts which we will need in subsequent lectures.

To help in understanding these properties, we provide one small example. It can be calculated that $\widetilde{H}_{(22)}(\mathbf{x};q,t)=s_{(4)}(\mathbf{x})+(q+t+qt)s_{(31)}(\mathbf{x})+(q^2+t^2)s_{(22)}(\mathbf{x})+(q^2t+qt^2+qt)s_{(211)}(\mathbf{x})+q^2t^2s_{(1111)}(\mathbf{x})$. We can draw this as a picture as in Figure 1.

First we describe the action of ω. For a partition μ, let $n(\mu)=\sum_i(i-1)\mu_i$. Then

$$\omega\widetilde{H}_\mu(\mathbf{x};q,t)=t^{n(\mu)}q^{n(\mu')}\widetilde{H}_\mu(\mathbf{x};q^{-1},t^{-1}).$$

Here $n(\mu)$ and $n(\mu')$ appear because they are respectively the top degrees of t and q found in $\widetilde{H}_\mu(\mathbf{x};q,t)$, so the multiplication by $t^{n(\mu)}q^{n(\mu')}$ normalizes the right hand side to be a polynomial in q and t with nonzero constant term.

There is also the **Macdonald specialization formula**, which states that

$$\widetilde{H}_\mu[1-u;q,t]=\Omega[-uB_u]=\prod_{(i,j)\in\mu}(1-uq^jt^i).$$

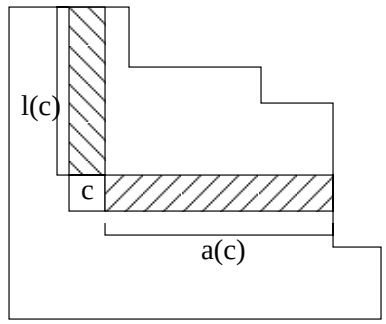


Figure 2. Arm and Leg of a cell $c \in \lambda$.

We can use this formula to derive $\widetilde{K}_{\lambda\mu}$ when λ a hook shape, that is, if $\lambda = (n-r, 1^r)$ for some r. Specifically,

$$\widetilde{K}_{(n-r,1^r),\mu} = e_r[B_\mu - 1].$$

Finally, we describe a q,t-analog of the Hall inner product and give a corresponding Cauchy formula for Macdonald polynomials. Define

$$\langle f, g\rangle_* := \langle f[X(1-q); q, t], \omega g[X(1-t); q, t]\rangle,$$

where the inner product on the right is the usual Hall inner product (with respect to the x variables). Then $\langle \widetilde{H}_\lambda(\mathbf{x}; q, t), \widetilde{H}_\mu(\mathbf{x}; q, t)\rangle_* = \langle \widetilde{H}_\lambda[X(1-q); q, t], \omega \widetilde{H}_\mu[X(1-t); q, t]\rangle$, and, expanding both parts of the inner product in terms of the orthonormal basis of Schur functions, we see that $\langle \widetilde{H}_\lambda(\mathbf{x}; q, t), \widetilde{H}_\mu(\mathbf{x}; q, t)\rangle_* \neq 0$ iff $\{\nu : \nu \geq \lambda\} \cap \{\nu : \nu' \geq \mu'\} \neq \emptyset$ iff $\lambda \leq \mu$. By symmetry of the inner product (which follows from ω being an isometry and $\Pi_{(1-q)}$ and $\Pi_{1/(1-q)}$ being adjoint), we also have $\lambda \geq \mu$, so $\langle \widetilde{H}_\lambda(\mathbf{x}; q, t), \widetilde{H}_\mu(\mathbf{x}; q, t)\rangle_* = 0$ if $\lambda \neq \mu$.

Let c be a cell in the diagram of some partition λ. The **arm** and **leg** of c, respectively denoted $a(c)$ and $l(c)$, are the number of boxes strictly to the right of, and respectively the number of boxes strictly above, the box c in the diagram of λ, as illustrated in Figure 2. It turns out that

$$\langle \widetilde{H}_\mu(\mathbf{x}; q, t), \widetilde{H}_\mu(\mathbf{x}; q, t)\rangle_* = t^{n(\mu)} q^{n(\mu')} \prod_{c\in\mu} (1 - t^{l(c)+1} q^{-a(c)})(1 - t^{-l(c)} q^{a(c)+1}).$$

Therefore, we have

$$\Omega[XY] = \sum_\mu \frac{t^{-n(\mu)} q^{-n(\mu')} \widetilde{H}_\mu[X(1-q); q, t] \omega \widetilde{H}_\mu[Y(1-t); q, t]}{\prod_{c\in\mu}(1 - t^{l(c)+1} q^{-a(c)})(1 - t^{-l(c)} q^{a(c)+1})},$$

or, after substituting $X/(1-q)$ for X and $-Y/(1-t)$ for Y, taking the degree n piece, and multiplying both sides by $(-1)^n$,

$$e_n\left[\frac{XY}{(1-q)(1-t)}\right] = \sum_{|\mu|=n} \frac{t^{-n(\mu)} q^{-n(\mu')} \widetilde{H}_\mu(\mathbf{x}; q, t) \widetilde{H}_\mu(\mathbf{y}; q, t)}{\prod_{c\in\mu}(1 - t^{l(c)+1} q^{-a(c)})(1 - t^{-l(c)} q^{a(c)+1})}.$$

3.7. Exercises

(1) Let $X = x_1 + x_2 + \cdots$ and $Y = y_1 + y_2 + \cdots$. Express $e_n[X - Y]$ in terms of symmetric functions separately in the x and y variables.

(2) Show that the graded Frobenius series of $\mathbb{C}[x_1, \cdots, x_n]$ as an S_n representation, that is, $\sum_d t^d F_{\mathbb{C}[x_1,\cdots,x_n]_d}$ (where $\mathbb{C}[x_1, \cdots, x_n]_d$ denotes the polynomials of degree d), is $h_n[X/(1-t)]$.

(3) Prove Proposition 5.

(4) Show that $\Delta \mid_{t=1}$ is a derivation on $\Lambda_{\mathbb{Q}(q)}$ by showing that

$$\Delta(p_\mu p_\nu) \mid_{t=1} = \Delta(p_\mu) \mid_{t=1} p_\nu + p_\mu \Delta(p_\nu) \mid_{t=1} .$$

(5) Show that

$$\omega \widetilde{H}_\mu(\mathbf{x}; q, t) = t^{n(\mu)} q^{n(\mu')} \widetilde{H}_\mu(\mathbf{x}; q^{-1}, t^{-1}).$$

(You will need to use the Macdonald specialization formula.)

(6) Use the Macdonald specialization formula to show that $\widetilde{K}_{\lambda\mu}(q,t) = e_r[B_\mu - 1]$ when $\lambda = (n-r, 1^r)$.

(7) (a) Prove that for any expression A

$$\frac{e_n[(1-u)A]}{1-u} \mid_{u=1} = (-1)^{n-1} p_n[A].$$

(b) For the Macdonald operator Δ, show that

$$\Delta\left((-1)^{n-1} p_n\left[\frac{X}{(1-q)(1-t)}\right]\right) = \frac{e_n[X]}{(1-q)(1-t)}.$$

(c) Let $\Pi_\mu(q,t) = \prod_{(i,j)\in\mu\setminus(0,0)}(1 - q^j t^i)$. Now use parts (a) and (b), the Macdonald specialization, and the Cauchy formula to prove that

$$e_n(\mathbf{x}) = \sum_{|\mu=n|} \frac{t^{-n(\mu)} q^{-n(\mu')} (1-q)(1-t) \Pi_\mu(q,t) B_\mu(q,t) \widetilde{H}_\mu(\mathbf{x}; q, t)}{\prod_{c\in\mu}(1 - t^{l(c)+1} q^{-a(c)})(1 - t^{-l(c)} q^{a(c)+1})}.$$

LECTURE 4

Connecting Macdonald Polynomials and q-Lagrange Inversion; (q,t)-Analogs

In this lecture we will take expressions which at first appear to be relatively unmotivated symmetric functions and show that in fact they are a (q,t)-analog of the $k_n(q)$ which solved the q-Lagrange inversion problem in Lecture 2. Most of the lecture will be devoted to this proof which includes some complicated calculations. They have been included because they reflect many of the calculational techniques which are important in this subject. The main general reference for this section is [**7**].

4.1. The Operator ∇ and a (q,t)-Analog of $k_n(q)$

Recall from Lecture 3 and specifically Exercise 3.7(7) that

$$e_n(\mathbf{x}) = \sum_{|\mu=n|} \frac{t^{-n(\mu)}q^{-n(\mu')}(1-q)(1-t)\Pi_\mu(q,t)B_\mu(q,t)\widetilde{H}_\mu(\mathbf{x};q,t)}{\prod_{c\in\mu}(1-t^{l(c)+1}q^{-a(c)})(1-t^{-l(c)}q^{a(c)+1})},$$

where $B_\mu(q,t) = \sum_{(i,j)\in\lambda} t^i q^j$, $\Pi_\mu(q,t) = \Omega[1-B_\mu] = \prod_{(i,j)\in\mu\setminus(0,0)}(1-t^iq^j)$, and, for c a cell in the diagram of μ, $a(c)$ and $l(c)$ denote respectively the arm and leg of c.

Define an operator ∇ on $\Lambda_{\mathbb{Q}(q,t)}$ by letting

$$\nabla\widetilde{H}_\mu := t^{n(\mu)}q^{n(\mu')}\widetilde{H}_\mu$$

and extending by linearity. Applying this operator to the above expansion of e_n gives

$$\nabla e_n = \sum_{|\mu=n|} \frac{(1-q)(1-t)\Pi_\mu(q,t)B_\mu(q,t)\widetilde{H}_\mu(\mathbf{x};q,t)}{\prod_{c\in\mu}(1-t^{l(c)+1}q^{-a(c)})(1-t^{-l(c)}q^{a(c)+1})}.$$

Now we calculate $\langle\nabla e_n, e_n\rangle$. Notice that

$$\begin{aligned}
\langle\widetilde{H}_\mu(\mathbf{x};q,t), e_n\rangle &= \langle\widetilde{H}_\mu(\mathbf{x};q,t), s_{(1^n)}\rangle \\
&= \widetilde{K}_{(1^n),\mu} \\
&= e_{n-1}[B_\mu - 1] \\
&= e_{n-1}[\sum_{(i,j)\in\mu\setminus(0,0)} t^iq^j] = \prod_{(i,j)\in\mu\setminus(0,0)} t^iq^j = t^{n(\mu)}q^{n(\mu')},
\end{aligned}$$

where the third equality comes from the Macdonald specialization formula as discussed in Lecture 3. Therefore,

$$\langle \nabla e_n, e_n \rangle = \sum_{|\mu=n|} \frac{t^{n(\mu)} q^{n(\mu')} (1-q)(1-t)\Pi_\mu(q,t) B_\mu(q,t)}{\prod_{c\in\mu}(1-t^{l(c)+1}q^{-a(c)})(1-t^{-l(c)}q^{a(c)+1})}.$$

Define $C_n(q,t)$ to be this rational function $\langle \nabla e_n, e_n \rangle$. It turns out that $C_n(q,t)$ is a polynomial with positive integer coefficients, and that $C_n(q,1) = C_n(q)$, the q-analog of the Catalan numbers discussed in Lecture 2. Furthermore, $C_n(q,t)$ is symmetric under exchanging q and t; that is, $C_n(q,t) = C_n(t,q)$. For example, $C_3(q,t) = q^3+q^2t+qt+qt^2+t^3$, and specializing to $t=1$ gives $C_3(q) = q^3+q^2+2q+1$ which is what we had earlier. Therefore, it makes sense to think of $C_n(q,t)$ as a (q,t)-analog of the Catalan numbers.

Now notice that $C_n(q) = k_n(q)|_{e_k\mapsto 1}$, as we saw at the end of Lecture 2. Since $h_n = \sum_{|\mu|=n} m_\mu$ and $\{h_\mu\}$ and $\{m_\mu\}$ are dual bases, $\langle h_\mu, h_n\rangle = 1$ for all μ, and consequently, since ω is an isometry with respect to the Hall inner product, $\langle e_\mu, e_n\rangle = 1$ for all μ. Therefore, if we pretend that the e_k in $k_n(q)$ actually stand for elementary symmetric functions, then $C_n(q) = \langle k_n(q), e_n\rangle$.

Comparing the equations $C_n(q) = \langle k_n(q), e_n\rangle$ and $C_n(q,t) = \langle \nabla e_n, e_n\rangle$ hints at a possible connection between $k_n(q)$ and ∇e_n. It turns out that there is indeed a connection given by the following theorem, which we will spend most of the remainder of this lecture proving.

Theorem 7. *Interpreting the e_k in $k_n(q)$ as elementary symmetric functions, we have that*

$$\nabla e_n|_{t=1} = k_n(q).$$

Before we go into the proof, let us mention two corollaries giving (q,t)-analogs of our main examples from Lecture 2. The first corollary follows from the discussion above. To prove the second, recall that $h_{(1^n)} = \sum_{|\mu|=n} \binom{n}{\mu_1,\ldots,\mu_l} m_\mu$, so $\langle e_\mu, e_{(1^n)}\rangle = \binom{n}{\mu_1,\ldots,\mu_l}$.

Corollary 1. *Define $C_n(q,t)$, as above, by $C_n(q,t) = \langle \nabla e_n, e_n\rangle$. Then*

$$C_n(q,1) = C_n(q).$$

Corollary 2. *Define $P_n(q,t)$ by $P_n(q,t) := \langle \nabla e_n, e_{(1^n)}\rangle$. Then*

$$P_n(q,1) = P_n(q) = \sum_{\lambda\in\delta(n)} q^{\binom{n}{2}-|\lambda|} \binom{n}{\alpha_0,\alpha_1,\ldots,\alpha_n}.$$

4.2. Proof of Theorem 7

Let $K(z) = \sum_n k_n(q) z^n$, and $E(z) = \sum_n e_n z^n$. Identifying the e_n with the elementary symmetric functions $e_n(\mathbf{x})$, we have $E(z) = \omega\Omega[zX] = \sum_n e_n[X] z^n$. For convenience, let us define $E := \omega\Omega$ as a symmetric power series, so $E(z) = E[zX]$. Notice that $E[zX] = \prod_i (1+zx_i)$, so $E[z(A+B)] = E[zA]E[zB]$ for any expressions A and B. Consequently, since $1 = E[0] = E[z(A-A)] = E[zA]E[-zA]$, we have that $E[-zA] = 1/E[zA]$.

Now recall that $K(z)$ is in fact the solution to the q-Lagrange inversion problem $z/E[zX] \circ_q zK(qz) = z$, or, equivalently by Theorem 5, $z = zK(qz) \circ_{q^{-1}} z/E[zX]$.

For convenience, let g_n be the coefficient of z^n in $zK(qz)$, so $g_n = q^{n-1}k_{n-1}(q)$. We can now calculate that

$$\begin{aligned} z &= zK(qz) \circ_{q^{-1}} z/E[zX] \\ &= \sum_n g_n \frac{z}{E[zX]} \frac{q^{-1}z}{E[q^{-1}zX]} \cdots \frac{q^{-(n-1)}z}{E[q^{-(n-1)}zX]} \\ &= \sum_n g_n z^n q^{-\binom{n}{2}} \frac{1}{E[z(1+q^{-1}+\cdots+q^{-(n-1)})X]} \\ &= \sum_n g_n z^n q^{-\binom{n}{2}} \frac{1}{E[z(1-q^{-n})X/(1-q^{-1})]} \\ &= \sum_n g_n z^n q^{-\binom{n}{2}} \frac{E[zq^{-n}X/(1-q^{-1})]}{E[zX/(1-q^{-1})]}. \end{aligned}$$

Hence

$$\sum_n g_n z^n q^{-\binom{n}{2}} E\left[\frac{q^{-n}zX}{1-q^{-1}}\right] = zE\left[\frac{zX}{1-q^{-1}}\right]. \tag{1}$$

For any series $\Psi(z) = \sum_n \Psi_n z^n$, define ${}^\vee\Psi(z) = \sum_n \Psi_n q^{\binom{n}{2}} z^n = \Psi(z) \circ_q z$. Now we need a lemma about the behavior of ${}^\vee$.

Lemma 3. *We have the identities*

(1) ${}^\vee(z^n q^{-\binom{n}{2}} \Psi(q^{-n}z)) = z^n {}^\vee\Psi(qz)$
(2) ${}^\vee(z\Psi(z)) = z^\vee\Psi(qz)$.

Proof. By linearity, it suffices to prove this for $\Psi(z) = z^r$ (for all r) in both cases. We see that

$$\begin{aligned} {}^\vee(z^n q^{-\binom{n}{2}} (q^{-n}z)^r) &= q^{\binom{n+r}{2}} q^{-\binom{n}{2}-nr} z^{n+r} \\ &= q^{\binom{r}{2}} z^{n+r} \\ &= z^n {}^\vee z^r. \end{aligned}$$

Also,

$$\begin{aligned} {}^\vee(z^{r+1}) &= q^{\binom{r+1}{2}} z^{r+1} \\ &= zq^{\binom{r}{2}} q^r z^r \\ &= z^\vee((qz)^r). \end{aligned}$$

□

Apply the operator ${}^\vee$ to both sides of equation 1. Using the first part of the lemma on the left hand side and the second part on the right hand side, we get

$$\sum_n g_n z^n {}^\vee E\left[\frac{zX}{1-q^{-1}}\right] = z^\vee E\left[\frac{qzX}{1-q^{-1}}\right].$$

Hence,

$$zK(qz) = G(z) = \sum_n g_n z^n = \frac{z^\vee E\left[qzX/(1-q^{-1})\right]}{{}^\vee E\left[zX/(1-q^{-1})\right]},$$

and, substituting $q^{-1}z$ for z,

$$K(z) = \sum_n k_n(q) z^n = \frac{{}^\vee E\left[zX/(1-q^{-1})\right]}{{}^\vee E\left[q^{-1}zX/(1-q^{-1})\right]}.$$

Specializing to $q = 1$ appropriately here gives us $k_n = [z^n]\frac{E(z)^{n+1}}{n+1}$, that is, the last formula is actually a q-analog of the classical formula in Theorem 4.

To understand $K(z)$ more explicitly, we make two further definitions; neither is strictly necessary but they will both make our notation significantly more compact. First, for each partition μ, define the symmetric functions $f_\mu(\mathbf{x})$ (sometimes known as the **forgotten symmetric functions**) by the identity

$$e_n[XY] := \sum_{|\mu|=n} h_\mu[X] f_\mu[Y].$$

Equivalently, we could also define f_μ as the dual basis to $\{e_\mu\}$ under the Hall inner product, or by letting $f_\mu := \omega m_\mu$. Secondly, we introduce a fictitious alphabet A such that

$$h_n[A] := q^{\binom{n}{2}} h_n[X/(1-q)].$$

Now we produce the following identity to simplify our expression for $K(z)$:

$$\begin{aligned}
{}^\vee E\left[\frac{q^{-1}zX}{1-q^{-1}}\right] &= {}^\vee E\left[\frac{-zX}{1-q}\right] \\
&= {}^\vee\sum_n (-1)^n \omega e_n[X/(1-q)] z^n \\
&= {}^\vee\sum_n h_n[X/(1-q)](-z)^n \\
&= \sum_n q^{\binom{n}{2}} h_n[X/(1-q)](-z)^n \\
&= \sum_n h_n[A](-z)^n = \sum_n e_n[-zA] = \sum_n E[-zA] = 1/E[zA].
\end{aligned}$$

From this identity, our previous equation for $K(z)$ reduces to

$$K(z) = \frac{E[zA]}{E[qzA]} = E[z(1-q)A] = \sum_\mu h_\mu[A] f_\mu[z(1-q)],$$

the last equality coming from our definition of f_μ. By our definition of A,

$$K(z) = \sum_\mu \left(\prod_{i=1}^{l(\mu)} q^{\binom{\mu_i}{2}} h_{\mu_i}\left[X/(1-q)\right] \right) f_\mu[z(1-q)].$$

Extracting on both sides the coefficient of z^n, we end up with

$$k_n(q) = \sum_{|\mu|=n} q^{n(\mu')} h_\mu[X/(1-q)] f_\mu[1-q].$$

Finally, recall from Lecture 3 that

$$\widetilde{H}_\mu(\mathbf{x}; q, 1) = \prod_i \widetilde{H}_{\mu_i}(\mathbf{x}; q, 1) = \left(\prod_i (1-q)\cdots(1-q^{\mu_i}) \right) h_\mu\left[X/(1-q)\right],$$

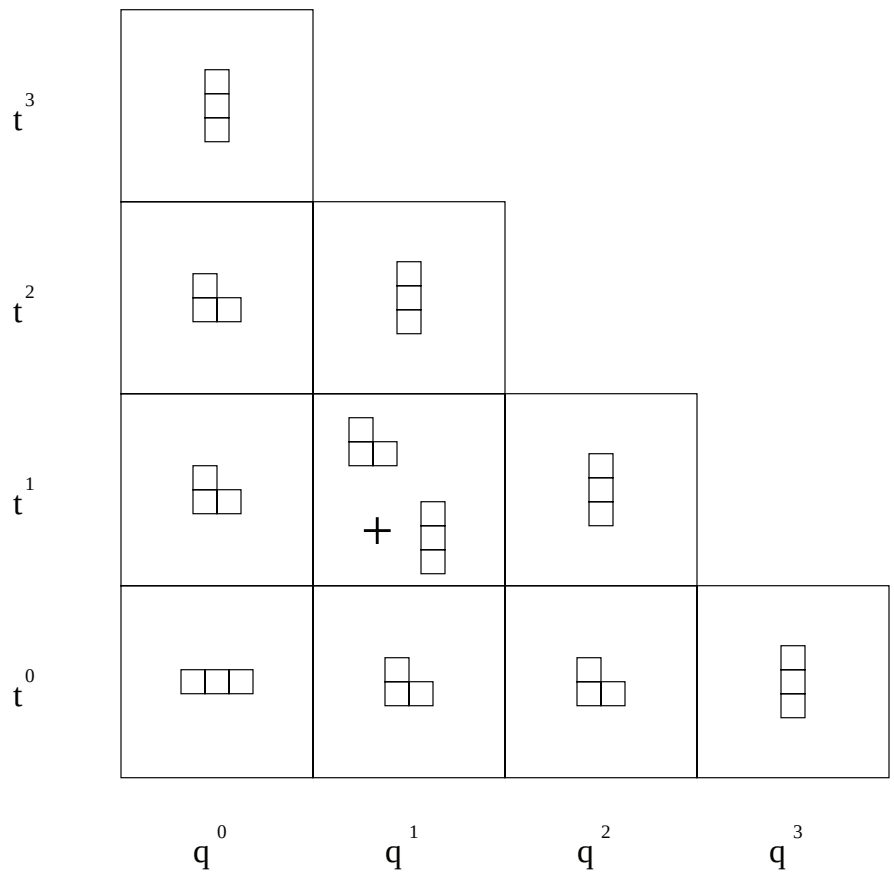


Figure 1. ∇e_3

which means that

$$\nabla|_{t=1} h_\mu[X/(1-q)] = q^{n(\mu')} h_\mu[X/(1-q)].$$

Hence,

$$k_n(q) = \nabla|_{t=1} \left(\sum_{|\mu|=n} h_\mu[X/(1-q)] f_\mu[1-q] \right) = \nabla e_n(\mathbf{x})|_{t=1},$$

as desired. □

4.3. First Remarks on Positivity

Making some calculations, we see that $\nabla e_3 = s_{(3)} + (q + t + q^2 + qt + t^2)s_{(21)} + (qt + q^3 + q^2t + qt^2 + t^3)s_{(111)}$. This is pictured in Figure 1.

Looking at the $s_{(111)} = e_3$ part gives us $C_3(q,t) = qt + q^3 + q^2t + qt^2 + t^3$, while taking the Hall inner product with $e_{(111)}$ gives us $P_3(q,t) = 1 + 2q + 2t + 2q^2 + 3qt + 2t^2 + q^3 + q^2t + qt^2 + t^3$, since $\langle s_{(3)}, e_{(111)} \rangle = \langle s_{(111)}, e_{(111)} \rangle = 1$, while $\langle s_{(21)}, e_{(111)} \rangle = 2$.

Notice that $P_3(q,t)$ and $C_3(q,t)$ are both polynomials in q and t with positive integer coefficients. This in turn follows from the coefficients of s_λ in the Schur function expansion of ∇e_3 all being polynomials with positive integer coefficients. This and further calculations suggest that $\langle \nabla e_n, s_\mu \rangle$ should always be a polynomial with positive integer coefficients. One can hope to prove this positivity in two ways. First, one can hope that ∇e_n has a combinatorial interpretation under which one can calculate $\langle \nabla e_n, s_\lambda \rangle$ by counting some set of objects (associated with the partition λ) with appropriate weights. More precisely, there should be combinatorially defined sets $\mathcal{S}_\lambda$ and functions $\mathrm{qwt}, \mathrm{twt} : \mathcal{S}_\lambda \to \mathbb{N}$ such that $\langle \nabla e_n, s_\lambda \rangle = \sum_{s \in \mathcal{S}_\lambda} q^{\mathrm{qwt}(s)} t^{\mathrm{twt}(s)}$. Secondly, one can hope ∇e_n has a representation theoretic interpretation by which ∇e_n is the bi-graded Frobenius characteristic $F_{V_n}^{(q,t)}$ for some naturally defined family of bi-graded S_n representations V_n.

Since the Macdonald polynomials $\widetilde{H}_\mu(\mathbf{x}; q, t)$ are also Schur-positive, that is, have only positive integer polynomial coefficients in their Schur function expansions, there should also be similar interpretations of the Macdonald polynomials.

At present, there are known interpretations of the Macdonald polynomials and of ∇e_n in terms of S_n-representation theory. Both ∇e_n and $\widetilde{H}_\mu$ turn out to be the Frobenius characteristics of certain finite dimensional quotients of the rings $\mathbb{C}[x_1, \ldots, x_n, y_1, \ldots, y_n]$ which we will describe in the last lecture. Although these quotient rings can be defined in an elementary way, the existing proofs of these theorems require some fairly sophisticated algebraic geometry involving the Hilbert scheme of points in the plane [**16, 17**].

As for combinatorial interpretations, those relating to ∇e_n are known and proved only for $C_n(q,t)$ and some related specializations. Some recent conjectures have, however, shed further light on this subject. These will be the main topic of the final lecture.

4.4. Exercises

(1) Prove that $\nabla e_n|_{t=0} = \widetilde{H}_{(n)}$, and that therefore $P_n(q,0) = \langle \widetilde{H}_{(n)}, e_{(1^n)} \rangle = [n]_q!$, where by definition $[n]_q! = [n]_q [n-1]_q \cdots [1]_q$ and $[k]_q = \frac{q^k - 1}{q-1}$.

LECTURE 5
Positivity and Combinatorics?

5.1. Representation Theory of $\widetilde{H}_\mu(\mathbf{x};q,t)$

Recall the Frobenius characteristic of an S_n representation V is defined as

$$F_V(\mathbf{x}) = \sum_{|\mu|=n} (\mathrm{dim} V^{S_\mu}) m_\mu(\mathbf{x}).$$

Recall also that

$$F_{V_\lambda} = s_\lambda(\mathbf{x})$$

for the irreducible representation V_λ and that Frobenius characteristic is additive on direct sums (of representations).

If V is **graded**, that is, $V = \oplus_{i\in\mathbb{N}} V_i$ where each V_i is an S_n representation, then we can define

$$F_V(\mathbf{x};q) = \sum_{i\in\mathbb{N}} F_{V_i}(\mathbf{x}) q^i.$$

Similarly, if V is **bi-graded** with $V = \oplus_{i,j\in\mathbb{N}} V_{i,j}$, we can define

$$F_V(\mathbf{x};q,t) = \sum_{i\in\mathbb{N}} F_{V_{i,j}}(\mathbf{x}) q^i t^j.$$

By construction, $\langle F_V(\mathbf{x};q,t), s_\lambda\rangle \in \mathbb{N}[q,t]$ for every λ. Therefore, one method for showing that a symmetric function $f \in \Lambda_{\mathbb{Q}(q,t)}$ has the property that $\langle f, s_\lambda\rangle \in \mathbb{N}[q,t]$ for every λ is to show that $f = F_V(\mathbf{x};q,t)$ for some bi-graded representation V.

In this section we will construct this representation V for $f = \widetilde{H}_\mu(\mathbf{x};q,t)$, which shows that $\widetilde{K}_{\lambda\mu} \in \mathbb{N}[q,t]$. In the next section we will do the same for $f = \nabla e_n$. Although we will be able to explicitly describe these representations, the proof that they have the right Frobenius characteristic involves fairly sophisticated algebraic geometry involving the Hilbert scheme of points in the plane, and would require another entire series of lectures to present. No elementary proof that these representations have the right Frobenius characteristic is known.

Given a partition μ with $|\mu| = n$, let $\{(p_1,q_1),\ldots,(p_n,q_n)\}$ be the coordinates of the boxes in its diagram. Now define

$$\Delta_\mu(x_1,x_2,\ldots,x_n,y_1,y_2,\ldots,y_n) = \det\left[x_i^{p_j} y_i^{q_j}\right].$$

For example, for $\mu = (3,2)$,

$$\Delta_{(3,2)}(\mathbf{x},\mathbf{y}) = \det \begin{bmatrix} 1 & x_1 & y_1 & x_1y_1 & y_1^2 \\ 1 & x_2 & y_2 & x_2y_2 & y_2^2 \\ 1 & x_3 & y_3 & x_3y_3 & y_3^2 \\ 1 & x_4 & y_4 & x_4y_4 & y_4^2 \\ 1 & x_5 & y_5 & x_5y_5 & y_5^2 \end{bmatrix}.$$

For $\mu = (1^n)$, only the x variables are involved, and $\Delta_{(1^n)}(\mathbf{x},\mathbf{y})$ is just the classical Vandermonde determinant

$$\Delta(\mathbf{x}) = \det \left[x_i^{j-1}\right]_{i,j=1}^n = \prod_{i>j}(x_i - x_j).$$

(Note that the convention is for powers of x to increase along the vertical axis in the partition diagram and for powers of y to increase along the horizontal axis, contrary to the usual expectation for Cartesian coordinates. Our peculiar convention has become established in the literature because the rings R_μ we will soon define were first studied in the case of Hall-Littlewood polynomials, and these are conventionally written in terms of t and the x-variables, setting q and the y-variables to 0.)

Now let S denote the ring $\mathbb{C}[x_1,\ldots,x_n,y_1,\ldots,y_n]$, bi-graded so that its (i,j)-th graded piece consists of polynomials homogeneous of degree j in the x variables and degree i in the y variables. (In the lectures and in a number of places in the literature, $\mathbb{Q}$ is used instead of $\mathbb{C}$ here. This is an irrelevant difference since the representation theory of S_n is exactly the same over the two fields. We have reverted to using $\mathbb{C}$ since that is more consistent with earlier lectures and the general study of representation theory.) Now for each partition μ with $|\mu| = n$, define an ideal J_μ of S by

$$J_\mu = \left\{ f : f(\frac{\partial}{\partial \mathbf{x}}, \frac{\partial}{\partial \mathbf{y}})\Delta_\mu(\mathbf{x},\mathbf{y}) = 0 \right\}.$$

In other words, J_μ consists of all polynomials that, when considered as partial differentiation operators, annihilate Δ_μ. Now let $R_\mu = S/J_\mu$.

The simplest example is $\mu = (1^n)$. As mentioned before, $\Delta_{(1^n)}$ is the classical Vandermonde determinant, and $J_{(1^n)} = \langle y_1,\ldots,y_n,e_1(\mathbf{x}),\ldots,e_n(\mathbf{x})\rangle$. Therefore,

$$R_{(1^n)} = \mathbb{C}[\mathbf{x}]/\langle \mathbb{C}[\mathbf{x}]_+^{S_n}\rangle,$$

or, in words, the polynomial ring in the x variables modulo the ideal generated by all homogeneous non-constant symmetric functions. This ring is known as the ring of **covariants**, and it is a classical theorem that $R_{(1^n)} \cong_{S_n} \mathbb{C}\cdot S_n \cong \mathbb{C}\uparrow_1^{S^n}$, and, furthermore, that

$$F_{R_{(1^n)}}(\mathbf{x};q,t) = (1-t)(1-t^2)\cdots(1-t^n)h_n\left[X/(1-t)\right] = \widetilde{H}_{(1^n)}(\mathbf{x};q,t).$$

Generalizing this, we have the following theorem.

Theorem 8 ([**16**]). *There holds the identity*

$$F_{R_\mu}(\mathbf{x};q,t) = \widetilde{H}_\mu(\mathbf{x};q,t).$$

Since, as computed in Lecture 3, $\widetilde{H}_\mu(\mathbf{x};1,1) = h_{(1^n)}(\mathbf{x})$, this means that $R_\mu \cong \mathbb{C}\cdot S_n$ as S_n representations. In particular, $\dim(R_\mu) = n!$. This was the "$n!$ conjecture," which turned out to be the most difficult point in the proof of Theorem 8.

5.2. Representation Theory of ∇e_n

A different quotient of the ring S gives a module whose Frobenius series is ∇e_n. Let J_n be the ideal of S generated $\mathbb{C}[\mathbf{x},\mathbf{y}]_+^{S_n}$, that is, the ideal generated by all homogeneous non-constant functions symmetric with respect to the diagonal action of S_n on the x and y variables. One set of generators for J_n is the **polarized elementary symmetric functions**, defined by

$$e_{a,b}(\mathbf{x},\mathbf{y}) = \sum_{\substack{I,J\in\{1,\ldots,n\}\\ I\cap J=\emptyset\\ \#I=a,\#J=b}} \prod_{i\in I} x_i \prod_{j\in J} y_j.$$

Now let $R_n = \mathbb{C}/J_n$, the coinvariant ring for the diagonal action. We have the following theorem.

Theorem 9 ([**17**]). *There holds the identity*

$$F_{R_n}(\mathbf{x};q,t) = \nabla e_n.$$

Corollary 3.

(1) $\nabla e_n \in \mathbb{N}[q,t]\cdot\{s_\lambda : |\lambda| = n\}$.
(2) $C_n(q,t) = \langle \nabla e_n, e_n\rangle = \sum_{i,j} \dim(R_n^\epsilon)_{i,j} t^i q^j \in \mathbb{N}[q,t]$.
(3) $P_n(q,t) = \langle \nabla e_n, e_{(1^n)}\rangle = \sum_{i,j} \dim(R_n)_{i,j} t^i q^j \in \mathbb{N}[q,t]$.

Proof. (1) holds because the Frobenius series of any (positively graded) S_n-module is in $\mathbb{N}[q,t]\cdot\{s_\lambda : |\lambda| = n\}$. Since $\langle f, e_n\rangle$ picks out the coefficient of $e_n = s_{(1^n)}$ in the expansion of f in the Schur function basis, if $f = F_V$ for some S_n representation V, $\langle f, e_n\rangle$ gives the multiplicity of the sign representation in V. Since the sign representation is 1-dimensional, (2) follows. Finally, for any S_n representation V, $\langle F_V, e_{(1^n)}\rangle = \langle F_V, h_{(1^n)}\rangle$, which is the coefficient of $m_{(1^n)}$ in the monomial expansion of F_V. By definition, this is the dimension of the subspace of V fixed by the trivial group, which is all of V, giving (3). $\square$

5.3. Combinatorics of ∇e_n

In this section we discuss a combinatorial interpretation of ∇e_n in terms of the tableaux used to represent parking functions, although we will allow tableaux of any content instead of just content (1^n). We will give two functions qwt_n, twt_n from $\bigcup_{\lambda\subseteq\delta(n)} \mathrm{SSYT}(\lambda + (1^n)/\lambda)$ to $\mathbb{N}$ such that, conjecturally,

$$\nabla e_n = \sum_T q^{\mathrm{qwt}_n(T)} t^{\mathrm{twt}_n(T)} x^T.$$

It will not be obvious at first glance that these are indeed symmetric functions, but that has been proven in [**13**]. Note that this is not exactly the desired combinatorial interpretation, as it gives an expansion of ∇e_n in terms of monomial symmetric functions rather than Schur functions, but it may be a useful first step.

Since setting $t = 1$ should give us $k_n(q)$, the desired function qwt_n should simply be $T \mapsto \binom{n}{2} - |\lambda|$, where T is a tableau of skew shape $\lambda + (1^n)/\lambda$. The appropriate function twt_n is much more subtle. We will describe this function first in terms of a combinatorial interpretation of the (q,t)-Catalan numbers $C_n(q,t)$.

Let $\lambda \subseteq \delta(n)$, and $\widehat{\lambda} = \lambda + (1^n)/\lambda$. We say that a cell $(i,j) \in \widehat{\lambda}$ **attacks** $(i',j') \in \widehat{\lambda}$ if either $i + j = i' + j'$ and $j < j'$, or $i + j = i' + j' + 1$ and $j > j'$.

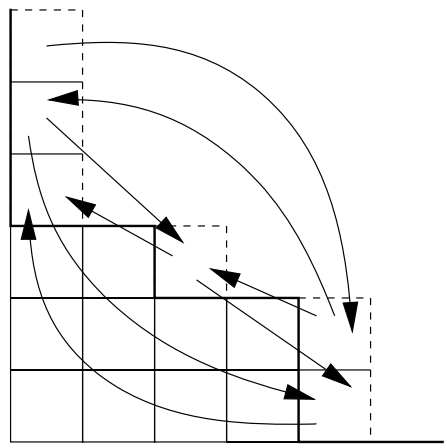


Figure 1. Attacking pairs of cells for $\lambda = (4,4,2)$ (and $n = 6$)

More pictorially, a cell c attacks c' if either c and c' are on the same diagonal with c to the left of c', or c is one diagonal above and strictly to the right of c'. Now simply let $\mathrm{twt}_n(\lambda) = \#\{(c,c')|c,c' \in \widehat{\lambda} \text{ and } c \text{ attacks } c'\}$. Figure 1 shows that $\mathrm{twt}_6((4,4,2)) = 9$.

Now we have the following theorem.

Theorem 10 ([**6, 13**])**.** *There holds the identity*

$$C_n(q,t) = \sum_{\lambda \in \delta(n)} q^{\mathrm{qwt}_n(\lambda)} t^{\mathrm{twt}_n(\lambda)}.$$

Although $C_n(q,t)$ is invariant under switching q and t, it is still an open problem to find an involution on partitions which would combinatorially explain this symmetry. More precisely, there should be a combinatorially defined involution I such that $\mathrm{qwt}_n(I(\lambda)) = \mathrm{twt}_n(\lambda)$ and $\mathrm{twt}_n(I(\lambda)) = \mathrm{qwt}_n(\lambda)$, but no such involution is known.

Now we come to the conjectured combinatorial description of ∇e_n. As stated earlier, let T be a tableau of shape $\widehat{\lambda}$, and let $\mathrm{qwt}(T) = \mathrm{qwt}_n(\lambda) = \binom{n}{2} - |\lambda|$. Now use the notion of attack defined earlier to define $\mathrm{twt}(T) = \#\{(c,c')|c,c' \in \widehat{\lambda}, T(c) > T(c'), \text{ and } c \text{ attacks } c'\}$.

Theorem 11. *For each $\lambda \subseteq \delta(n)$,*

$$D_\lambda(\mathbf{x};t) = \sum_{T \in \mathrm{SSYT}(\widehat{\lambda})} t^{\mathrm{twt}(T)} x^T$$

is a symmetric function, and $D_\lambda(\mathbf{x};t) \in \mathbb{N}[t] \cdot \{s_\lambda\}$.

In fact, $D_\lambda(\mathbf{x};t)$ is shown in [**13**] to be an example of an LLT polynomial, as defined by Lascoux, Leclerc and Thibon in [**19**].

Conjecture 1.

$$\nabla e_n = \sum_{\lambda \subseteq \delta(n)} q^{\mathrm{qwt}_n(\lambda)} D_\lambda(\mathbf{x};t) = \sum_{\substack{\lambda \subseteq \delta(n) \\ T \in \mathrm{SSYT}(\widehat{\lambda})}} q^{\mathrm{qwt}(T)} t^{\mathrm{twt}(T)} x^T.$$

This conjecture, if true, would have the following corollary; recall that a tableau of skew shape $\widehat{\lambda}$ and content (1^n) corresponds directly to a parking function.

Corollary 4 (to Conjecture 1)**.**

$$P_n(q,t) = \langle \nabla e_n, h_{(1^n)} \rangle = \sum_{\substack{\lambda \subseteq \delta(n) \\ T \in \mathrm{SSYT}(\widehat{\lambda},(1^n))}} q^{\mathrm{qwt}(T)} t^{\mathrm{twt}(T)}.$$

It is also mysterious why this should be symmetric under switching q and t, and what connection these combinatorics may have with the ring R_n described above.

It can at least be shown that insofar as $C_n(q,t)$ is concerned, Conjecture 1 agrees with Theorem 10. First of all, in keeping with how ω usually acts on objects indexed by tableaux, it can be shown that

$$\omega D_\lambda(\mathbf{x};t) = \sum_{T \in \mathrm{SSYT}_-(\widehat{\lambda})} t^{\mathrm{twt}(T)} x^T,$$

where $\mathrm{SSYT}_-(\widehat{\lambda})$ denotes the set of **imaginary** tableaux T of shape $\widehat{\lambda}$, whose "imaginary" entries increase weakly along columns and strictly along rows (the requirement on rows is irrelevant in the case of the shapes $\widehat{\lambda}$ occurring in the above formula). For imaginary tableaux, $\mathrm{twt}(T)$ is redefined to allow a contribution from a pair of cells (c,c') if c attacks c and $T(c) \geq T(c')$ (instead of requiring $T(c) > T(c')$).

For each $\lambda \subseteq \delta(n)$, there is a unique imaginary tableau T of shape $\widehat{\lambda}$ with all entries being $\overline{1}$, and for this imaginary tableau, $\mathrm{twt}(T) = \mathrm{twt}_n(\lambda)$. Therefore, $\langle \sum_{\lambda \in \delta(n)} q^{\mathrm{qwt}_n(\lambda)} \omega D_\lambda(\mathbf{x};t), h_n \rangle = C_n(q,t)$. Since $\langle \sum_{\lambda \subseteq \delta(n)} q^{\mathrm{qwt}_n(\lambda)} D_\lambda(\mathbf{x};t), e_n \rangle = \langle \sum_{\lambda \subseteq \delta(n)} q^{\mathrm{qwt}_n(\lambda)} \omega D_\lambda(\mathbf{x};t), h_n \rangle$, the theorem for $C_n(q,t)$ agrees with the conjecture.

5.4. Combinatorics of $\widetilde{H}_\mu(\mathbf{x};q,t)$

This topic was addressed, not in these lectures, but in a satellite lecture by Jim Haglund. We will comment briefly on the latest developments. Haglund conjectured, and discussed in his lecture, a combinatorial formula analogous to Conjecture 1 for the monomial expansion of $\widehat{H}_\mu(\mathbf{x};q,t)$. Like Conjecture 1, Haglund's formula can be expressed as a q-weighted sum of LLT polynomials in the parameter t, which shows in particular that it is in fact a symmetric function. (This also shows, subject to a general Schur-positivity conjecture for LLT polynomials, that Haglund's formula is Schur-positive. The special case of the LLT positivity conjecture required for Schur-positivity of the formula in Conjecture 1 is known to hold.) Between the the PCMI meeting and the preparation of the final version of these notes, Haglund's conjecture has been proven by Haglund, Haiman and Loehr, who verify directly that Haglund's formula satisfies the defining axioms for Macdonald polynomials in Theorem-Definition 2. For details, see [**12**].

5.5. Exercises

Show that Conjecture 1 gives the correct predictions for the following.

(1) $\nabla e_n \mid_{t=1} = k_n(q)$
(2) $\nabla e_n \mid_{q=0} = (1-t)(1-t^2)\cdots(1-t^n) h_n\left[X/(1-t)\right] = \widetilde{H}_{(1^n)}(\mathbf{x};q,t)$
(3) $\nabla e_n \mid_{t=0}$ (This one is trickier.)

Proving that the conjecture gives the correct prediction for $\nabla e_n \mid_{q=1}$ is an open problem. Using the first exercise, this is presumably a special case for showing combinatorially that the conjecture gives a function symmetric under switching q and t.

BIBLIOGRAPHY

1. F. Bergeron, A. M. Garsia, M. Haiman, and G. Tesler, *Identities and positivity conjectures for some remarkable operators in the theory of symmetric functions*, Methods Appl. Anal. **6** (1999), no. 3, 363–420.
2. R. Bezrukavnikov and D. Kaledin, *McKay equivalence for symplectic resolutions of singularities*, Electronic preprint, arXiv:math.AG/0401002, 2004.
3. L. Carlitz and J. Riordan, *Two element lattice permutation numbers and their q-generalization*, Duke Math. J. **31** (1964), 371–388.
4. Ivan Cherednik, *Diagonal coinvariants and double affine Hecke algebras*, Int. Math. Res. Not. (2004), no. 16, 769–791, math.QA/0305245.
5. Pavel Etingof and Victor Ginzburg, *Symplectic reflection algebras, Calogero-Moser space, and deformed Harish-Chandra homomorphism*, Invent. Math. **147** (2002), no. 2, 243–348, arXiv:math.AG/0011114.
6. A. M. Garsia and J. Haglund, *A proof of the q,t-Catalan positivity conjecture*, Discrete Math. **256** (2002), no. 3, 677–717, LaCIM 2000 Conference on Combinatorics, Computer Science and Applications (Montreal, QC).
7. A. M. Garsia and M. Haiman, *A remarkable q,t-Catalan sequence and q-Lagrange inversion*, J. Algebraic Combin. **5** (1996), no. 3, 191–244.
8. Adriano M. Garsia, *A q-analogue of the Lagrange inversion formula*, Houston J. Math. **7** (1981), no. 2, 205–237.
9. Ira Gessel, *A noncommutative generalization and q-analog of the Lagrange inversion formula*, Trans. Amer. Math. Soc. **257** (1980), no. 2, 455–482.
10. Victor Ginzburg, *Principal nilpotent pairs in a semisimple Lie algebra. I*, Invent. Math. **140** (2000), no. 3, 511–561.
11. Iain Gordon, *On the quotient ring by diagonal invariants*, Invent. Math. **153** (2003), no. 3, 503–518, arXiv:math.RT/0208126.
12. J. Haglund, M. Haiman, and N. Loehr, *A combinatorial formula for Macdonald polynomials*, J. Amer. Math. Soc. **18** (2005), no. 3, 735–761 (electronic), arXiv:math.CO/0409538.
13. J. Haglund, M. Haiman, N. Loehr, J. B. Remmel, and A. Ulyanov, *A combinatorial formula for the character of the diagonal coinvariants*, Duke Math. J. **126** (2005), no. 2, 195–232, arXiv:math.CO/0310424.
14. Mark Haiman, *t,q-Catalan numbers and the Hilbert scheme*, Discrete Math. **193** (1998), no. 1-3, 201–224, Selected papers in honor of Adriano Garsia (Taormina, 1994).

15. ———, *Macdonald polynomials and geometry*, New perspectives in geometric combinatorics (Billera, Björner, Greene, Simion, and Stanley, eds.), MSRI Publications, vol. 38, Cambridge University Press, 1999, pp. 207–254.
16. ———, *Hilbert schemes, polygraphs and the Macdonald positivity conjecture*, J. Amer. Math. Soc. **14** (2001), no. 4, 941–1006, arXiv:math.AG/0010246.
17. ———, *Vanishing theorems and character formulas for the Hilbert scheme of points in the plane (abbreviated version)*, Physics and combinatorics, 2000 (Nagoya), World Sci. Publishing, River Edge, NJ, 2001, pp. 1–21.
18. ———, *Combinatorics, symmetric functions, and Hilbert schemes*, Current developments in mathematics, 2002, Int. Press, Somerville, MA, 2003, pp. 39–111.
19. Alain Lascoux, Bernard Leclerc, and Jean-Yves Thibon, *Ribbon tableaux, Hall-Littlewood functions, quantum affine algebras, and unipotent varieties*, J. Math. Phys. **38** (1997), no. 2, 1041–1068, arXiv:q-alg/9512031.
20. I. G. Macdonald, *A new class of symmetric functions*, Actes du 20e Séminaire Lotharingien, vol. 372/S-20, Publications I.R.M.A., Strasbourg, 1988, pp. 131–171.
21. ———, *Symmetric functions and Hall polynomials*, second ed., The Clarendon Press, Oxford University Press, New York, 1995, With contributions by A. Zelevinsky, Oxford Science Publications.
22. Richard P. Stanley, *Enumerative combinatorics. Vol. 2*, Cambridge University Press, Cambridge, 1999, With a foreword by Gian-Carlo Rota and appendix 1 by Sergey Fomin.

Chromatic Numbers, Morphism Complexes, and Stiefel-Whitney Characteristic Classes

Dmitry N. Kozlov

IAS/Park City Mathematics Series
Volume 14, 2004

Chromatic Numbers, Morphism Complexes, and Stiefel-Whitney Characteristic Classes

Dmitry N. Kozlov

Preamble

Combinatorics, in particular graph theory, has a rich history of being a domain of successful applications of tools from other areas of mathematics, including topological methods. Here, we survey the study of the `Hom`-complexes, and the ways these can be used to obtain lower bounds for the chromatic numbers of graphs, presented in a recent series of papers [**BK03a, BK03b, BK04, CK04a, CK04b, Ko04, Ko05b**].

The structural theory is developed and put in the historical context, culminating in the proof of the Lovász Conjecture, which can be stated as follows:

> *For a graph G, such that the complex* $\mathtt{Hom}\,(C_{2r+1}, G)$ *is k-connected for some $r, k \in \mathbb{Z}$, $r \geq 1$, $k \geq -1$, we have $\chi(G) \geq k+4$.*

Beyond the, more customary in this area, cohomology groups, the algebro-topological concepts involved are spectral sequences and Stiefel-Whitney characteristic classes. Complete proofs are included for all the new results appearing in this survey for the first time.

[1]Institute of Theoretical Computer Science / Department of Mathematics, Eidgenössische Technische Hochschule - Zürich, CH-8006 Zürich, Switzerland.
E-mail address: `dkozlov@inf.ethz.ch`.
The author would like to thank the Swiss National Science Foundation and Mathematical Science Research Institute, Berkeley for the generous support.

©2007 American Mathematical Society

LECTURE 1
Introduction

1.1. The Chromatic Number of a Graph

1.1.1. The Definition and Applications

Unless stated otherwise, all graphs are undirected, loops are allowed, whereas multiple edges are not. We shall occasionally stress these conventions, to avoid the possibility of misunderstanding.

For a graph G, $V(G)$ denotes the set of its vertices, and $E(G)$ denotes the set of its edges. If convenient, we think of $E(G)$ as a $\mathbb{Z}_2$-invariant subset of $V(G) \times V(G)$, where $\mathbb{Z}_2$ acts on $V(G) \times V(G)$ by switching the coordinates: $(x, y) \mapsto (y, x)$. Under this convention, a looped vertex x is encoded by the diagonal element (x, x), while the edge from x to y (for $x \neq y$) is encoded by the pair $(x, y), (y, x) \in V(G) \times V(G)$. For example, the edge set of the graph with 2 vertices connected by an edge, were the first vertex is looped, and the second one is not, is encoded by the set $\{(1, 1), (1, 2), (2, 1)\}$.

Definition 1.1.1. *Let G be a graph. A* **vertex-coloring** *of G is a set map $c : V(G) \to S$ such that $(x, y) \in E(G)$ implies $c(x) \neq c(y)$.*

Clearly, a vertex coloring exists if and only if G has no loops.

Definition 1.1.2. *The* **chromatic number** *of G, $\chi(G)$, is the minimal cardinality of a finite set S, such that there exists a vertex-coloring $c : V(G) \to S$.*

An example of a graph with chromatic number 4 is shown on Figure 1.1.1. If no such finite set S exists, for example, if G has loops, we use the convention $\chi(G) = \infty$.

The literature devoted to the applications of computing the chromatic number of a graph is very extensive. Two of the basic applications are the *frequency assignment problem* and the *task scheduling problem*.

The first one concerns a collection of transmitters, with certain pairs of transmitters required to have different frequencies (e.g., because they are too close). Clearly, the minimal number of frequencies required for such an assignment is precisely the chromatic number of the graph, whose vertices correspond to the transmitters, with two vertices connected by an edge if and only if the corresponding transmitters are requested to have different frequencies.

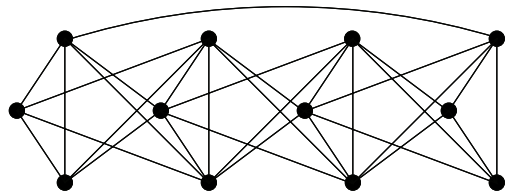


Figure 1.1.1. A graph with chromatic number 4, which does not contain K_4 as an induced subgraph.

The second problem concerns a collection of tasks which need to be performed. Each task has to be performed exactly once, and the tasks are to be performed in regularly allocated slots (e.g., hours). The only constraint is that certain tasks cannot be performed simultaneously. Again, the minimal number of slots required for the task scheduling is equal to the chromatic number of the graph, whose vertices correspond to tasks, with two vertices connected by an edge if and only if the corresponding tasks cannot be performed simultaneously.

1.1.2. The Hadwiger Conjecture

The question of computing $\chi(G)$ has a long history. In 1852, F. Guthrie, [**Gut80**], asked whether it is true that any planar map of connected countries can be colored with 4 colors, so that every pair of countries, which share a (non-point) boundary segment, receive different colors. The first time this question appeared in print was in a paper by Cayley, [**Cay78**], and it became known as the Four-Color Problem, one of the most famous questions in graph theory, as well as a popular brain-teaser. There is very extensive literature on the subject, see e.g., [**Har69, KS77, MSTY, Ore67, Th98**].

The apparently first proof, offered by A. Kempe in 1880, [**Kem79**], turned out to be false, as did many later ones. The flaw was noticed in 1890 by P. Heawood, [**Hea90**], who also proved the weaker Five-Color Theorem. After several important contributions, most notably by G. Birkhoff, and H. Heesch, [**Hee69**], the latter reduced the Four-Color Theorem to the analysis of the large, but finite set of "unavoidable" configurations, the original conjecture has been proved 1976 by Appel & Haken, using computer computations, see [**AH76**] for the original announcement and [**AH89**] for last reprint. A new, shorter and more structural proof (though still relying on computers) has been obtained in 1997, see [**RSST**]. The usual way to formulate this theorem is to dualize the map to obtain a planar graph, coloring vertices instead of the countries.

Theorem 1.1.3 (**The Four-Color Theorem).** (Appel & Haken, [**AH89**]; revised proof by Robertson, Sanders, Seymour & Thomas, [**RSST**]).
Every planar graph is four colorable.

In 1943, Hadwiger, [**Had43**], stated a conjecture closely related to the Four-Color Theorem. Recall that a graph H is called a *minor* of another graph G, if H can be obtained from a subgraph of G by a sequence of edge-contractions. Let K_n denote an unlooped complete graph on n vertices, that is, $V(K_n) = [n]$, $E(K_n) = \{(x, y) \mid x, y \in [n], x \neq y\}$.

Conjecture 1.1.4 (Hadwiger Conjecture).
For every positive integer t, if a graph has no K_{t+1} minor, then it has a t-coloring, in other words, every graph G has $K_{\chi(G)}$ as its minor.

The Hadwiger conjecture is proved for $\chi(G) \leq 5$. Indeed, it is trivial for $\chi(G) = 1$, as K_1 is a minor of any graph. For $\chi(G) = 2$ it just says that K_2 is a minor of an arbitrary graph containing an edge. If $\chi(G) = 3$, then G contains an odd cycle, in particular it has K_3 as a minor. The case $\chi(G) = 4$ is reasonably easy, and was shown by Hadwiger, [**Had43**], and Dirac, [**Dir52**]. Finally, it was shown in 1937 by Wagner, [**Wag37**], that the case $\chi(G) = 5$ of the Hadwiger Conjecture is equivalent to the Four-Color Theorem.

1.1.3. The Complexity of Computing the Chromatic Number

The problem of computing the chromatic number of a graph is NP-complete, implying that the worst-case performance of any algorithm is, most likely, exponential in the number of vertices. Stronger, already the, seemingly much more special, problem of deciding whether a given planar graph is 3-colorable is NP-complete, see e.g., [**GJ79**]. Recently, it has been shown that even coloring a 3-colorable graph with 4 colors is NP-complete, see [**KLS93**].

We note that deciding whether $\chi(G) = 2$ (i.e., whether the graph is bipartite) is computationally much easier. The plain depth-first search yields $O(|V(G)|+|E(G)|)$ performance time. Another good news is that one can 4-color a planar graph in polynomial time: quartic time was obtained in [**AH89**], and later improved to quadratic time, see [**RSST**].

The situation is not getting much better if we switch to considering approximations. For example, it was shown by Garey & Johnson, [**GJ76**], that if a polynomial time approximate algorithm for graph coloring exists (in the precise formulation, meaning that the output of the algorithm does not differ by more than a constant factor, which is smaller than 2, from the actual value of the chromatic number), then there exists a polynomial time algorithm for graph coloring, which, of course, is not very likely.

Much of the same can be said about computing lower bounds. Even the most trivial lower bound for the chromatic number, given by the clique number, is not good, since computing the clique number of a graph is also an NP-complete problem. For fixed clique size, the lower bound based on clique number is polynomially computable, but is not very interesting.

The original Lovász bound, by virtue of being based on computing the connectivity of a simplicial complex, also has a very high computational complexity, since determining the triviality of the homotopy groups is an extremely hard problem, even in low dimensions.

It is then a positive and welcome surprise, that our bounds, based on the Stiefel-Whitney characteristic classes are both nontrivial and polynomially computable; here we fix the test graph and the tested dimension and consider the computational complexity with respect to the number of vertices of the graph which is being tested. The crucial difference is that, as opposed to homotopy, the cohomology groups (and the functorial invariants contained therein) may be computed by means of simple linear algebra.

1.2. The Category of Graphs

1.2.1. Graph Homomorphisms and the Chromatic Number

The following notion is the gist in recasting the various coloring questions in the functorial language.

Definition 1.2.1. *For two graphs T and G, a* **graph homomorphism** *from T to G is a map $\varphi : V(T) \to V(G)$, such that if $x, y \in V(T)$ are connected by an edge, then $\varphi(x)$ and $\varphi(y)$ are also connected by an edge.*

In other words, the map of the vertex sets $\varphi : V(T) \to V(G)$ induces the product map $\varphi \times \varphi : V(T) \times V(T) \to V(G) \times V(G)$, and the condition for the set map φ to be a graph homomorphism translates into

$$(\varphi \times \varphi)(E(T)) \subseteq E(G).$$

Expressed verbally: *edges map to edges.*

The study of graph homomorphisms is a classical and well-developed subject within combinatorics. The interested reader may want to consult the textbooks [**GR01**] and [**HN04**].

Clearly, for any two positive integers m and n, a graph homomorphism $\varphi : K_m \to K_n$ exists if and only if $m \leq n$. More generally, we can now restate Definition 1.1.2 in the language of graph homomorphisms.

Definition 1.2.2. *The* **chromatic number** *of G, $\chi(G)$, is the minimal positive integer n, such that there exists a graph homomorphism $\varphi : G \to K_n$.*

In this sense, the problem of vertex-colorings and computing chromatic numbers corresponds to choosing a particular family of graphs, namely unlooped complete graphs, fixing a valuation on this family, here we are mapping K_n to n, and then searching for a graph homomorphism from a given graph to the chosen family, which would minimize the fixed valuation. Using the intuition from statistical mechanics we call such a family of graphs *state graphs*.

A natural question arises: are there any other choices of families of state graphs and valuations which correspond to other natural and well-studied classes of graph problems. The answer is yes, and we shall describe two examples in the following subsections.

1.2.2. The Fractional Chromatic Number

First, we define an important family of graphs.

Definition 1.2.3. *Let n, k be positive integers, $n \geq 2k$. The* **Kneser graph** *$K_{n,k}$ is defined to be the graph whose set of vertices is the set of all k-subsets of $[n]$, and the set of edges is the set of all pairs of disjoint k-subsets.*

Examples:

- $K_{2k,k}$ is a matching on $\binom{2k}{k}$ vertices;
- $K_{n,1}$ is the unlooped complete graph K_n;
- $K_{5,2}$ is the Petersen graph.

We can now define the fractional chromatic number by means of graph homomorphisms.

Definition 1.2.4. *Let G be a graph. The* **fractional chromatic number** *of G, $\chi_f(G)$, is defined by*

$$\chi_f(G) = \inf_{(n,k)} \frac{n}{k} ,$$

where the infimum is taken over all pairs (n, k) such that there exists a graph homomorphism from G to $K_{n,k}$.

Here, the state graphs are the Kneser graphs, $\{K_{n,k}\}_{n\geq 2k}$, and the chosen valuation on this family is $K_{n,k} \mapsto n/k$.

1.2.3. The Circular Chromatic Number

Again, we start by defining the appropriate family of graphs.

Definition 1.2.5. *Let r be a real number, $r \geq 2$. R_r is defined to be the graph whose set of vertices is the set of unit vectors in the plane pointing from the origin, and two vertices x and y are connected by an edge if and only if $2\pi/r \leq \alpha$, where α is the sharper of the two angles between x and y (or π if these two angles are equal).*

Note that both the number of vertices and valencies of the vertices (if $r > 2$) are infinite.

Definition 1.2.6. *Let G be a graph. The* **circular chromatic number** *of G is $\chi_c(G) = \inf r$, where the infimum is taken over all positive reals r, such that there exists a graph homomorphism from G to R_r.*

In other words, the family of the state graphs is $\{R_r\}_{r\geq 2}$, and the chosen valuation is $R_r \mapsto r$.

It is possible to define $\chi_c(G)$ by using only finite state graphs.

Definition 1.2.7. *Let n, k be positive integers, $n \geq 2k$. $R_{n,k}$ is defined to be the graph whose set of vertices is $[n]$, and two vertices $x, y \in [n]$ are connected by an edge if and only if*

$$k \leq |x - y| \leq n - k.$$

Examples:

- $R_{2k,k}$ is a complete matching on $2k$ vertices;
- $R_{2k+1,k}$ is a cycle with $2k + 1$ vertices;
- $R_{n,1} = K_{n,1} = K_n$;
- $R_{n,2}$ is the unlooped complement of a cycle with n vertices.

The equivalent definition of $\chi_c(G)$ in terms of finite state graphs is also functorial.

Proposition 1.2.8. *Let G be a graph. We have the equality*

$$\chi_c(G) = \inf_{(n,k)} \frac{n}{k} ,$$

where the infimum is taken over all pairs (n, k) such that there exists a graph homomorphism from G to $R_{n,k}$.

Here, the state graphs are $\{R_{n,k}\}_{n\geq 2k}$, and the chosen valuation on this family is again $R_{n,k} \mapsto n/k$.

We remark, that for any graph G we have

$$\chi(G) - 1 < \chi_c(G) \leq \chi(G).$$

For more information on the circular chromatic number of a graph, see [**Vi88, Zhu01**].

1.2.4. The Category Graphs

As we have seen so far, graph homomorphisms are invaluable in formulating various coloring problems. The usual framework for studying a set of mathematical objects and the set of maps between them is that of *a category*. Before we get that, we need to check a few properties.
(1) Let T, G, H be three graphs, and let $\varphi : T \to G$ and $\psi : G \to H$ be two graph homomorphisms. The *composition* of the set maps $\psi \circ \varphi$ is again a graph homomorphism from T to H, as $(x, y) \in E(T)$ implies $(\varphi(x), \varphi(y)) \in E(G)$, which further implies $(\psi(\varphi(x)), \psi(\varphi(y))) \in E(H)$.
(2) The composition of set maps (and hence of graph homomorphisms) is *associative*.
(3) For any graph G, the *identity* map $\mathrm{id} : V(G) \to V(G)$ is a graph homomorphism.

Now we are ready to put it all together into one structure.

Definition 1.2.9. **Graphs** *is the category defined as follows:*

- *the objects of* **Graphs** *are all graphs;*
- *the morphisms* $\mathcal{M}(G, H)$ *for two objects* $G, H \in \mathcal{O}(\mathbf{Graphs})$ *are all graph homomorphisms from* G *to* H*.*

For a graph G, let G^o be the *looping* of G, i.e., $V(G^o) = V(G)$, $E(G^o) = E(G) \cup \{(x, x) \,|\, x \in V(G)\}$. Then, K_1^o is a graph consisting of one vertex and one loop, it is the *terminal object* of **Graphs**. The empty graph is the *initial object* of **Graphs**.

As a useful variation we also consider the category $\mathbf{Graphs_p}$(where $\mathbf{p}$ stand for "proper"), whose objects are all graphs, and whose morphisms are all proper graph homomorphisms. We call a graph homomorphism $\varphi : T \to G$ *proper* if $|\varphi^{-1}(g)|$ is finite for all $g \in V(G)$.

Proposition 1.2.10. *The direct product of graphs (see Definition 4.1.6) is a categorical product, while the disjoint union of graphs is a categorical coproduct in* **Graphs**.

Even more generally, we have the following property.

Proposition 1.2.11. **Graphs** *has all finite limits and colimits.*

A surprising result of Welzl, [**Wel84**], shows that this category is in a certain sense dense.

Theorem 1.2.12 (Welzl Theorem)**.**
Let T *and* G *be two arbitrary finite graphs, such that* $\chi(T) \geq 3$*, and there exists a graph homomorphism from* T *to* G*, but there is no graph homomorphism from* G *to* T*. Then, there exists a graph* H*, such that there exist graph homomorphisms from* T *to* H*, and from* H *to* G*, but there are no graph homomorphisms from* H *to* T*, or from* G *to* H*.*

In the setting of this category, we can think of the following generalization of various coloring problems: given a category $\mathcal{C}$, and a subcategory $\widetilde{\mathcal{C}}$, determine the set of morphisms from a given object A to the objects in $\widetilde{\mathcal{C}}$. In other words, we need to study obstructions to the existence of morphisms between certain objects.

1.2.5. Test Objects

Due to the lack of structure, it is rather forbidding to study obstructions in the category **Graphs** directly. Instead, we consider a functor $\mathcal{F} : \mathbf{Graphs} \to \mathcal{T}$, where $\mathcal{T}$ is some category with a well-developed obstruction theory.

For two graphs $A, B \in \mathcal{O}(\mathbf{Graphs})$, if there exists a graph homomorphism $\varphi : A \to B$, then, since $\mathcal{F}$ is a functor, we have an induced morphism $\mathcal{F}(\varphi) : \mathcal{F}(A) \to \mathcal{F}(B)$. If, on the other hand, by some general obstruction arguments in $\mathcal{T}$, there can be no morphism $\mathcal{F}(A) \to \mathcal{F}(B)$, then we have gotten a contradiction, hence $\mathcal{M}(A, B) = \emptyset$.

The question then becomes: how does one find "good" functors $\mathcal{F}$, i.e., functors which yield nontrivial obstructions to the existence of morphisms in **Graphs**.

The centerpiece of this survey is the following choice of $\mathcal{F}$: choose a test graph T, and map every $G \in \mathbf{Graphs}$ to a topological space which is derived from the set of graph homomorphisms $\varphi : T \to G$. The idea of topologizing the set of graph homomorphisms between two given graphs is due to L. Lovász and will be presented in detail in Section 2.1.

Let us recall the following standard construction in category theory, cf. [**McL98**, Section II.6]. For a category $\mathcal{C}$ and an object $a \in \mathcal{O}(\mathcal{C})$, the *category of objects under* a, denoted $a \downarrow \mathcal{C}$, is defined as follows:

- the objects of $a \downarrow \mathcal{C}$ are all pairs (m, b), where m is a morphism from a to b;
- for $b_1, b_2 \in \mathcal{O}(\mathcal{C})$, and $m_1 \in \mathcal{M}(a, b_1)$, $m_2 \in \mathcal{M}(a, b_2)$, the morphisms in $a \downarrow \mathcal{C}$ from (m_1, b_1) to (m_2, b_2) are all morphisms $m : b_1 \to b_2$, such that $m \circ m_1 = m_2$ (in other words, all ways to complement m_1, m_2 to an appropriate commutative triangle).

The interesting and crucial detail here is the additional topological structure which we have on top of the more usual comma category construction.

LECTURE 2
The Functor $\mathtt{Hom}\,(-,-)$

2.1. Complexes of Graph Homomorphisms

2.1.1. Complex of Complete Bipartite Subgraphs and the Neighborhood Complex

First, we state the well-known definition in the generality which we need here.

Definition 2.1.1. *Let $A, B \subseteq V(G)$, $A, B \neq \emptyset$. We call (A, B) a* **complete bipartite subgraph** *of G, if for any $x \in A$, $y \in B$, we have $(x, y) \in E(G)$, i.e., $A \times B \subseteq E(G)$.*

In particular, note that all vertices in $A \cap B$ are required to have loops, and that the edges between the vertices of A (or of B) are allowed. An example is shown on Figure 2.1.1.

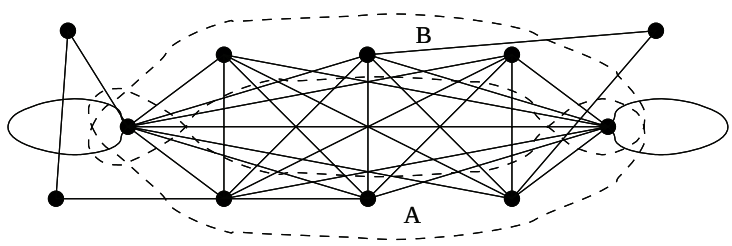


Figure 2.1.1. A complete bipartite subgraph.

Let G be a (possibly infinite) graph. Let $\Delta^{V(G)}$ be a simplex whose set of vertices is $V(G)$, in particular, the simplices of $\Delta^{V(G)}$ can be identified with the finite subsets of $V(G)$. We stress here, that we take as an infinite simplex, the colimit of the standard inclusion sequence of finite simplices:

$$\Delta^0 \hookrightarrow \Delta^1 \hookrightarrow \Delta^2 \hookrightarrow \dots.$$

Under this convention, the points of $\Delta^{V(G)}$ are all convex combinations of the points $V(G)$, where only finitely many points have nonzero coefficients.

A direct product of regular CW complexes is again a regular CW complex. Even stronger, $\Delta^{V(G)} \times \Delta^{V(G)}$ can be thought of as a polyhedral complex, whose cells are direct products of two simplices.

Definition 2.1.2. $\mathrm{Bip}\,(G)$ *is the subcomplex of* $\Delta^{V(G)} \times \Delta^{V(G)}$ *defined by the following condition:* $\sigma \times \tau \in \mathrm{Bip}\,(G)$ *if and only if* (σ, τ) *is a complete bipartite subgraph of* G.

Note that if (A, B) is a complete bipartite subgraph of G, and $\widetilde{A} \subseteq A$, $\widetilde{B} \subseteq B$, $\widetilde{A}, \widetilde{B} \neq \emptyset$, then $(\widetilde{A}, \widetilde{B})$ is also a complete bipartite subgraph of G. This verifies that $\mathrm{Bip}\,(G)$ is actually a subcomplex.

$\mathrm{Bip}\,(G)$ is a CW complex, whose closed cells are isomorphic to direct products of simplices (in the particular case here, they are in fact products of two simplices). We call complexes satisfying that property *prodsimplicial*.

In 1978 Lovász proposed the following construction.

Definition 2.1.3. *Let* G *be a graph. The* **neighborhood complex** *of* G *is the simplicial complex* $\mathcal{N}(G)$ *defined as follows: its vertices are all non-isolated vertices of* G*, and its simplices are all the subsets of* $V(G)$ *which have a common neighbor.*

Let $\mathcal{N}(v)$ denote the set of neighbors of v, i.e.,

$$\mathcal{N}(v) = \{x \in V(G) \,|\, (v, x) \in E(G)\}.$$

Then, the maximal simplices of $\mathcal{N}(G)$ are precisely $\mathcal{N}(v)$, for $v \in V(G)$. The complexes $\mathrm{Bip}\,(G)$ and $\mathcal{N}(G)$ are closely related.

Proposition 2.1.4. *Let* G *be an arbitrary graph.*

(a) ([**BK03b**, Proposition 4.2]).

$\mathrm{Bip}\,(G)$ *is homotopy equivalent to* $\mathcal{N}(G)$.

(b) ([**Ko05b**, Theorem 7.2]).

Even stronger, $\mathrm{Bip}\,(G)$ *and* $\mathcal{N}(G)$ *have the same simple homotopy type.*

$\mathrm{Bip}\,(G)$ is our first example of the $\mathtt{Hom}\,(-,-)$-construction, namely, $\mathrm{Bip}\,(G)$ is isomorphic, as a polyhedral complex, to $\mathtt{Hom}\,(K_2, G)$.

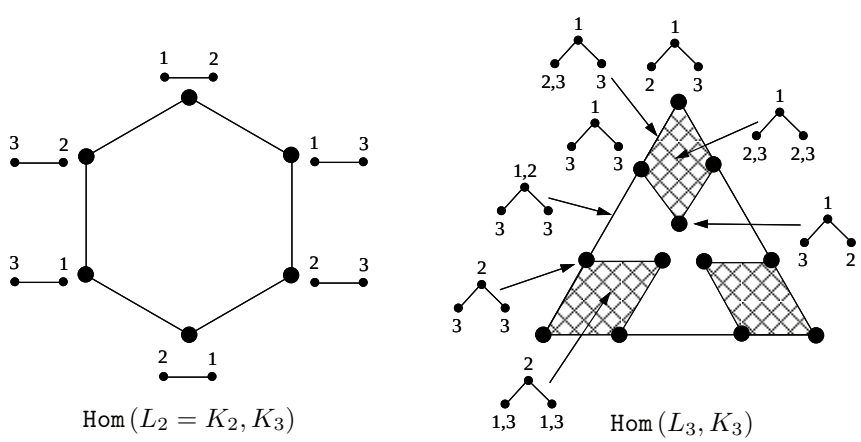


Figure 2.1.2. 3-coloring complexes of an edge and of a 3-string.

2.1.2. $\mathtt{Hom}$-Construction for Graphs

We shall now define $\mathtt{Hom}\,(T,G)$ for an arbitrary pair of graphs T and G. As a model, we take the definition of Bip (G). Let again $\Delta^{V(G)}$ be a simplex whose set of vertices is $V(G)$. Let $C(T,G)$ denote the *weak* direct product $\prod_{x\in V(T)}\Delta^{V(G)}$, i.e., the copies of $\Delta^{V(G)}$ are indexed by vertices of T.

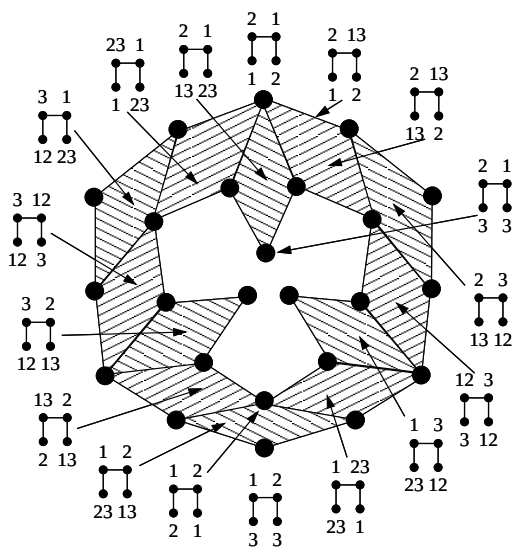
By the weak direct product we mean the following construction: a cell in $C(T,G)$ is a direct product of cells $\prod_{x\in V(T)}\sigma_x$, with the extra condition that $\dim\sigma_x \geq 1$ for only finitely many x. The dimension of this cell is $\sum_{x\in V(T)}\dim\sigma_x$, in particular, it is finite.

Definition 2.1.5. $\mathtt{Hom}\,(T,G)$ *is the subcomplex of* $C(T,G)$ *defined by the following condition:* $\sigma = \prod_{x\in V(T)}\sigma_x \in \mathtt{Hom}\,(T,G)$ *if and only if for any* $x,y \in V(T)$*, if* $(x,y)\in E(T)$*, then* (σ_x,σ_y) *is a complete bipartite subgraph of* G*.*

Let us make a number of simple, but fundamental, observations about the complexes $\mathtt{Hom}\,(T,G)$.
(1) The topology of $\mathtt{Hom}\,(T,G)$ is inherited from the product topology of $C(T,G)$. By this inheritance, the cells of $\mathtt{Hom}\,(T,G)$ are products of simplices.
(2) $\mathtt{Hom}\,(T,G)$ is a polyhedral complex whose cells are indexed by all functions $\eta : V(T) \to 2^{V(G)}\setminus\{\emptyset\}$, such that if $(x,y)\in E(T)$, then $\eta(x)\times\eta(y)\subseteq E(G)$.

The closure of a cell η consists of all cells indexed by $\tilde{\eta} : V(T)\to 2^{V(G)}\setminus\{\emptyset\}$, which satisfy $\tilde{\eta}(v)\subseteq\eta(v)$, for all $v\in V(T)$.

Throughout this survey, we shall make extensive use of the η-notation.

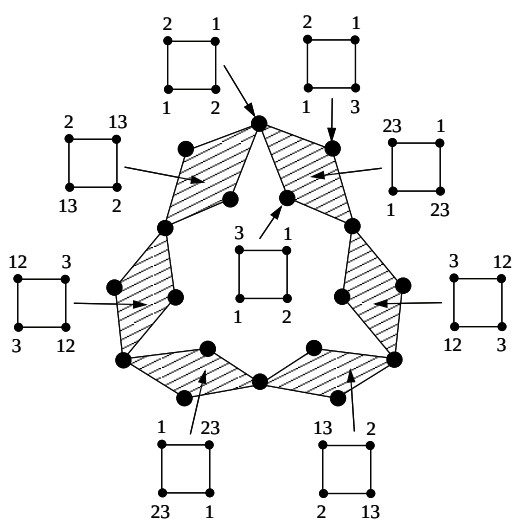


$\mathtt{Hom}\,(L_4,K_3)$

Figure 2.1.3. 3-coloring complex of a 4-string.

(3) In the literature there are several different notations for the set of all graph homomorphisms from a graph T to the graph G. Since an untangling of the definitions shows that this set is precisely the set of vertices of $\mathtt{Hom}\,(T,G)$, i.e., its 0-skeleton, it feels natural to denote it by $\mathtt{Hom}_0(T,G)$.
(4) On the intuitive level, one can think of each $\eta : V(T) \to 2^{V(G)} \setminus \{\emptyset\}$, satisfying the conditions of the Definition 2.1.5, as associating non-empty lists of vertices of G to vertices of T with the condition on this collection of lists being that any choice of one vertex from each list will yield a graph homomorphism from T to G.
(5) The standard way to turn a polyhedral complex into a simplicial one is to take the barycentric subdivision. This is readily done by taking the face poset and then taking its nerve (order complex). So, here, if we consider the partially ordered set $\mathcal{F}(\mathtt{Hom}\,(T,G))$ of all η as in Definition 2.1.5, with the partial order defined by $\tilde{\eta} \le \eta$ if and only if $\tilde{\eta}(v) \subseteq \eta(v)$, for all $v \in V(T)$, then we get that the order complex $\Delta(\mathcal{F}(\mathtt{Hom}\,(T,G)))$ is a barycentric subdivision of $\mathtt{Hom}\,(T,G)$. A cell τ of $\mathtt{Hom}\,(T,G)$ corresponds to the union of all the simplices of $\Delta(\mathcal{F}(\mathtt{Hom}\,(T,G)))$ labeled by the chains with the maximal element τ.

Some examples are shown on Figures 2.1.2, 2.1.3, 2.1.4, and 2.1.5. On these figures we used the following notations: L_n denotes an n-string, i.e., a tree with n vertices and no branching points, C_m denotes a cycle with m vertices, i.e., $V(C_m) = \mathbb{Z}_m$, $E(C_m) = \{(x, x+1), (x+1, x) \,|\, x \in \mathbb{Z}_m\}$.



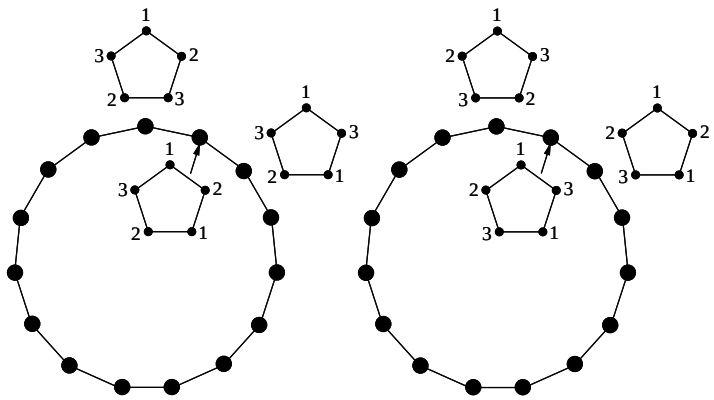
$\mathtt{Hom}\,(C_4, K_3)$

Figure 2.1.4. 3-coloring complex of a 4-cycle.

2.2. Morphism Complexes

2.2.1. General Construction

As mentioned above, one way to interpret Definition 2.1.5 is the following: the cells are indexed by the maps $\eta : V(T) \to 2^{V(G)} \setminus \{\emptyset\}$, such that any choice $\varphi : V(T) \to$



$\mathtt{Hom}\,(C_5, K_3)$

Figure 2.1.5. 3-coloring complex of a 5-cycle.

$V(G)$, satisfying $\varphi(x) \in \eta(x)$, for all $x \in V(T)$, defines a graph homomorphism $\varphi \in \mathtt{Hom}_0(T, G)$. One can generalize this as follows.

Let A and B be two sets, and let M be a collection of some set maps $\varphi : A \to B$. Let $C(A, B) = \prod_{x \in A} \Delta^B$, where Δ^B is the simplex having B as a vertex set, and copies in the direct product are indexed by the elements of A (the direct product is taken in the same weak sense as in the subsection 2.1.2).

Definition 2.2.1. *Let $\mathtt{Hom}_M(A, B)$ be the subcomplex of $C(A, B)$ consisting of all $\sigma = \prod_{x \in A} \sigma_x$, such that any choice $\varphi : A \to B$ satisfying $\varphi(x) \in \sigma_x$, for all $x \in A$, yields a map in M.*

Intuitively one can think of the map φ as the *section* of σ, and the condition can then be verbally stated: *all sections lie in M.* Clearly, the complex $\mathtt{Hom}_M(A, B)$ is always prodsimplicial.

An important fact is that $\mathtt{Hom}_{_}(-,-)$ complexes are fully determined by the low-dimensional data. This result did not previously appear in the literature, so we include here a complete proof.

Proposition 2.2.2. *All complexes $\mathtt{Hom}_{_}(-,-)$ with isomorphic 1-skeletons are isomorphic to each other as polyhedral complexes.*

More precisely, the 1-skeleton determines the complex in the following way: every product of simplices, whose 1-skeleton is in the 1-skeleton of the complex, itself belongs to the complex.

Proof. Let us consider a complex X of the type $\mathtt{Hom}_M(A, B)$, where A, B and M are a priori unknown. Trivially, the 0-skeleton of X is the set M itself. Furthermore, the 1-skeleton tells us which pairs of set maps $\varphi, \psi : A \to B$ differ precisely in one element of A.

Clearly, for $\sigma = \prod_{x \in A} \sigma_x \in C(A, B)$ to belong to $\mathtt{Hom}_M(A, B)$, it is required that the 1-skeleton of σ is a subgraph of the 1-skeleton of $\mathtt{Hom}_M(A, B)$. Let us show that the converse of this statement is true as well.

Let Γ be the 1-skeleton of X. For every edge e in Γ let $\lambda(e) \in A$ denote the element in which the value of the function is changed along e. Since we do not

know the set A, we can only make the statements of the type *the labels of these two edges are the same/different*.

Assume we have $S \subseteq M$, such that S can be written as a direct product $S = S_1 \times S_2 \times \cdots \times S_t$. Assume furthermore that the subgraph of Γ induced by S is precisely the 1-skeleton of the corresponding cell.

First, consider 3 elements $a, b, c \in S_1 \times \cdots \times S_t$, which have the same indices in all S_i's except for exactly one. Then, by our assumption on S, the subgraph of Γ induced on the vertices a, b, and c, is a triangle. Clearly, if 3 changes of a value of a function result in the same function, then the changes were done in the same element of A, i.e., $\lambda(a, b) = \lambda(a, c) = \lambda(b, c)$.

Next, consider 4 elements $a, b, c, d \in S_1 \times \cdots \times S_t$, such that pairs of vertices $(a, b), (b, c), (c, d)$, and (a, d), have the same indices in all S_i's except for exactly one. Assume further that this index is not the same for (a, b) and (b, c): say a and b differ in S_1, and b and c differ in S_2.

According to our assumption on S, $(a, b), (b, c), (c, d)$, and (a, d) are edges of Γ. If $\lambda(a, b) = \lambda(b, c)$, then Γ contains the edge (a, c), and $\lambda(a, b) = \lambda(a, c)$, which contradicts our choice of S.

If, on the other hand, $\lambda(a, b) \neq \lambda(b, c)$, then, since changes of functions along the paths $a \to b \to c$ and $a \to d \to c$ should give the same answer, we are left with the only possibility: namely $\lambda(a, b) = \lambda(c, d)$, and $\lambda(b, c) = \lambda(a, d)$.

Let $a \in S_1 \times \cdots \times S_t$, $a = (a_1, \ldots, a_t)$. By our first argument, if $b \in S_1 \times \cdots \times S_t$, $b = (a_1, \ldots, \tilde{a}_i, \ldots, a_t)$, then $\lambda(a, b)$ does not depend on a_i and $\tilde{a}_i$. Furthermore, let $c, d \in S_1 \times \cdots \times S_t$, $d = (a_1, \ldots, \tilde{a}_j, \ldots, a_t)$, $c = (a_1, \ldots, \tilde{a}_i, \ldots, \tilde{a}_j, \ldots, a_t)$, for $i \neq j$. By our second argument, applied to a, b, c, d, we get that $\lambda(a, b) = \lambda(c, d)$. If iterated for various j, this implies that $\lambda(a, b)$ does not depend on $a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_t$ either; thus it may depend only on the index i.

Finally, this label should be different for different i's, as otherwise, by the same argument as above, we would get more edges in the subgraph of Γ induced by S, than what we allowed by our assumptions.

Summarizing, we have shown that the cell $\sigma = \prod_{x \in A} \sigma_x \in C(A, B)$ belongs to $\mathtt{Hom}_M(A, B)$ if and only if the 1-skeleton of σ is a subgraph of Γ. This implies that $\mathtt{Hom}_M(A, B)$ is uniquely determined by its 1-skeleton. $\square$

Intuitively, one can interpret Proposition 2.2.2 as saying that, with respect to its 1-skeleton, $\mathtt{Hom}_M(A, B)$ is the polyhedral analog of the flag complex construction.

2.2.2. Specifying the Parameters in the General Construction

(1) As mentioned above, if we take A and B to be the sets of vertices of two graphs T and G, and then take M to be the set of graph homomorphisms from T to G, then $\mathtt{Hom}_M(A, B)$ will coincide with $\mathtt{Hom}(T, G)$.

(2) We think of a *directed* graph G as a pair of sets $(V(G), E(G))$, such that $E(G) \subseteq V(G) \times V(G)$.

Definition 2.2.3. *For two directed graphs T and G, a* **directed graph homomorphism** *from T to G is a map $\varphi : V(T) \to V(G)$, such that $(\varphi \times \varphi)(E(T)) \subseteq E(G)$.*

Let A and B be the sets of vertices of two directed graphs T and G, and let M to be the set of directed graph homomorphisms from T to G, then $\mathtt{Hom}_M(A, B)$ is the analog of $\mathtt{Hom}(T, G)$ for directed graphs. An example is shown on Figure 2.2.1.

For a directed graph G, let $u(G)$ be the undirected graph obtained from G by forgetting the directions, and identifying the multiple edges. We remark that for any two directed graphs G and H, the complexes $\mathtt{Hom}(G,H)$ and $\mathtt{Hom}(u(G),u(H))$ are isomorphic, if $E(H)$ is $\mathbb{Z}_2$-invariant.

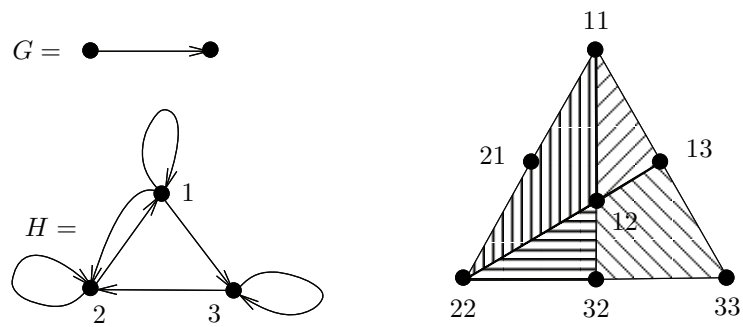


Figure 2.2.1. An example of a $\mathtt{Hom}$ complex for two directed graphs.

(3) Let A and B be the vertex sets of simplicial complexes Δ_1 and Δ_2, and let M be the set of simplicial maps from Δ_1 to Δ_2, then $\mathtt{Hom}_M(A,B)$ is the analog of $\mathtt{Hom}(T,G)$ for simplicial complexes.

(4) Recall that a hypergraph with the vertex set V is a subset $\mathcal{H} \subseteq 2^V$. Let A and B be the vertex sets of hypergraphs $\mathcal{H}_1$ and $\mathcal{H}_2$. There are various choices for when to call a map $\varphi : A \to B$ a *hypergraph homomorphism*. Two possibilities which we mention here are: one could require that $\varphi(\mathcal{H}_1) \subseteq \mathcal{H}_2$, or one could ask that for any $H_1 \in \mathcal{H}_1$, there exists $H_2 \in \mathcal{H}_2$, such that $\varphi(H_1) \subseteq H_2$. The example (3) is a special case of both. Either way, the corresponding complex $\mathtt{Hom}_M(A,B)$ provides us with an analog of $\mathtt{Hom}(T,G)$ for hypergraphs.

(5) Let A and B be the vertex sets of posets P and Q, and let M be the set of order-preserving maps from P to Q, then $\mathtt{Hom}_M(A,B)$ is the analog of $\mathtt{Hom}(T,G)$ for posets.

2.3. Historic Detour

2.3.1. The Kneser-Lovász Theorem

The Kneser conjecture was posed in 1955, see [**Kn55**], and concerned chromatic numbers of a specific family of graphs, later called *Kneser graphs*. We denoted these graphs $K_{n,k}$, see Definition 1.2.3. In 1978 L. Lovász solved the Kneser conjecture by finding geometric obstructions of Borsuk-Ulam type to the existence of graph colorings.

Theorem 2.3.1 (Kneser-Lovász, [**Kn55, Lov78**]).
For arbitrary positive integers n, k, such that $n \geq 2k$, we have $\chi(K_{n,k}) = n-2k+2$.

To show the inequality $\chi(K_{n,k}) \geq n-2k+2$ Lovász associated the neighborhood complex $\mathcal{N}(G)$, see Definition 2.1.3, to an arbitrary graph G, and then used the connectivity information of the topological space $\mathcal{N}(G)$ to find obstructions to the colorability of G. More precisely, he proved the following statement.

Theorem 2.3.2 (Lovász, [**Lov78**]).
Let G be a graph, such that $\mathcal{N}(G)$ is k-connected for some $k \in \mathbb{Z}$, $k \geq -1$, then $\chi(G) \geq k+3$.

The main topological tool which Lovász employed was the Borsuk-Ulam theorem. Shorter proofs were obtained by Bárány, [**Bar78**], and Greene, [**Gr02**], both using some versions of the Borsuk-Ulam theorem, see also [**GR01**]. A nice brief survey of these can be found in [**dL03**].

Since these developments, the topological equivariant methods have gained ground and became a part of the standard repertoire in combinatorics. We refer the reader to the series of papers [**Ziv96, Ziv97, Ziv98**] for an excellent introduction to the subject.

We have seen in Proposition 2.1.4, that the complexes $\mathcal{N}(G)$ and $\mathtt{Hom}\,(K_2, G)$ have the same simple homotopy type. This fact leads one to consider the family of $\mathtt{Hom}$ complexes as a natural context in which to look for further obstructions to the existence of graph homomorphisms.

2.3.2. Later Developments

2.3.2.1. *The Vertex-Critical Subgraphs of Kneser Graphs.*

For a graph G and a vertex $v \in V(G)$ we introduce the following notation: $G - v$ denotes the graph which is obtained from G by deleting the vertex v and all edges adjacent to v, i.e., $V(G-v) = V(G) \setminus \{v\}$, and $E(G-v) = E(G) \cap (V(G-v) \times V(G-v))$.

Shortly after Lovász' result, Schrijver, in [**Sch78**], has sharpened Theorem 2.3.1. To formulate his result, we recall that a graph G is called *vertex-critical* if, for any vertex $v \in V(G)$, we have $\chi(G) = \chi(G-v) + 1$.

Definition 2.3.3. *Let n, k be positive integers, $n \geq 2k$. The* **stable Kneser graph** *$K^{\mathrm{stab}}_{n,k}$ is defined to be the graph whose set of vertices is the set of all k-subsets S of $[n]$, such that if $i \in S$, then $i+1 \notin S$, and, if $n \in S$, then $1 \notin S$. Two subsets are joined by an edge if and only if they are disjoint.*

Clearly, $K^{\mathrm{stab}}_{n,k}$ is an induced subgraph of $K_{n,k}$.

Theorem 2.3.4 (Schrijver, [**Sch78**]).
$K^{\mathrm{stab}}_{n,k}$ is a vertex-critical subgraph of $K_{n,k}$, i.e., $K^{\mathrm{stab}}_{n,k}$ is a vertex-critical graph, and $\chi(K^{\mathrm{stab}}_{n,k}) = n - 2k + 2$.

2.3.2.2. *Chromatic Numbers of Kneser Hypergraphs.*

In 1986, Alon-Frankl-Lovász, [**AFL86**], have generalized Theorem 2.3.1 to the case of hypergraphs. To start with, recall the standard way to extend the notion of the chromatic number to hypergraphs.

Definition 2.3.5. *For a hypergraph $\mathcal{H}$, the chromatic number $\chi(\mathcal{H})$ is, by definition, the minimal number of colors needed to color the vertices of $\mathcal{H}$ so that no hyperedge is monochromatic.*

Next, there is a standard way to generalize Definition 1.2.3 to the case of hypergraphs.

Definition 2.3.6. *Let n, k, r be positive integers, such that $r \geq 2$, and $n \geq rk$. The* **Kneser r-hypergraph** *$K^r_{n,k}$ is the r-uniform hypergraph, whose ground set*

consists of all k-subsets of $[n]$, and the set of hyperedges consists of all r-tuples of disjoint k-subsets.

Using the introduced notations, we can now formulate the generalization of Theorem 2.3.1.

Theorem 2.3.7 (Alon-Frankl-Lovász, [**AFL86**]).
For arbitrary positive integers n, k, r, such that $r \geq 2$, and $n \geq rk$, we have

$$\chi(K^r_{n,k}) = \left\lceil \frac{n - rk + r}{r - 1} \right\rceil .$$

Theorem 2.3.7 can be proved using the generalization of the Borsuk-Ulam theorem from [**BSS81**].

2.3.2.3. *Further References.*

There has been a substantial body of further important work, which, due to space constraints, we do not pursue in detail in this survey, some of the references are [**Dol88, Kr92, Kr00, Ma04, MZ04, Sar90, Zie02**].

There have also been multiple constructions, such as box complexes, designed to generalize the original Lovász neighbourhood complexes. However, as later research showed, the bounds obtained in that way were essentially convertible, since the $\mathbb{Z}_2$-homotopy types of these complexes were very closely related, either by simply being the same, or by means of one being the suspension of another, or something close to that. This means that all these constructions are avatars of the same object, as explained in [**Ziv04**].

2.4. More about the $\mathtt{Hom}$-Complexes

2.4.1. Coproducts

For any three graphs G, H, and K, we have

$$\mathtt{Hom}(G \coprod H, K) = \mathtt{Hom}(G, K) \times \mathtt{Hom}(H, K), \tag{2.4.1}$$

and, if G is connected, and $G \neq K_1$, then also

$$\mathtt{Hom}(G, H \coprod K) = \mathtt{Hom}(G, H) \coprod \mathtt{Hom}(G, K),$$

where the equality denotes isomorphism of polyhedral complexes.

The first formula is obvious. To verify the second one, note that, for any graph homomorphism $\eta : V(G) \to 2^{V(H) \cup V(K)} \setminus \{\emptyset\}$, and any $x, y \in V(G)$, such that $(x, y) \in E(G)$, if $\eta(x) \cap V(H) \neq \emptyset$, then $\eta(y) \subseteq V(H)$, which under the assumptions on G implies that $\bigcup_{x \in V(G)} \eta(x) \subseteq V(H)$.

2.4.2. Products

For any three graphs G, H, and K, we have the following homotopy equivalence, see [**Ba05**]:

$$\mathtt{Hom}(G, H \times K) \simeq \mathtt{Hom}(G, H) \times \mathtt{Hom}(G, K). \tag{2.4.2}$$

In fact, the formula (2.4.2) can be strengthened to state that the left hand side is simple homotopy equivalent (in the sense of Whitehead, see [**Co73**]) to the right hand side. Since this simple homotopy equivalence result is new, we include a complete argument, as promised in the abstract.

Consider the following three maps $2^{p_H} : 2^{V(H)\times V(K)} \to 2^{V(H)}$, $2^{p_K} : 2^{V(H)\times V(K)} \to 2^{V(K)}$, and $c : 2^{V(H)} \times 2^{V(K)} \to 2^{V(H)\times V(K)}$, where 2^{p_H} and 2^{p_K} are induced by the standard projection maps $p_H : V(H) \times V(K) \to V(H)$ and $p_K : V(H) \times V(K) \to V(K)$, and c is given by $c(A, B) = A \times B$. We let $\psi : 2^{V(H)\times V(K)} \to 2^{V(H)\times V(K)}$ denote the composition map $\psi(S) = c(2^{p_H}(S), 2^{p_K}(S)) = 2^{p_H}(S) \times 2^{p_K}(S)$.

Given a cell of $\mathtt{Hom}\,(G, H \times K)$ indexed by $\eta : V(G) \to 2^{V(H)\times V(K)} \setminus \{\emptyset\}$, one can see that the composition function $\psi \circ \eta : V(G) \to 2^{V(H)\times V(K)} \setminus \{\emptyset\}$ will also index a cell. Indeed, for any $(x, y) \in E(G)$ we know that $(\eta(x), \eta(y))$ is a complete bipartite subgraph of $H \times K$, which is the same as to say that, for any $\alpha \in \eta(x)$, and $\beta \in \eta(y)$, we have $(p_H(\alpha), p_H(\beta)) \in E(H)$, and $(p_K(\alpha), p_K(\beta)) \in E(K)$. If we now choose $\tilde{\alpha} \in \psi(\eta(x))$, and $\tilde{\beta} \in \psi(\eta(y))$, we have $p_H(\tilde{\alpha}) = p_H(\alpha)$, for some $\alpha \in \eta(x)$, and $p_H(\tilde{\beta}) = p_H(\beta)$, for some $\beta \in \eta(y)$, hence verifying that $(p_H(\tilde{\alpha}), p_H(\tilde{\beta})) \in E(H)$. The fact that $(p_K(\tilde{\alpha}), p_K(\tilde{\beta})) \in E(K)$ can be proved analogously.

This means that we have a map $\varphi : \mathcal{F}(\mathtt{Hom}\,(G, H \times K)) \to \mathcal{F}(\mathtt{Hom}\,(G, H \times K))$. It is easy to see that φ is order-preserving and ascending (meaning $\varphi(x) \geq x$, for any $x \in \mathcal{F}(\mathtt{Hom}\,(G, H \times K))$). It follows from [**Ko05a**, Theorem 3.1] that $\Delta(\mathcal{F}(\mathtt{Hom}\,(G, H \times K))) = \mathrm{Bd}\,(\mathtt{Hom}\,(G, H \times K))$ collapses onto $\Delta(\mathrm{im}\,\varphi)$.

On the other hand, $\mathcal{F}(\mathtt{Hom}\,(G, H)) \times \mathcal{F}(\mathtt{Hom}\,(G, K)) \cong \mathrm{im}\,\varphi$ with the isomorphism given by the map $(\eta_1, \eta_2) \mapsto \eta$, where $\eta(x) = \eta_1(x) \times \eta_2(x)$, for any $x \in V(G)$. Thus we conclude that $\mathrm{Bd}\,(\mathtt{Hom}\,(G, H \times K))$ collapses onto $\Delta(\mathcal{F}(\mathtt{Hom}\,(G, H)) \times \mathcal{F}(\mathtt{Hom}\,(G, K))) \cong \Delta(\mathcal{F}(\mathtt{Hom}\,(G, H))) \times \Delta(\mathcal{F}(\mathtt{Hom}\,(G, K))) = \mathrm{Bd}\,(\mathtt{Hom}\,(G, H)) \times \mathrm{Bd}\,(\mathtt{Hom}\,(G, K)) \cong \mathtt{Hom}\,(G, H) \times \mathtt{Hom}\,(G, K)$, and our argument is now complete.

For the analog of the formula (2.4.2), where the direct product is taken on the left, we need the following additional standard notion.

Definition 2.4.1. *For two graphs H and K, the* **power graph** K^H *is defined by*

- $V(K^H)$ *is the set of all set maps* $f : V(H) \to V(K)$*;*
- $(f, g) \in E(K^H)$*, for* $f, g : V(H) \to V(K)$*, if and only if, whenever* $(v, w) \in E(H)$*, we also have* $(f(v), g(w)) \in E(K)$*.*

It is easy to see that the power graph notion is introduced precisely so that for any triple of graphs the following adjunction relation holds:

$$\mathtt{Hom}_0(G \times H, K) = \mathtt{Hom}_0(G, K^H). \tag{2.4.3}$$

In our topological situation the formula (2.4.3) generalizes up to homotopy. More precisely, we have the following homotopy equivalence, see [**Ba05**],

$$\mathtt{Hom}\,(G \times H, K) \simeq \mathtt{Hom}\,(G, K^H). \tag{2.4.4}$$

The formula (2.4.4) can as well be strengthened to yield a simple homotopy equivalence. Below we include a complete argument.

Define a map $\psi : 2^{V(K^H)} \to 2^{V(K^H)}$, $\psi : \Omega \mapsto \psi(\Omega)$, as follows: $g \in \psi(\Omega)$ if and only if $g(x) \in \{f(x) \mid f \in \Omega\}$, for all $x \in V(H)$. In other words, we use the collection of functions Ω to specify the sets of values, which functions from $\psi(\Omega)$ are allowed to take. Clearly, we have $\psi(\Omega) \supseteq \Omega$. Take a cell of $\mathtt{Hom}\,(G, K^H)$, $\eta : V(G) \to 2^{V(K^H)} \setminus \{\emptyset\}$, and consider the composition map $\psi \circ \eta : V(G) \to 2^{V(K^H)} \setminus \{\emptyset\}$.

Since η is a cell, we know that if $(x,y) \in E(G)$, and $\alpha \in \eta(x)$, $\beta \in \eta(y)$, then $(\alpha,\beta) \in E(K^H)$, i.e., whenever $(v,w) \in E(H)$, we have $(\alpha(v),\beta(w)) \in E(K)$.

Choose $\tilde{\alpha} \in \psi(\eta(x))$, and $\tilde{\beta} \in \psi(\eta(y))$. To check that $(\tilde{\alpha},\tilde{\beta}) \in E(K^H)$, we need to check that for any $(v,w) \in E(H)$, we have $(\tilde{\alpha}(v),\tilde{\beta}(w)) \in E(K)$. However, by the definition of ψ, we know that $\tilde{\alpha}(v) = \alpha(v)$, for some $\alpha \in \eta(x)$, and $\tilde{\beta}(w) = \beta(w)$, for some $\beta \in \eta(y)$. It follows that $(\tilde{\alpha}(v),\tilde{\beta}(w)) = (\alpha(v),\beta(w)) \in E(K)$, and hence $\psi \circ \eta$ is again a cell.

As a consequence, the composition gives us an order-preserving ascending map $\varphi : \mathcal{F}(\mathtt{Hom}(G,K^H)) \to \mathcal{F}(\mathtt{Hom}(G,K^H))$. The image of this map is isomorphic to $\mathcal{F}(\mathtt{Hom}(G \times H,K))$. The isomorphism map takes the poset element $\eta : V(G) \times V(H) \to 2^{V(K)} \setminus \{\emptyset\}$ to the poset element $\tilde{\eta} : V(G) \to 2^{V(K^H)} \setminus \{\emptyset\}$ defined by

$$\tilde{\eta}(x) = \{f : V(H) \to V(K) \mid f(v) \in \eta(x,v), \text{ for all } v \in V(H)\},$$

for all $x \in V(G)$. By [**Ko05a**, Theorem 3.1], we conclude that the complex $\Delta(\mathcal{F}(\mathtt{Hom}(G,K^H))) = \mathrm{Bd}(\mathtt{Hom}(G,K^H))$ collapses onto its subcomplex $\Delta(\operatorname{im}\varphi) = \mathrm{Bd}(\mathtt{Hom}(G \times H,K))$.

We obtain an interesting special case of the formula (2.4.4) when substituting $G = K_1^o$ (which means a graph with one looped vertex). Since $K_1^o \times H = H$, for any graph H, we conclude that $\mathtt{Hom}(H,K) \simeq \mathtt{Hom}(K_1^o,K^H)$ for any two graphs H and K. As seen directly, for an arbitrary graph G, $\mathtt{Hom}(K_1^o,G)$ is the clique complex of the looped part of G, i.e., of the subgraph induced by the set of vertices which have loops. In particular, the complex $\mathtt{Hom}(K_1^o,G)$ is simplicial. On the other hand, a vertex $f \in V(K^H)$ has a loop if and only if f is a graph homomorphism. We can therefore conclude that for arbitrary graphs H and K the complex $\mathtt{Hom}(H,K)$ is homotopy equivalent to the clique complex of the subgraph of K^H, induced by the set of all graph homomorphisms from H to K.

2.4.3. Associated Covariant and Contravariant Functors

For an arbitrary polyhedral complex X, we let $\mathcal{F}(X)$ denote its face poset ordered by inclusion. The notion of a *link* of a vertex of a polyhedral complex allows several interpretations, so let us fix our convention here. Let v be a vertex of X, the link of v, denoted $\mathrm{lk}_X(v)$, is the cell complex whose face poset is given by $\mathcal{F}_{>v}(X)$. Geometrically, $\mathrm{lk}_X(v)$ can be obtained as follows. Realize faces of X as polyhedra, in a coherent manner, and take ε to be a positive number which is smaller then the minimal length of an edge from v. Each face F containing v can be truncated at vertex v, by cutting it along the set of points at distance ε from v. These cuts fit coherently to form the desired cell complex. For example, the link of any vertex of a cube is a triangle, and in general, the link of any vertex of a polytope K is the polytope obtained by truncating K at the vertex v.

Let T, G, and K be three arbitrary graphs, and let φ be a graph homomorphism from G to K. Then the composition induces a poset map $f : \mathcal{F}(\mathtt{Hom}(T,G)) \to \mathcal{F}(\mathtt{Hom}(T,K))$, namely, for $\eta : V(T) \to 2^{V(G)} \setminus \{\emptyset\}$, we have $f(\eta) = 2^{\varphi} \circ \eta$, where 2^{φ} is the map induced on the subsets.

Recall that for arbitrary regular CW complexes A and B, a poset map $f : \mathcal{F}(A) \to \mathcal{F}(B)$ comes from a cellular map $\varphi : A \to B$ (meaning that $f = \mathcal{F}(\varphi)$), if and only if $\mathrm{rk}\,\varphi(x) \le \mathrm{rk}\,x$, for all $x \in \mathcal{F}(A)$. It is not difficult to check that this condition is satisfied by the poset map $f : \mathcal{F}(\mathtt{Hom}(T,G)) \to \mathcal{F}(\mathtt{Hom}(T,K))$

defined above. Hence, we can conclude that this f comes from a cellular map from $\mathtt{Hom}\,(T,G)$ to $\mathtt{Hom}\,(T,K)$, which we denote by φ^T.

Moreover, a detailed pointwise analysis of the polyhedral structure of $\mathtt{Hom}\,(T,G)$ shows that cells (direct products of simplices) map surjectively to other cells, and that this map is a product map induced by the corresponding maps on the simplices. Therefore, φ^T is a polyhedral map.

The situation is slightly more complicated if one considers the functoriality in the first argument. Let us choose some *proper* graph homomorphism ψ from T to G, and let K be some graph. Again, by using composition we can define a poset map $g : \mathcal{F}(\mathtt{Hom}\,(G,K)) \to \mathcal{F}(\mathtt{Hom}\,(T,K))$, namely, for $\eta : V(G) \to 2^{V(K)} \setminus \{\emptyset\}$, and $v \in V(T)$, we have $g(\eta)(v) = \eta(\psi(v))$. This map is well-defined, since, first if $v, w \in V(T)$, and $(v,w) \in E(T)$, then $(\psi(v), \psi(w)) \in E(G)$, and therefore, for any $x \in \eta(\psi(v))$, and $y \in \eta(\psi(w))$, we have $(x,y) \in E(K)$, and second, by the properness assumption, $\sum_{v\in V(T)}(|\eta(\psi(v))| - 1) < \infty$. Furthermore, this map is order-preserving: if $\tau \geq \eta$, i.e., if $\tau(w) \supseteq \eta(w)$, for any $w \in V(T)$, then $g(\tau)(w) = \tau(\psi(w)) \supseteq \eta(\psi(w)) = g(\eta)(w)$.

Intuitively, one can think of the map g as the pullback map. It is important to remark that, if ψ is not injective, it may happen that $\dim g(\eta) > \dim \eta$.

For an arbitrary regular CW complex X, let $\mathrm{Bd}\,(X)$ denote the *barycentric subdivision* of X. Since g is an order-preserving map, the induced map

$$\Delta(g) : \mathrm{Bd}\,(\mathtt{Hom}\,(G,K)) \to \mathrm{Bd}\,(\mathtt{Hom}\,(T,K))$$

is simplicial and gives the corresponding map of topological spaces, which we denote ψ_K. However, g does not always come from a cellular map. In fact, one can check that there exists a cellular map $\psi_K : \mathtt{Hom}\,(G,K) \to \mathtt{Hom}\,(T,K)$, such that $\mathcal{F}(\psi_K) = g$, if ψ is injective on the vertices of T.

In any case, we see that $\mathtt{Hom}\,(T,-)$ is a covariant functor from **Graphs** to **Top**, while $\mathtt{Hom}\,(-,K)$ is a contravariant functor from $\mathbf{Graphs_p}$ to **Top**; here **Top** denotes the category whose objects are topological spaces, and whose morphisms are all continuous maps.

2.4.4. Composition of Hom's

For three arbitrary graphs T, G, and K, there is a composition map

$$\xi : \mathcal{F}(\mathtt{Hom}\,(T,G)) \times \mathcal{F}(\mathtt{Hom}\,(G,K)) \longrightarrow \mathcal{F}(\mathtt{Hom}\,(T,K)),$$

whose detailed description is as follows: for graph homomorphisms $\alpha : V(T) \to 2^{V(G)}\setminus\{\emptyset\}$, and $\beta : V(G) \to 2^{V(K)}\setminus\{\emptyset\}$, define the map $\tilde{\beta} : 2^{V(G)}\setminus\{\emptyset\} \to 2^{V(K)}\setminus\{\emptyset\}$ by

$$\text{for } S \in 2^{V(G)} \setminus \{\emptyset\}, \quad \tilde{\beta}(S) := \cup_{x\in S}\beta(x),$$

and then set $\xi(\alpha,\beta) := (\tilde{\beta} \circ \alpha : V(T) \to 2^{V(K)} \setminus \{\emptyset\})$. It is easy to check that this map is well-defined. Indeed, let $x, y \in V(T)$, such that $(x,y) \in E(T)$, choose arbitrary $a \in \alpha(x)$, and $b \in \alpha(y)$, and then choose arbitrary $\tilde{a} \in \beta(a)$, and $\tilde{b} \in \beta(b)$. Clearly, $(x,y) \in E(T)$ implies $(a,b) \in E(G)$, since α is a graph homomorphism, which then implies $(\tilde{a},\tilde{b}) \in E(K)$, since β is a graph homomorphism.

Applying the nerve functor Δ to the poset map ξ we get a simplicial map

$$\Delta(\xi) : \Delta(\mathcal{F}(\mathtt{Hom}\,(T,G)) \times \mathcal{F}(\mathtt{Hom}\,(G,K))) \longrightarrow \mathrm{Bd}\,(\mathtt{Hom}\,(T,K)),$$

and hence, since for any posets P_1 and P_2, the simplicial complex $\Delta(P_1 \times P_2)$ is homeomorphic to the polyhedral complex $\Delta(P_1) \times \Delta(P_2)$ (in fact it is its subdivision), we have a corresponding topological map

$$\mathtt{Hom}(T,G) \times \mathtt{Hom}(G,K) \longrightarrow \mathtt{Hom}(T,K).$$

2.4.5. Action of Automorphism Groups

An important special case of the situation described in the Subsection 2.4.4 is when the considered graph homomorphisms are actually isomorphisms. In other words, for arbitrary graphs T and G, the elements $\varphi \in \mathrm{Aut}(T)$ and $\psi \in \mathrm{Aut}(G)$ induce polyhedral maps $\varphi_G : \mathtt{Hom}(T,G) \to \mathtt{Hom}(T,G)$ and $\psi^T : \mathtt{Hom}(T,G) \to \mathtt{Hom}(T,G)$, which are easily shown to be isomorphisms.

Summarizing, we have a polyhedral action of the group $\mathrm{Aut}(T) \times \mathrm{Aut}(G)$ on the polyhedral complex $\mathtt{Hom}(T,G)$. As an example, we have $\mathcal{S}_m \times \mathcal{S}_n$-action on $\mathtt{Hom}(K_m,K_n)$, and $\mathcal{D}_m \times \mathcal{S}_n$-action on $\mathtt{Hom}(C_m,K_n)$, where $\mathcal{S}_n$ is the n-th symmetric group, and $\mathcal{D}_n$ is the n-th dihedral group.

We note the following useful fact: if for some vertex v there exists a group element $\varphi \in \mathrm{Aut}(T)$, such that $(v,\varphi(v)) \in E(T)$ (for example, if φ flips an edge in T), then the induced map $\varphi_G : \mathtt{Hom}(T,G) \to \mathtt{Hom}(T,G)$ is fixed-point free for an arbitrary graph G without loops. For example, $\mathbb{Z}_2$ (corresponding to an arbitrary reflection from $\mathcal{D}_{2r+1}$) acts freely on $\mathtt{Hom}(C_{2r+1},K_n)$.

2.4.6. Universality

In topological combinatorics it happens very often that the family of the studied objects is universal with respect to the invariants which one is interested in computing. This is also the case not only for the $\mathtt{Hom}$-complexes, but even for the $\mathtt{Hom}(K_2,-)$-complexes. The following result is due to Csorba, [**Cs04a**], and, independently, to Živaljević, [**Ziv04**].

Theorem 2.4.2 ([**Cs04a, Ziv04**]).
For each finite, free $\mathbb{Z}_2$-complex X, there exists a graph G, such that $\mathtt{Hom}(K_2,G)$ is $\mathbb{Z}_2$-homotopy equivalent to X.

We note that Theorem 2.4.2 can be verified by combining [**Ziv04**, Theorem 32], with a remark in the beginning of [**Ziv04**, Section 7].

2.5. Folds

2.5.1. Sequences of Collapses Induced by Folds

$\mathtt{Hom}$-complexes behave well with respect to the following standard operation from graph theory.

Definition 2.5.1. *For a graph G and $v \in V(G)$, $G - v$ is called a* **fold** *of G if there exists $u \in V(G)$, $u \neq v$, such that $\mathcal{N}(u) \supseteq \mathcal{N}(v)$.*

Let $G - v$ be a fold of G. We let $i : G - v \hookrightarrow G$ denote the inclusion homomorphism, and let $f : G \to G - v$ denote the folding homomorphism defined by

$$f(x) = \begin{cases} u, & \text{for } x = v; \\ x, & \text{for } x \neq v. \end{cases}$$

Note that i is a graph homomorphism for an arbitrary choice of $v \in V(G)$, whereas f is a graph homomorphism if and only if $G - v$ is a fold, in particular, this could be taken as an alternative definition of the fold.

Let X be a polyhedral complex. Recall that an *elementary collapse* of X is a removal of a pair of polyhedra (σ, τ), such that σ is a maximal polyhedron, $\dim \tau = \dim \sigma - 1$, and σ is the only maximal polyhedron containing τ. Furthermore, let Y be a subcomplex of X. We say that X *collapses onto* Y if there exists a sequence of elementary collapses leading from X to Y. If X collapses onto Y, then Y is a strong deformation retract of X.

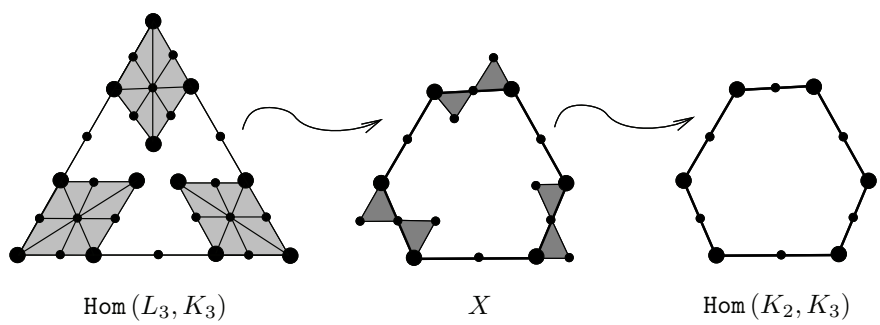


Figure 2.5.1. A two-step folding of the first argument in $\mathtt{Hom}\,(L_3, K_3)$.

Theorem 2.5.2 ([**Ko04**, Theorem 3.3])**.**
Let $G - v$ be a fold of G and let H be some graph. Then

(1) $\mathrm{Bd}\,\mathtt{Hom}\,(G, H)$ *collapses onto* $\mathrm{Bd}\,\mathtt{Hom}\,(G - v, H)$*;*
(2) $\mathtt{Hom}\,(H, G)$ *collapses onto* $\mathtt{Hom}\,(H, G - v)$*.*

The maps i_H and f^H are strong deformation retractions.

Figure 2.5.1 shows an example of the collapsing sequence appearing in the proof of Theorem 2.5.2 (1).

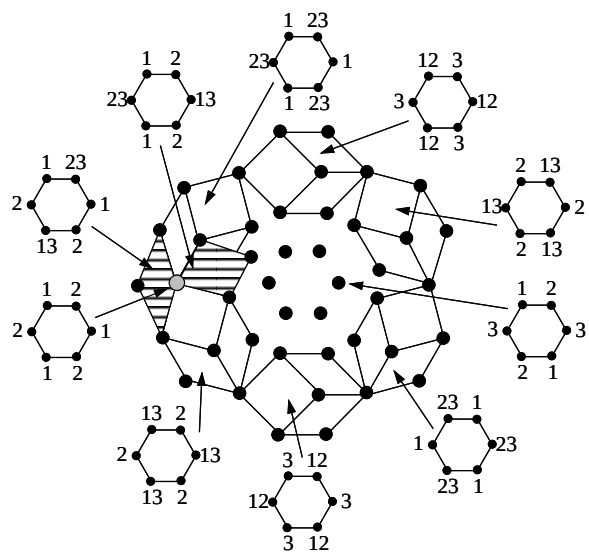
Note that Theorem 2.5.2 cannot be generalized to encompass arbitrary graph homomorphisms ϕ of G onto H, where H is a subgraph of G, and ϕ is identity on H. For example, $\mathtt{Hom}\,(C_6, K_3) \not\simeq \mathtt{Hom}\,(K_2, K_3)$, see Figures 2.5.2 and 2.5.3, despite of the existence of the sequence of graph homomorphisms $K_2 \to C_6 \to K_2$ which compose to give an identity.

We remark that, for the sake of transparency, the striped rectangles are shown on Figure 2.5.2 only around one of the 6 joining vertices, and only two out of the three. The big connected component corresponds to the graph homomorphisms $\varphi : C_6 \to K_3$ having the winding number 0. The isolated points correspond to the 6 possible tight windings of C_6 around K_3. Observe also that the cubes are solid.

2.5.2. Applications

When G is a graph, and H is an induced subgraph of G, we say that G *reduces to* H, if there exists a sequence of folds leading from G to H.

Corollary 2.5.3 ([**BK03b**, Corollary 5.3])**.**
Let G be a graph, and $S \subseteq V[G]$, such that G reduces to $G[S]$. Assume S is Γ-invariant for some $\Gamma \subseteq \mathrm{Aut}(G)$. Then the inclusion $i : G[S] \hookrightarrow G$ induces a Γ-invariant homotopy equivalence $i_H : \mathtt{Hom}(G,H) \to \mathtt{Hom}(G[S],H)$ for an arbitrary graph H.



$\mathtt{Hom}(C_6, K_3)$

Figure 2.5.2. The figure depicts the polyhedral complex of all 3-colorings of a 6-cycle.

The Theorem 2.5.2 can be used to obtain complete understanding of the homotopy type of the $\mathtt{Hom}$-complexes for certain specific families of graphs.

Proposition 2.5.4. *If T is a tree with at least one edge, and G an arbitrary graph, then $\mathtt{Hom}(T,G)$ is homotopy equivalent to $\mathcal{N}(G)$. As a consequence, if F is a forest, and $T_1, \ldots, T_k$ are all its connected components consisting of at least 2 vertices, then $\mathtt{Hom}(F,G) \simeq \prod_{i=1}^{k} \mathcal{N}(G)$.*

An even more special case was important in [**BK03b, BK04**] for the proof of Lovász Conjecture.

Corollary 2.5.5 ([**BK03b**, Proposition 5.4])**.**
If T is a finite tree with at least one edge, then the map $i_{K_n} : \mathtt{Hom}(T,K_n) \to \mathtt{Hom}(K_2,K_n)$ induced by any inclusion $i : K_2 \hookrightarrow T$ is a homotopy equivalence, in particular $\mathtt{Hom}(T,K_n) \simeq S^{n-2}$.
If F is a finite forest, and $T_1, \ldots, T_k$ are all its connected components consisting of at least 2 vertices, then $\mathtt{Hom}(F,K_n) \simeq \prod_{i=1}^{k} S^{n-2}$.

In this case, Corollary 2.5.3 can be applied to describe the $\mathbb{Z}_2$-homotopy type as well. First, some new notations: let S^n_a, resp. S^n_t, denote the n-dimensional sphere, equipped with an antipodal, resp. trivial $\mathbb{Z}_2$-action, where n is a nonnegative integer, or infinity.

Proposition 2.5.6 ([**BK03b**, Proposition 5.5])**.**
Let T be a finite tree with at least one edge and a $\mathbb{Z}_2$-action determined by an invertible graph homomorphism $\gamma : T \to T$. If γ flips an edge in T, then $\mathtt{Hom}\,(T, K_n) \simeq_{\mathbb{Z}_2} S_a^{n-2}$*, otherwise* $\mathtt{Hom}\,(T, K_n) \simeq_{\mathbb{Z}_2} S_t^{n-2}$.

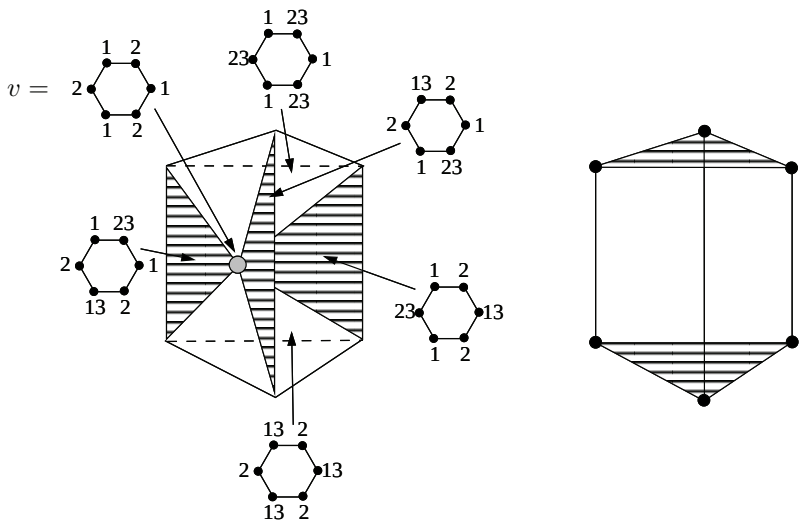


Figure 2.5.3. The figure on the left shows the neighbourhood of the vertex v of $\mathtt{Hom}\,(C_6, K_3)$. The figure on the right shows the link of this vertex.

Curiously, another computable special case is that of an unlooped complement of a forest.

Proposition 2.5.7. *Let F be a finite forest, and let G be an arbitrary graph, then* $\mathtt{Hom}\,(\overline{F}, G) \simeq \mathtt{Hom}\,(K_m, G)$*, where m is the maximal cardinality of an independent set in F.*

In particular, as was shown in [**BK03b**, Proposition 5.6] that $\mathtt{Hom}\,(\overline{F}, K_n)$ is homotopy equivalent to $\mathtt{Hom}\,(K_m, K_n)$, and hence, by Theorem 3.3.3, to a wedge of $(n-m)$-dimensional spheres.

Remark 2.5.8. *Recently, folds gained further prominence in connection with* $\mathtt{Hom}$ *complexes. One has discovered, see* [**Ba05**]*, that it is possible to introduce a natural Quillen model category structure, see* [**GM96**, Chapter V]*, and* [**Qu73**]*, on the category of graphs, such that the weak equivalences are precisely the maps, which allow factorizations into a sequence of folds and unfolds (which therefore may be viewed as trivial homotopy equivalences).*

LECTURE 3
Stiefel-Whitney Classes and First Applications

3.1. Elements of the Principal Bundle Theory

3.1.1. Spaces with a Free Action of a Finite Group and Special Cohomology Elements

Consider a regular CW complex X with a cellular action of a finite group Γ. If desired, the Γ-action can be made to be simplicial by passing to the barycentric subdivision, (cf. [**Bre72, Hat02**]).

For the interested reader we remark here that sometimes one takes the barycentric subdivision even if the original action already was simplicial. The main point of this is that one can make the action enjoy an additional property: *if a simplex is preserved by one of the group elements, then it must be pointwise fixed by this element.*

Next, assume that Γ acts freely. In this case X is called a Γ-*space*. We know that, by the general theory of principal Γ-bundles, see e.g., [**tD87**], there exists a Γ-equivariant map $w : X \to \mathbf{E}\Gamma$, and, that the induced quotient map $w/\Gamma : X/\Gamma \to \mathbf{E}\Gamma/\Gamma = \mathbf{B}\Gamma$ is unique up to homotopy. Here $\mathbf{E}\Gamma$ denotes a contractible space on which the group Γ acts freely, and $\mathbf{B}\Gamma$ denotes the *classifying space* of Γ (also known as the associated Eilenberg-MacLane space, or $K(\Gamma, 1)$-space, see e.g., [**AM94, Bre93, GM96, Hat02, McL95, May99, Wh78**]).

Passing to cohomology, we see that the induced map $(w/\Gamma)^* : H^*(\mathbf{B}\Gamma) \to H^*(X/\Gamma)$ does not depend on the choice of w, and thus, the image of $(w/\Gamma)^*$ consists of some canonically distinguished cohomology elements. For $z \in H^*(\mathbf{B}\Gamma)$, we let $w(z, X)$ denote the element $(w/\Gamma)^*(z)$, which we call *characteristic class* associated to z.

Let Y be another regular CW complex with a free action of Γ, and assume that $\varphi : X \to Y$ is a Γ-equivariant map. By what is said above, there exists a Γ-map $v : Y \to \mathbf{E}\Gamma$. Hence, in addition to the map $w : X \to \mathbf{E}\Gamma$, we also have a composition map $v \circ \varphi$. Passing on to the quotient map and then to the induced map on cohomology, we get yet another map $((v \circ \varphi)/\Gamma)^* : H^*(\mathbf{B}\Gamma) \to H^*(X/\Gamma)$. However, as we mentioned above, the map on the cohomology algebras does not depend on the choice of the original map to $\mathbf{E}\Gamma$. Thus, since $((v \circ \varphi)/\Gamma)^* = (\varphi/\Gamma)^* \circ (v/\Gamma)^*$,

we have commuting diagrams, see Figure 3.1.1, and therefore

$$w(z, X) = (\varphi/\Gamma)^*(w(z, Y)),$$

where z is an arbitrary element of $H^*(\mathbf{B}\Gamma)$.

$$\begin{tikzcd} X \arrow[r, "\varphi"] \arrow[dr, "v \circ \varphi"'] & Y \arrow[d, "v"] \\ & \mathbf{E}\Gamma \end{tikzcd} \qquad \begin{tikzcd} H^*(X/\Gamma) & H^*(Y/\Gamma) \arrow[l, "(\varphi/\Gamma)^*"'] \\ & H^*(\mathbf{B}\Gamma) \arrow[ul, "((v \circ \varphi)/\Gamma)^*"] \arrow[u, "(v/\Gamma)^*"'] \end{tikzcd}$$

Figure 3.1.1. Functoriality of characteristic classes.

In other words the characteristic classes associated to a finite group action are *natural*, or, as one sometimes says, *functorial*.

We refer the reader to the wonderful book of tom Dieck, [**tD87**], for further details on equivariant maps and associated bundles. We also recommend the classical book of Milnor&Stasheff, [**MS74**], as an excellent source for the theory of characteristic classes of vector bundles. The generalities on bundles, including principal bundles, can be found in [**Ste51**].

3.1.2. $\mathbb{Z}_2$-spaces and the Definition of Stiefel-Whitney Classes

Let now X be a $\mathbb{Z}_2$-space, i.e., a CW complex equipped with a fixed point free involution. Specifying $\Gamma = \mathbb{Z}_2$ in the considerations above, we get a map $w : X \to S_a^\infty = \mathbf{E}\mathbb{Z}_2$. Furthermore, we have the induced quotient map

$$w/\mathbb{Z}_2 : X/\mathbb{Z}_2 \to S_a^\infty/\mathbb{Z}_2 = \mathbb{RP}^\infty = \mathbf{B}\mathbb{Z}_2.$$

In this particular case, the induced $\mathbb{Z}_2$-algebra homomorphism

$$(w/\mathbb{Z}_2)^* : H^*(\mathbb{RP}^\infty; \mathbb{Z}_2) \to H^*(X/\mathbb{Z}_2; \mathbb{Z}_2)$$

is determined by very little data. Namely, let z denote the nontrivial cohomology class in $H^1(\mathbb{RP}^\infty; \mathbb{Z}_2)$. Then $H^*(\mathbb{RP}^\infty; \mathbb{Z}_2) \simeq \mathbb{Z}_2[z]$ as a graded $\mathbb{Z}_2$-algebra, with z having degree 1. We denote the image $(w/\mathbb{Z}_2)^*(z) \in H^1(X/\mathbb{Z}_2; \mathbb{Z}_2)$ by $\varpi_1(X)$. Obviously, the whole map $(w/\mathbb{Z}_2)^*$ is determined by the element $\varpi_1(X)$.

This is the *Stiefel-Whitney class* of the $\mathbb{Z}_2$-space X. Clearly, $\varpi_1^k(X) = (w/\mathbb{Z}_2)^*(z^k)$. Furthermore, by the general observation, if Y is another $\mathbb{Z}_2$-space, and $\varphi : X \to Y$ is a $\mathbb{Z}_2$-map, then $(\varphi/\mathbb{Z}_2)^*(\varpi_1(Y)) = \varpi_1(X)$.

As an example we can quickly compute $\varpi_1(S_a^n)$, for an arbitrary nonnegative integer n. First, for dimensional reasons, $\varpi_1(S_a^0) = 0$. So we assume $n \geq 1$. Next, we have $S_a^n/\mathbb{Z}_2 = \mathbb{RP}^n$. The cohomology algebra $H^*(\mathbb{RP}^n; \mathbb{Z}_2)$ is generated by one element $\beta \in H^1(\mathbb{RP}^n; \mathbb{Z}_2)$, with a single relation $\beta^{n+1} = 0$. Finally, the standard inclusion map $\iota : S_a^n \hookrightarrow S_a^\infty$ is $\mathbb{Z}_2$-equivariant, and induces another standard inclusion $\iota/\mathbb{Z}_2 : \mathbb{RP}^n \hookrightarrow \mathbb{RP}^\infty$. Identifying $\mathbb{RP}^n$ with the image of $\iota/\mathbb{Z}_2$, we can think of it as the n-skeleton of $\mathbb{RP}^\infty$. Thus the induced $\mathbb{Z}_2$-algebra homomorphism $(\iota/\mathbb{Z}_2)^* : H^*(\mathbb{RP}^\infty; \mathbb{Z}_2) \to H^*(\mathbb{RP}^n; \mathbb{Z}_2)$ maps the canonical generator of $H^1(\mathbb{RP}^\infty; \mathbb{Z}_2)$ to β, and hence we can conclude that $\varpi_1(S_a^n) = \beta$.

3.2. Properties of Stiefel-Whitney Classes

3.2.1. Borsuk-Ulam Theorem, Index, and Coindex

The Stiefel-Whitney classes can be used to determine the nonexistence of certain $\mathbb{Z}_2$-maps. The following theorem is an example of such situation.

Theorem 3.2.1 (Borsuk-Ulam)**.**
Let n and m be nonnegative integers. If there exists a $\mathbb{Z}_2$-map $\varphi : S^n_a \to S^m_a$, then $n \leq m$.

Proof. Choose representations for the cohomology algebras $H^*(\mathbb{RP}^n; \mathbb{Z}_2) = \mathbb{Z}_2[\alpha]$, and $H^*(\mathbb{RP}^m; \mathbb{Z}_2) = \mathbb{Z}_2[\beta]$, with the only relations on the generators being $\alpha^{n+1} = 0$, and $\beta^{m+1} = 0$. Since the Stiefel-Whitney classes are functorial, we get $(\varphi/\mathbb{Z}_2)^*(\varpi_1(S^m_a)) = \varpi_1(S^n_a)$.

On the other hand, by the computation in the subsection 3.1.2, we have $\varpi_1(S^n_a) = \alpha$, and $\varpi_1(S^m_a) = \beta$. So $(\varphi/\mathbb{Z}_2)^*(\beta) = \alpha$, and hence $\alpha^{m+1} = (\varphi/\mathbb{Z}_2)^*(\beta)^{m+1} = (\varphi/\mathbb{Z}_2)^*(\beta^{m+1}) = 0$. Since $\alpha^{n+1} = 0$ is the only relation on α, this yields the desired inequality $m \geq n$. □

The Borsuk-Ulam Theorem makes the following terminology useful for formulating further obstructions to maps between $\mathbb{Z}_2$-spaces.

Definition 3.2.2. *Let X be a $\mathbb{Z}_2$-space.*
- *The* **index** *of X, denoted* $\operatorname{Ind} X$*, is the minimal integer n, for which there exists a $\mathbb{Z}_2$-map from X to S^n_a.*
- *The* **coindex** *of X, denoted* $\operatorname{Coind} X$*, is the maximal integer n, for which there exists a $\mathbb{Z}_2$-map from S^n_a to X.*

Assume that we have two $\mathbb{Z}_2$-spaces X and Y, and that $\gamma : X \to Y$ is a $\mathbb{Z}_2$-map, then, we have the inequality $\operatorname{Coind} X \leq \operatorname{Ind} Y$. Indeed, if there exists $\mathbb{Z}_2$-maps $\varphi : S^n_a \to X$, and $\psi : Y \to S^m_a$, then the composition

$$S^n_a \xrightarrow{\varphi} X \xrightarrow{\gamma} Y \xrightarrow{\psi} S^m_a$$

yields a $\mathbb{Z}_2$-map between two spheres with antipodal actions, hence, by the Borsuk-Ulam Theorem, we can conclude that $n \leq m$. In particular, taking $Y = X$, and $\varphi = \mathrm{id}$, we get the inequality $\operatorname{Coind} X \leq \operatorname{Ind} X$, for an arbitrary $\mathbb{Z}_2$-space.

3.2.2. Higher Connectivity and Stiefel-Whitney Classes

Many results giving topological obstructions to graph colorings had the k-connectivity of some space as the crucial assumption. We notice here an important connection between this condition and non-nullity of powers of Stiefel-Whitney classes.

First, it is trivial, that if X is a non-empty $\mathbb{Z}_2$-space, then one can equivariantly map S^0_a to X. It is possible to extend this construction inductively to an arbitrary $\mathbb{Z}_2$-space.

Proposition 3.2.3. *Let X and Y be two simplicial complexes with a free $\mathbb{Z}_2$-action, such that for some $k \geq 0$, we have $\dim X \leq k$, and Y is $(k-1)$-connected. Assume further that we have a $\mathbb{Z}_2$-map $\psi : X^{(d)} \to Y$, for some $d \geq -1$. Then, there exists a $\mathbb{Z}_2$-map $\varphi : X \to Y$, such that φ extends ψ.*

Please note the following convention used in the formulation of Proposition 3.2.3: $d = -1$ means we have no map ψ (in other words, $X^{-1} = \emptyset$), hence no additional conditions on the map φ.

Proof. Choose a $\mathbb{Z}_2$-invariant simplicial structure on X. We construct φ inductively on i-skeleton of X, for $i \geq d+1$. If $d = -1$, we start by defining φ on the 0-skeleton as follows: for each orbit $\{a, b\}$ consisting of two vertices of X, simply map a to an arbitrary point $y \in Y$, and then map b to $\gamma(y)$, where γ is the free involution of X.

Assume now that φ is defined on the $(i-1)$-skeleton of X, and extend the construction to the i-skeleton as follows. Let (σ, τ) be a pair of i-dimensional simplices of X, such that $\gamma\sigma = \tau$. The boundary $\partial\sigma$ is a $(i-1)$-dimensional sphere. By our assumptions $i - 1 \leq \dim X - 1 \leq k - 1$, hence the restriction of φ to $\partial\sigma$ extends to σ. Finally, we extend φ to the second simplex τ by applying the involution γ: $\varphi|_\tau := (\varphi|_\sigma) \circ \gamma$. □

Corollary 3.2.4. *Let X be a $\mathbb{Z}_2$-space, and assume X is $(k-1)$-connected, for some $k \geq 0$. Then there exists a $\mathbb{Z}_2$-map $\varphi : S^k_a \to X$. In particular, we have $\varpi^k_1(X) \neq 0$.*

Proof. S^k_a is k-dimensional, hence the statement follows immediately from Proposition 3.2.3. To see that $\varpi^k_1(X) \neq 0$, recall that, since the Stiefel-Whitney classes are functorial, we have $(\varphi/\mathbb{Z}_2)^*(\varpi^k_1(X)) = \varpi^k_1(S^k_a)$, and the latter has been verified to be nontrivial. □

The Corollary 3.2.4 explains the rule of thumb that, whenever dealing with $\mathbb{Z}_2$-spaces, the condition of k-connectivity can be replaced by the weaker condition that the $(k+1)$-th power of the appropriate Stiefel-Whitney class is different from 0.

3.2.3. Combinatorial Construction of Stiefel-Whitney Classes

Let us describe how the construction used in the proof of Proposition 3.2.3 can be employed to obtain an explicit combinatorial description of the Stiefel-Whitney classes.

Let X be a regular CW complex and a $\mathbb{Z}_2$-space, and denote the fixed point free involution on X by γ. As mentioned above, one can choose a simplicial structure on X, such that γ is a simplicial map. We define a $\mathbb{Z}_2$-map $\varphi : X \to S^\infty_a$ following the recipe above.

Take the standard $\mathbb{Z}_2$-equivariant cell decomposition of S^∞_a with two antipodal cells in each dimension. Divide $X^{(0)}$, the set of the vertices of X, into two disjoint sets $X^{(0)} = A \cup B$, such that every orbit of the $\mathbb{Z}_2$-action contains exactly one element from A and one element from B. Let $\{a, b\}$ be the 0-skeleton of S^∞_a, and map all the points in A to a, and all the points in B to b. Call the edges having one vertex in A, and one vertex in B, *multicolored*, and the edges connecting two vertices in A, resp. two vertices in B, A-internal, resp. B-internal.

Let $\{e_1, e_2\}$ be the 1-skeleton of S^∞_a. One can then extend φ to the 1-skeleton as follows. Map the A-internal edges to a, map the B-internal edges to b. Note that the multicolored edges form $\mathbb{Z}_2$-orbits, 2 edges in every orbit. For each such orbit, map one of the edges to e_1 (there is some arbitrary choice involved here), and map the other one to e_2.

Since the $\mathbb{Z}_2$-action on the space X is free, the generators of the cochain complex $C^*(X/\mathbb{Z}_2; \mathbb{Z}_2)$ can be indexed with the orbits of simplices. For an arbitrary simplex

δ we denote by τ_δ the generator corresponding to the orbit of δ; in particular, $\tau_{\gamma(\delta)} = \tau_\delta$.

The induced quotient cell decomposition of $\mathbb{RP}^\infty$ is the standard one, with one cell in each dimension. The cochain z^*, corresponding to the unique edge of $\mathbb{RP}^\infty$, is the generator (and the only nontrivial element) of $H^1(\mathbb{RP}^\infty; \mathbb{Z}_2)$. Its image under $(\varphi/\mathbb{Z}_2)^*$ is simply the sum of all orbits of the multicolored edges:

$$\varpi_1(X) = (\varphi/\mathbb{Z}_2)^*(z^*) = \sum_{\text{multicolored } e} \tau_e, \tag{3.2.1}$$

where the sum is taken over representatives of $\mathbb{Z}_2$-orbits of multicolored edges, one representative per orbit.

To describe the powers of the Stiefel-Whitney classes, $\varpi_1^k(X)$, we need to recall how the cohomology multiplication is done simplicially. In fact, to evaluate $\varpi_1^k(X)$ on a k-simplex $(v_0, v_1, \ldots, v_k)$, we need to evaluate $\varpi_1(X)$ on each of the edges (v_i, v_{i+1}), for $i = 0, \ldots, k-1$, and then multiply the results. Thus, the only k-simplices, on which the power $\varpi_1^k(X)$ evaluates nontrivially, are those whose ordered set of vertices has alternating elements from A and from B. We call these simplices *multicolored.* We summarize

$$\varpi_1^k(X) = \sum_{\text{multicolored } \sigma} \tau_\sigma, \tag{3.2.2}$$

where the sum is taken over representatives of $\mathbb{Z}_2$-orbits of multicolored k-dimensional simplices, one representative per orbit.

3.3. First Applications of Stiefel-Whitney Classes to Lower Bounds of Chromatic Numbers of Graphs

3.3.1. Complexes of Complete Multipartite Subgraphs

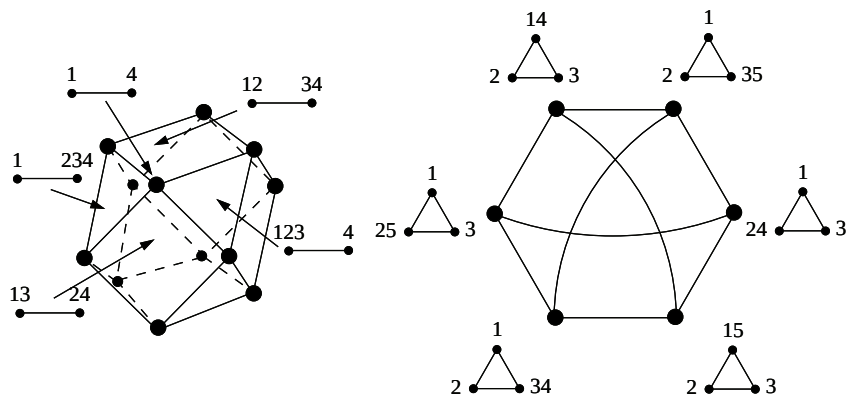
We have seen in the subsection 2.1.1 that the complex $\texttt{Hom}\,(K_2, G)$ is simply homotopy equivalent to the neighbourhood complex of G. In particular, $\texttt{Hom}\,(K_2, K_n)$ is simply homotopy equivalent to S^{n-2}. The following proposition provides us with a more complete information.

Proposition 3.3.1 ([**BK03b**, Proposition 4.3])**.**
(a) $\texttt{Hom}\,(K_2, K_{n+1})$ *is isomorphic as a cell complex to the boundary complex of the Minkowski sum* $\Delta^n + (-\Delta^n)$.
(b) *The* $\mathbb{Z}_2$*-action on* $\texttt{Hom}\,(K_2, K_{n+1})$*, induced by the flip action of* $\mathbb{Z}_2$ *on* K_2*, corresponds under this isomorphism to the central symmetry.*

Proposition 3.3.1 is illustrated with Figure 3.3.1. Perhaps the easiest way to Proposition 3.3.1 is by means of the following notion.

Definition 3.3.2. *Let* $X_1, \ldots, X_t$ *be a family of simplicial complexes with isomorphic sets of vertices. The* **deleted product** *of this family is the subcomplex of the direct product of* $X_1, \ldots, X_t$*, consisting of all cells* $\tau_1 \times \cdots \times \tau_t$*, satisfying* $\tau_i \cap \tau_j = \emptyset$*, for any* $i \neq j$.

Clearly, $\texttt{Hom}\,(K_m, K_n)$ can be viewed as a deleted product of m copies of $(n-1)$-dimensional simplices, see e.g., [**Ma03**]. In this context Proposition 3.3.1 is well-known, probably due to van Kampen.



$\mathtt{Hom}\,(K_2, K_4)$ Link of a vertex in $\mathtt{Hom}\,(K_3, K_5)$

Figure 3.3.1. Complexes of graph homomorphisms between complete graphs.

Also, for $m \geq 3$, $\mathtt{Hom}\,(K_m, G)$ can be thought of as complexes consisting of all complete m-partite subgraphs of G. Even in the case $G = K_n$, it seems complicated to understand $\mathtt{Hom}\,(K_m, G)$ up to homeomorphism. However, we still obtain a good description of the homotopy type.

Theorem 3.3.3 ([**BK03b**, Proposition 4.5])**.**
The cell complex $\mathtt{Hom}\,(K_m, K_n)$ *is homotopy equivalent to a wedge of* $(n-m)$*-dimensional spheres.*

Introducing a new piece of notation, let us say that the complex $\mathtt{Hom}\,(K_m, K_n)$ is homotopy equivalent to a wedge of $f(m,n)$ spheres. Let $S(-,-)$ denote the *Stirling numbers* of the second kind, and $SF_k(x) = \sum_{n \geq k} S(n,k)x^n$ denote the generating function (in the first variable) for these numbers. It is well-known, see for example [**Sta97**, p. 34], that

$$SF_k(x) = x^k/((1-x)(1-2x)\dots(1-kx)).$$

For $m \geq 1$, let $F_m(x) = \sum_{n \geq 1} f(m,n)x^n$ be the generating function (in the second variable) for the number of the spheres. Clearly, $F_1(x) = 0$, and $F_2(x) = x^2/(1-x)$.

Proposition 3.3.4 ([**BK03b**, Proposition 4.6])**.**
The numbers $f(m,n)$ *satisfy the following recurrence relation*

$$f(m,n) = mf(m-1,n-1) + (m-1)f(m,n-1), \tag{3.3.1}$$

for $n > m \geq 2$*; with the boundary values* $f(n,n) = n! - 1$*,* $f(1,n) = 0$ *for* $n \geq 1$*, and* $f(m,n) = 0$ *for* $m > n$*.*

Then, the generating function $F_m(x)$ *is given by the equation:*

$$F_m(x) = (m! \cdot x \cdot SF_{m-1}(x) - x^m)/(1+x). \tag{3.3.2}$$

As a consequence, the following non-recursive formulae are valid:

$$f(m,n) = (-1)^{m+n+1} + m!(-1)^n \sum_{k=m}^{n} (-1)^k S(k-1, m-1), \tag{3.3.3}$$

and

$$f(m,n) = \sum_{k=1}^{m-1} (-1)^{m+k+1} \binom{m}{k+1} k^n, \tag{3.3.4}$$

for $n \geq m \geq 1$.

In particular, for small values of m, we obtain the following explicit formulae: $f(2,n) = 1$, for $n \geq 2$, $f(3,n) = 2^n - 3$, for $n \geq 3$, $f(4,n) = 3^n - 4 \cdot 2^n + 6$, for $n \geq 4$, $f(5,n) = 4^n - 5 \cdot 3^n + 10 \cdot 2^n - 10$, for $n \geq 5$.

3.3.2. Stiefel-Whitney Classes and Test Graphs

One connection between the non-nullity of the powers of Stiefel-Whitney characteristic classes and the lower bounds for graph colorings is provided by the following general observation.

Theorem 3.3.5. *Let G be a graph without loops, and let T be a graph with $\mathbb{Z}_2$-action which flips some edge in T. If, for some integers $k \geq 0$, $m \geq 1$, $\varpi_1^k(\mathtt{Hom}\,(T,G)) \neq 0$, but $\varpi_1^k(\mathtt{Hom}\,(T,K_m)) = 0$, then $\chi(G) \geq m+1$.*

Since this statement is crucial for all our applications, we provide here a short argument.

Proof of Theorem 3.3.5. We know that, under the assumptions of the theorem, $\mathtt{Hom}\,(T,H)$ is a $\mathbb{Z}_2$-space for any loopfree graph H. Assume now that the graph G is m-colorable, i.e., there exists a homomorphism $\varphi : G \to K_m$. It induces a $\mathbb{Z}_2$-map $\varphi^T : \mathtt{Hom}\,(T,G) \to \mathtt{Hom}\,(T,K_m)$. Since the Stiefel-Whitney classes are functorial and $\varpi_1^k(\mathtt{Hom}\,(T,K_m)) = 0$, the existence of the map φ^T implies that $\varpi_1^k(\mathtt{Hom}\,(T,G)) = 0$, which is a contradiction to the assumption of the theorem. $\square$

We can now use Theorem 3.3.3 to give lower bounds for chromatic numbers of graphs in terms of Stiefel-Whitney classes of complexes of graph homomorphisms from complete graphs.

Theorem 3.3.6. *Let G be a graph, and let $n, k \in \mathbb{Z}$, such that $n \geq 2$, $k \geq -1$. If $\varpi_1^k(\mathtt{Hom}\,(K_n,G)) \neq 0$, then $\chi(G) \geq k+n$.*

Proof. Indeed, substituting $T = K_n$, and $m = k+n-1$, in the Theorem 3.3.5, all we need to do is to see that $\varpi_1^k(\mathtt{Hom}\,(K_n,K_{k+n-1})) = 0$. By Theorem 3.3.3, $\mathtt{Hom}\,(K_n,K_{k+n-1})$ is homotopy equivalent to a wedge of $(k-1)$-dimensional spheres. Hence, by dimensional reasons we conclude that $\varpi_1^k(\mathtt{Hom}\,(K_n,K_{k+n-1})) = 0$. $\square$

LECTURE 4
The Spectral Sequence Approach

4.1. $\mathtt{Hom}_+$-construction

4.1.1. Various Definitions

We shall now define a complex $\mathtt{Hom}_+(T,G)$ which is related to $\mathtt{Hom}\,(T,G)$. It is easier to compute various algebro-topological invariants for this complex, however, it also has less bearing on our main problem: computation of the lower bounds for chromatic numbers. We shall then connect the $\mathtt{Hom}$- and $\mathtt{Hom}_+$-constructions by means of a spectral sequence.

4.1.1.1. *A Subcomplex of a Total Join.*

Let T and G be arbitrary graphs. We shall define $\mathtt{Hom}_+(T,G)$ analogously to $\mathtt{Hom}\,(T,G)$, replacing the direct product with the join. We note here that whenever talking about $\mathtt{Hom}_+(T,G)$ we always assume that the graph T is finite.

Let, as before, $\Delta^{V(G)}$ be a simplex whose set of vertices is $V(G)$. Let $J(T,G)$ denote the join $*_{x\in V(T)}\Delta^{V(G)}$, i.e., the copies of $\Delta^{V(G)}$ are indexed by vertices of T. A cell (simplex) in $J(T,G)$ is a join of (possibly empty) simplices $*_{x\in V(T)}\sigma_x$, the dimension of this simplex is $\sum_{x\in V(T)}(\dim\sigma_x+1)-1$. Remark that this number is finite, since we assumed that T is finite. Here we use the usual convention that $\dim\emptyset=-1$.

Definition 4.1.1. *For arbitrary graphs T and G, $\mathtt{Hom}_+(T,G)$ is the simplicial subcomplex of $J(T,G)$ defined by the following condition: $\sigma=*_{x\in V(T)}\sigma_x\in\mathtt{Hom}_+(T,G)$ if and only if for any $x,y\in V(T)$, if $(x,y)\in E(T)$, and both σ_x and σ_y are nonempty, then (σ_x,σ_y) is a complete bipartite subgraph of G.*

The intuition behind this definition is that we relax the conditions of the $\mathtt{Hom}$ case by allowing some of the "coloring lists" to be empty. One can think of $\mathtt{Hom}_+(T,G)$ as a simplicial structure imposed on the set of all *partial graph homomorphisms* from T to G, i.e., graph homomorphisms from an induced subgraph of T to the graph G.

In analogy with the $\mathtt{Hom}$ case, we can describe the simplices of $\mathtt{Hom}_+(T,G)$ directly: they are indexed by all $\eta:V(T)\to 2^{V(G)}$ satisfying the same condition as in Definition 2.1.5. The closure of η is also defined identical to how it was defined

for $\mathtt{Hom}$. So the only difference is that $\eta(x)$ is allowed to be an empty set, for $x \in V(T)$.

4.1.1.2. *A Link of a Vertex in an Auxiliary* $\mathtt{Hom}$*-Complex.*

The following construction is the graph analog of the topological coning.

Definition 4.1.2. *For an arbitrary graph G, let G_+ be the graph obtained from G by adding an extra vertex a, called the apex vertex, and connecting it by edges to all the vertices of G_+ including a itself, i.e., $V(G_+) = V(G) \cup \{a\}$, and $E(G_+) = E(G) \cup \{(x,a),(a,x) \mid x \in V(G_+)\}$.*

We note that, for an arbitrary polyhedral complex K, such that all faces of K are direct products of simplices, and a vertex x of K, the *link* of x, $\mathrm{lk}_K(x)$, is a simplicial complex. It follows from the fact that a link of any vertex in a hypercube is a simplex, and the identity $\mathrm{lk}_{(A\times B)}(v,w) = \mathrm{lk}_A(v) * \mathrm{lk}_B(w)$, for arbitrary polyhedral complexes A and B.

We are now ready to formulate another definition, which is equivalent to Definition 4.1.1.

Definition 4.1.3. *For arbitrary graphs T and G, the simplicial complex $\mathtt{Hom}_+(T,G)$ is defined to be the link in $\mathtt{Hom}\,(T,G_+)$ of the specific graph homomorphism α, which maps all vertices of T to the apex vertex of G_+.*

In short: $\mathtt{Hom}_+(T,G) = \mathrm{lk}_{\mathtt{Hom}\,(T,G_+)}(\alpha)$. An example is shown on Figure 4.1.1.

The equivalence of the definitions follows essentially from the following bijection: let $\eta \in \mathcal{F}(\mathtt{Hom}\,(T,G_+))_{>\alpha}$, and set $\tilde{\eta}(v) := \eta(v) \setminus \{a\}$, for any $v \in V(T)$. Clearly, $\tilde{\eta} \in \mathcal{F}(\mathtt{Hom}_+(T,G))$, and it is easily checked that this bijection produces an isomorphism of simplicial complexes.

4.1.1.3. *Functorial Properties of the* $\mathtt{Hom}_+$*-Construction.*

Just like in the case of the $\mathtt{Hom}\,(-,-)$-construction, $\mathtt{Hom}_+(T,-)$ is a covariant functor from **Graphs** to **Top**. For two arbitrary graphs G and K, and a graph homomorphism φ from G to K, we have an induced simplicial map $\varphi^T : \mathtt{Hom}_+(T,G) \to \mathtt{Hom}_+(T,K)$.

Again, as in the case of $\mathtt{Hom}\,(-,-)$, the situation is somewhat more complicated with the functoriality in the first argument. Let T, G, and K, be three arbitrary graphs. This time, for a graph homomorphism ψ from T to G to induce a topological map from $\mathtt{Hom}_+(G,K)$ to $\mathtt{Hom}_+(T,K)$, we must require that ψ is surjective on the vertices. We can define the topological map ψ_K in the same way as for the $\mathtt{Hom}\,(-,-)$ case, but if ψ is not surjective on the vertices, then we may end up mapping a non-empty cell to an empty one. If, in addition, we want a simplicial map $\psi_K : \mathtt{Hom}_+(G,K) \to \mathtt{Hom}_+(T,K)$, then, as before in the subsection 2.4.3, we must require that ψ is injective, hence bijective on the vertices.

In particular, we still have that the group $\mathrm{Aut}\,(T) \times \mathrm{Aut}\,(G)$ acts on the complex $\mathtt{Hom}_+(T,G)$ simplicially. The difference is that we do not have the freeness as easily as we had in the $\mathtt{Hom}\,(-,-)$ case. For example, for an involution γ of T to induce a free action γ_G on $\mathtt{Hom}_+(T,G)$ we need to require that all orbits of γ on $V(T)$ are of cardinality 2, and that the vertices in the same orbit are connected by an edge. For instance, the action of $\mathbb{Z}_2$ on $\mathtt{Hom}_+(K_2,G)$ is free, whereas the reflection $\mathbb{Z}_2$-action on $\mathtt{Hom}_+(C_{2r+1},G)$ is not.

4.1.2. Connection to Independence Complexes

The following is a standard construction in topological combinatorics, see e.g., [**Ko99, Mes03**].

Definition 4.1.4. *For an arbitrary graph G, the* **independence complex** *of G, $\mathrm{Ind}\,(G)$, is the simplicial complex, whose set of vertices is $V(G)$, and simplices are all the independent sets (anticliques) of G.*

Before we can make use of the $\mathrm{Ind}\,(-)$-construction in our context, we need more graph terminology.

Definition 4.1.5. *For an arbitrary graph G, the* **strong complement** $\complement G$ *is defined by $V(\complement G) = V(G)$, and $E(\complement G) = V(G) \times V(G) \setminus E(G)$.*

For example, $\complement K_n$ is the disjoint union of n loops.

Definition 4.1.6. *For arbitrary graphs G and H, the* **direct product** $G \times H$ *is defined by: $V(G \times H) = V(G) \times V(H)$, and $E(G \times H) = \{((x,y),(x',y')) \mid (x,x') \in E(G), (y,y') \in E(H)\}$.*

For example, $K_2 \times K_2$ is a disjoint union of two copies of K_2, whereas $G \times \complement K_1$ is isomorphic to G for an arbitrary graph G.

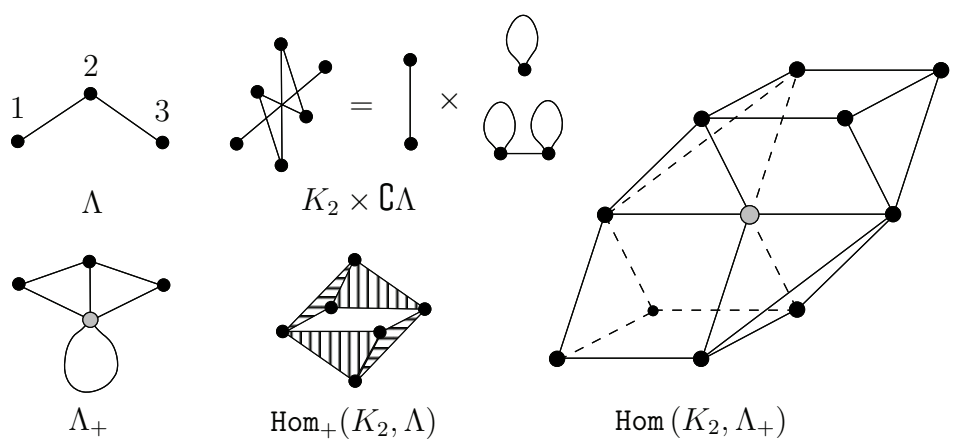


Figure 4.1.1. The $+$-construction.

Sometimes, it can be convenient to view $\mathtt{Hom}_+(G,H)$ as the independence complex of a certain graph.

Proposition 4.1.7 ([**BK04**, Proposition 3.2])**.**
For arbitrary graphs T and G, $\mathtt{Hom}_+(T,G)$ is isomorphic to $\mathrm{Ind}\,(T \times \complement G)$.

Specializing Proposition 4.1.7 to $G = K_n$, and taking into account $\complement K_n = \coprod_{i=1}^{n} \complement K_1$ (observed above), and the fact that for arbitrary graphs G_1 and G_2 we have

$$\mathrm{Ind}\,(G_1 \coprod G_2) = \mathrm{Ind}\,(G_1) * \mathrm{Ind}\,(G_2),$$

we obtain the following corollary.

Corollary 4.1.8. *For an arbitrary graph T, $\mathtt{Hom}_+(T, K_n)$ is isomorphic to the n-fold join* $\mathrm{Ind}\,(T)^{*n}$.

When G is loopfree, the dimension of the simplicial complex $\mathtt{Hom}_+(T, G)$ (unlike that of $\mathtt{Hom}\,(T, G)$) is easy to find, once the size of the maximal independent set of G is computed.

Proposition 4.1.9. *For an arbitrary graph T, and an arbitrary loopfree graph G, we have*

$$\dim(\mathtt{Hom}_+(T, G)) = |V(G)| \cdot (\dim(\mathrm{Ind}\,(T)) + 1) - 1.$$

Proof. Indeed, let $s = \dim(\mathrm{Ind}\,(T)) + 1$ be the size of the maximal independent set in T. Since G is loopfree, every vertex of G occurs in at most s of the sets $\eta(x)$, for $x \in V(T)$. On the other hand, we can choose an independent set $S \subseteq V(T)$, such that $|S| = s$, and then assign

$$\eta(x) = \begin{cases} V(G), & \text{for } x \in S; \\ \emptyset, & \text{otherwise.} \end{cases}$$

This gives a simplex of dimension $|V(G)| \cdot (\dim(\mathrm{Ind}\,(T)) + 1) - 1$. □

For example, $\dim(\mathtt{Hom}_+(C_{2r+1}, K_n)) = n \cdot ((r-1)+1) - 1 = nr - 1$.

4.1.3. The Support Map

For any topological space X and a set I, there is the standard support map from the join of I copies of X to the appropriate simplex

$$\mathrm{supp} : *_I X \longrightarrow \Delta^I,$$

which "forgets" the coordinates in X.

Specializing to our situation, for arbitrary graphs T and G, we get the restriction map $\mathrm{supp} : \mathtt{Hom}_+(T, G) \to \Delta^{V(T)}$. Explicitly, for each simplex of $\mathtt{Hom}_+(T, G)$, $\eta : V(T) \to 2^{V(G)}$, the support of η is given by $\mathrm{supp}\,\eta = V(T) \setminus \eta^{-1}(\emptyset)$. See Figure 4.1.2 for an example.

An important property of the support map is that the preimage of the barycenter of $\Delta^{V(T)}$ is homeomorphic to $\mathtt{Hom}\,(T, G)$. This is the crucial step in setting up a useful spectral sequence. The assumption that T is finite is crucial at this point, since an infinite simplex does not have a barycenter.

An alternative concise way to phrase the definition of supp is to consider the map $t^T : \mathtt{Hom}_+(T, G) \to \mathtt{Hom}_+(T, \complement K_1) \simeq \Delta^{V(T)}$ induced by the homomorphism $t : G \to \complement K_1$. Then, for each $\eta \in \mathtt{Hom}_+(T, G)$ we have $\mathrm{supp}\,\eta = t^T(\eta)$, where the simplices in $\Delta^{V(T)}$ are identified with the finite subsets of $V(T)$.

4.2. Spectral Sequence Generalities

Spectral sequences constitute an important tool of topological combinatorics in general. They have also been proved invaluable in the solution of the Lovász Conjecture. Taking into account the format of this article, we only give a short introduction here, aimed at setting up the notations and, hopefully, helping the intuition. We refer the interested reader to the excellent existing sources, see e.g., [**FFG86, McC01**].

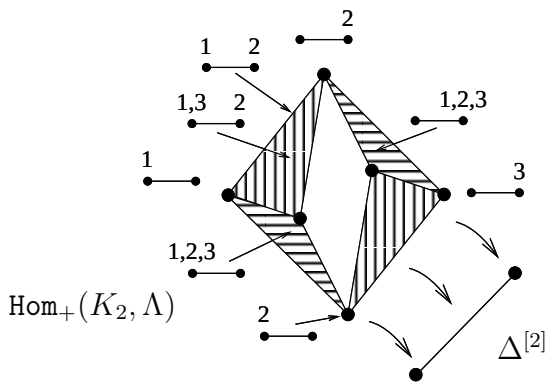


Figure 4.1.2. The support map from $\mathtt{Hom}_+(K_2, \Lambda)$ to $\Delta^{[2]}$.

4.2.1. Cochain Complexes and their Cohomology

Recall that a *cochain complex* is a sequence

$$\mathcal{C} = \cdots \xrightarrow{d^{i-2}} C^{i-1} \xrightarrow{d^{i-1}} C^i \xrightarrow{d^i} C^{i+1} \xrightarrow{d^{i+1}} \dots,$$

where C^i's are $\mathcal{R}$-modules, d^i's are $\mathcal{R}$-module homomorphisms (called differentials) satisfying $d^{i+1} \circ d^i = 0$, and $\mathcal{R}$ is a commutative ring with a unit. We shall use the notation $\mathcal{C} = (C^*, d^*)$. For all our purposes it is enough in the continuation to restrict one's attention to the cases $\mathcal{R} = \mathbb{Z}$, and $\mathcal{R} = \mathbb{Z}_2$.

We can associate a cochain complex $C^*(X)$ to a cell complex X in the standard way: $C^i(X)$ is taken to be a free $\mathcal{R}$-module with the generators indexed by the i-dimensional cells (the module of $\mathcal{R}$-valued functionals on cells of X), and the differential maps are given by the corresponding coboundary maps. Sometimes this particular cochain complex is called a *cellular cochain complex*.

Given two cochain complexes $\mathcal{C}_1 = (C_1^*, d_1^*)$ and $\mathcal{C}_2 = (C_2^*, d_2^*)$, a *cochain complex homomorphism* $\varphi : \mathcal{C}_1 \to \mathcal{C}_2$ (also called a *cochain complex map*) is a collection of $\mathcal{R}$-module homomorphisms $\varphi_i : C_1^i \to C_2^i$, for all integers i, such that the following diagram commutes

$$\begin{array}{ccc} C_1^i & \xrightarrow{d_1^i} & C_1^{i+1} \\ {\scriptstyle \varphi_i}\downarrow & & \downarrow{\scriptstyle \varphi_{i+1}} \\ C_2^i & \xrightarrow{d_2^i} & C_2^{i+1} \end{array} \tag{4.2.1}$$

For each choice of $\mathcal{R}$, cochain complexes together with cochain complex homomorphisms form a category.

Associated with a cochain complex, one has the *cohomology groups*

$$H^i(\mathcal{C}) = \ker d^i / \operatorname{im} d^{i-1}.$$

In our cases $H^i(\mathcal{C})$ is either an abelian group or a vector space over $\mathbb{Z}_2$.

Given two cochain complexes $\mathcal{C}_1 = (C_1^*, d_1^*)$ and $\mathcal{C}_2 = (C_2^*, d_2^*)$, and a cochain complex homomorphism $\varphi : \mathcal{C}_1 \to \mathcal{C}_2$, since φ_i's commute with the corresponding differentials, φ induces a map on the cohomology groups $\varphi_i^* : H^i(\mathcal{C}_1) \to H^i(\mathcal{C}_2)$.

The above facts mix well with the cellular structure. First, for a cell complex X, the cellular cohomology groups of X are by definition isomorphic to the cohomology groups of the associated cochain complex $C^*(X)$. Second, for two cell complexes X and Y, a cellular map $\varphi : X \to Y$ induces a cochain complex map between associated cochain complexes (but in the opposite direction!), and hence a map between corresponding cohomology groups.

4.2.2. Filtrations

In concrete situations it can be difficult to compute the cohomology groups $H^i(\mathcal{C})$ without auxiliary constructions. The idea behind spectral sequences is to break up this large task into smaller subtasks, with the formal machinery to help the bookkeeping. This "break up" is usually phrased in terms of a filtration.

A *cochain subcomplex* of $\mathcal{C}$ is a sequence

$$\widetilde{\mathcal{C}} = \cdots \xrightarrow{d^{i-2}} \widetilde{C}^{i-1} \xrightarrow{d^{i-1}} \widetilde{C}^{i} \xrightarrow{d^{i}} \widetilde{C}^{i+1} \xrightarrow{d^{i+1}} \dots,$$

where $\widetilde{C}^i$ is an $\mathcal{R}$-submodule of C^i, and the differentials are restrictions of those in $\mathcal{C}$. We shall simply write $\mathcal{C} \supseteq \widetilde{\mathcal{C}}$. In this situation, one can form the quotient cochain complex

$$\mathcal{C}/\widetilde{\mathcal{C}} = \cdots \xrightarrow{d^{i-2}} C^{i-1}/\widetilde{C}^{i-1} \xrightarrow{d^{i-1}} C^{i}/\widetilde{C}^{i} \xrightarrow{d^{i}} C^{i+1}/\widetilde{C}^{i+1} \xrightarrow{d^{i+1}} \dots.$$

The cohomology groups of this complex, $H^*(\mathcal{C}/\widetilde{\mathcal{C}})$, are usually denoted $H^*(\mathcal{C}, \widetilde{\mathcal{C}})$, and are called the *relative cohomology groups.*

If X is a cell complex, and Y its cell subcomplex, then the cellular cochain complex of Y is a cochain subcomplex of the cellular cochain complex of X. The corresponding cohomology groups of the quotient cochain complex are precisely the relative cohomology groups of the pair of topological spaces (X, Y).

Definition 4.2.1. *A (finite)* **filtration** *on a cochain complex $\mathcal{C}$ is a nested sequence of cochain complexes*

$$\mathcal{C}_j = \cdots \xrightarrow{d^{i-2}} C_j^{i-1} \xrightarrow{d^{i-1}} C_j^{i} \xrightarrow{d^{i}} C_j^{i+1} \xrightarrow{d^{i+1}} \dots,$$

for $j = 0, 1, 2, \dots, t$, such that $\mathcal{C} = \mathcal{C}_t \supseteq \mathcal{C}_{t-1} \supseteq \cdots \supseteq \mathcal{C}_0$ (that is why we suppressed the lower index in the differential).

In general, infinite filtrations can be considered, but in this article we limit our considerations to the finite ones. Given a filtration $\mathcal{C} = \mathcal{C}_t \supseteq \mathcal{C}_{t-1} \supseteq \cdots \supseteq \mathcal{C}_0$, we set $\mathcal{C}_{-1} = 0$, for the convenience of notations.

There are many standard filtrations of cochain complexes. For example, if a pure cochain complex is bounded, say $C^i = 0$, for $i < 0$, or $i > t$, then, the standard skeleton filtration is defined as follows:

$$C_j^i = \begin{cases} C^i, & \text{if } i \le j; \\ 0, & \text{otherwise.} \end{cases}$$

This filtration is not very interesting though, since computing the cohomology groups with its help is canonically equivalent with computing the cohomology groups from the cochain complex directly.

For a cell complex X, a classical way to define a filtration on its cellular cochain complex, is to choose a cell filtration on X, i.e., a sequence of cell subcomplexes $X = X_t \supseteq X_{t-1} \supseteq \cdots \supseteq X_0$ (again for the convenience of notations, we set

$X_{-1} = \emptyset$). As mentioned above, the corresponding cellular cochain complexes form a sequence of nested subcomplexes.

If the cell complex X is finite dimensional, then, taking X_i to be the i-th skeleton of X, we recover the standard skeleton filtration on $C^*(X)$, which explains the name of this filtration.

A much more interesting situation is the following.

Definition 4.2.2. *Assume that we have a cell map $\varphi : X \to Y$ and a filtration $Y = Y_t \supseteq Y_{t-1} \supseteq \cdots \supseteq Y_0$. Define a filtration on X as follows: $X_i := \varphi^{-1}(Y_i)$, for $i = 0, \ldots, t$. This filtration on X is called the* **pullback** *of the filtration on Y* **along φ.**

In the case when the filtration on Y is simply the skeleton filtration, the corresponding pullback filtration on X is called the *Serre filtration*. We use the same name for the corresponding filtration on the cellular cochain complex of X.

4.2.3. Spectral Sequence Terminology

Once we have fixed a filtration $\mathcal{C} = \mathcal{C}_t \supseteq \mathcal{C}_{t-1} \supseteq \cdots \supseteq \mathcal{C}_0$, $\mathcal{C}_i = (C_i^*, d^*)$, on a cochain complex, we can proceed to compute its cohomology groups by studying auxiliary algebraic gadgets derived from the filtration.

Rather than studying the 1-dimensional cochain complex directly, we study a sequence of 2-dimensional tableaux $E_n^{*,*}$, $n = 0, 1, 2, \ldots$. Our cochain complex had the usual differential, going one up in degree, which one can express symbolically by writing $d^1 : C^* \to C^{*+1}$. Instead, each tableau $E_n^{*,*}$ is equipped with a differential going almost diagonally, $d_n^{*,*} : E_n^{*,*} \to E_n^{*+n,*+1-n}$. One expresses this fact by saying that d_n is *a differential of bidegree* $(n, -n+1)$.

Each differential d_n is in a way derived from the original differential d, and furthermore, $E_{n+1}^{*,*}$ is the cohomology tableau of $E_n^{*,*}$ in the appropriate sense. The idea is then to compute the tableaux $E_n^{*,*}$ one by one, until they stabilize. The stabilized tableau is usually called $E_\infty^{*,*}$.

We would like to alert the reader at this point that even after the tableau $E_\infty^{*,*}$ was computed, it can still require additional work to determine the cohomology groups of the original cochain complex. Surely, if $\mathcal{R}$ is a field, the situation is easy. Namely one has

$$H^d(\mathcal{C}) = \bigoplus_{p+q=d} E_\infty^{p,q}.$$

However, if $\mathcal{R}$ is an arbitrary ring (for example $\mathcal{R} = \mathbb{Z}$), then one may need to solve a number of extension problems before obtaining the final answer. This has to do with the fact, that in a short exact sequence of $\mathcal{R}$-modules

$$0 \longrightarrow A \xrightarrow{\alpha} B \xrightarrow{\beta} C \longrightarrow 0,$$

B does not necessarily split as a direct sum of the submodule A and the quotient module C. This is not even true for $\mathcal{R} = \mathbb{Z}$, a classical example is to take $A = B = \mathbb{Z}$, $C = \mathbb{Z}_2$, to take $\alpha : x \mapsto 2x$ to be the doubling map (injective), and to take $\beta : x \mapsto x \mod 2$ to be the parity map (surjective).

Let us now describe more precisely how the tableaux $E_n^{*,*}$ and the differentials d_n are constructed. As auxiliary modules, set

$$Z_n^{p,q} := C_p^{p+q} \cap d^{-1}(C_{p+n}^{p+q+1}),$$

where d^{-1} denotes the inverse of the differential d, i.e., $Z_n^{p,q}$ consists of all elements of C_p^{p+q} whose boundary is in C_{p+n}^{p+q+1}; and set

$$B_n^{p,q} := C_p^{p+q} \cap d\,(C_{p-n}^{p+q-1}),$$

i.e., $B_n^{p,q}$ consists of all elements of C_p^{p+q} which constitute the image of d from C_{p-n}^{p+q-1}. These are the settings for $n \geq 0$. Finally, for $n = -1$, we use the following convention:

$$Z_{-1}^{p,q} := C_p^{p+q}, \qquad \text{and} \qquad B_{-1}^{p,q} := d(C_{p+1}^{p+q-1}).$$

With these notations, we set

$$E_n^{p,q} := Z_n^{p,q}/(Z_{n-1}^{p+1,q-1} + B_{n-1}^{p,q}), \tag{4.2.2}$$

for all $0 \leq n \leq \infty$.

It is an easy check, which we leave to the reader, that $d(Z_n^{p,q}) \subseteq Z_n^{p+n,q-n+1}$, and that $d(Z_{n-1}^{p+1,q-1} + B_{n-1}^{p,q}) \subseteq Z_{n-1}^{p+n+1,q-n} + B_{n-1}^{p+n,q-n+1}$. It follows that, via the quotient maps, the differential d induces a map from $E_n^{p,q}$ to $E_n^{p+n,q-n+1}$, which we choose to call $d_n^{p,q}$ (or just d_n, if it is clear what the coefficients p and q are).

One can view the tableau $(E_n^{*,*}, d_n)$ as a collection of (nearly) diagonal cochain complexes. This allows one to compute the cohomology groups, just like for the usual cochain complexes, by setting

$$H^{p,q}(E_n^{*,*}, d_n) = \ker(E_n^{p,q} \xrightarrow{d_n} E_n^{p+n,q-n+1})/\mathrm{im}\,(E_n^{p-n,q+n-1} \xrightarrow{d_n} E_n^{p,q}).$$

Now we can make the sense in which $E_{n+1}^{*,*}$ is the cohomology tableau of $E_n^{*,*}$ precise:

$$E_{n+1}^{p,q} = H^{p,q}(E_n^{*,*}, d_n). \tag{4.2.3}$$

Please note, that the equation (4.2.3) is not trivial, and needs a proof. It can be deduced directly from the equation (4.2.2), see e.g., [**McC01**].

Let us start with unwinding these definitions for $n = 0$. It follows from our conventions for $n = -1$, that

$$E_0^{p,q} = (C_p^{p+q} \cap d^{-1}(C_p^{p+q+1}))/(C_{p+1}^{p+q} + d(C_{p+1}^{p+q-1})) = C_p^{p+q}/C_{p+1}^{p+q}. \tag{4.2.4}$$

Furthermore, the differential $d : C_p^{p+q} \to C_p^{p+q+1}$ induces the differential $d_0 : E_0^{p,q} \to E_0^{p,q+1}$, which is nothing else but the differential of the relative cochain complex $(\mathcal{C}_p, \mathcal{C}_{p+1})$. By the equation (4.2.3), this yields

$$E_1^{p,q} = H^{p+q}(\mathcal{C}_p, \mathcal{C}_{p+1}). \tag{4.2.5}$$

Moreover, one can show that $d_1^{p,q} : E_1^{p,q} \longrightarrow E_1^{p+1,q}$ is the connecting homomorphism $\partial : H^{p+q}(\mathcal{C}_p, \mathcal{C}_{p+1}) \longrightarrow H^{p+q+1}(\mathcal{C}_{p+1}, \mathcal{C}_{p+2})$ in the long exact sequence of the triple $(\mathcal{C}_p, \mathcal{C}_{p+1}, \mathcal{C}_{p+2})$.

Unless some additional specific information is available, it is hard to say what happens in the tableaux for $n \geq 2$. The important thing is that with the setup above, the spectral sequence runs its course and eventually converges (modulo the extension difficulties outlined above) to the cohomology groups of the original cochain complex.

4.3. The Standard Spectral Sequence Converging to $H^*(\mathtt{Hom}_+(T,G))$

4.3.1. Filtration Induced by the Support Map

Let T and G be two graphs, and assume T is finite. As mentioned above, there is a simplicial map $\operatorname{supp} : \mathtt{Hom}_+(T,G) \to \Delta^{V(T)}$. Consider the Serre filtration of the cellular cochain complex $C^*(\mathtt{Hom}_+(T,G);\mathcal{R})$ associated with this map.

We order the vertices of T and of G, and then observe that the vertices of $\mathtt{Hom}_+(T,G)$ are indexed with pairs (x,y), where $x \in V(T)$, $y \in V(G)$, such that if x is looped, then so is y. Let us internally order these pairs lexicographically: $(x_1,y_1) \prec (x_2,y_2)$ if either $x_1 < x_2$, or $x_1 = x_2$ and $y_1 < y_2$. Orient each simplex of $\mathtt{Hom}_+(T,G)$ according to this order on the vertices. We call this orientation *standard*, and call the oriented simplex η_+. One can think of this simplex as a chain in the corresponding chain complex; we denote the dual cochain with η_+^*.

We can explicitly describe the considered filtration. Define the subcomplexes $F^p = F^pC^*(\mathtt{Hom}_+(T,G);\mathcal{R})$ of $C^*(\mathtt{Hom}_+(T,G);\mathcal{R})$ as follows:

$$F^p : \cdots \xrightarrow{\partial^{q-1}} F^{p,q} \xrightarrow{\partial^q} F^{p,q+1} \xrightarrow{\partial^{q+1}} \cdots,$$

where

$$F^{p,q} = F^pC^q(\mathtt{Hom}_+(T,G);\mathcal{R}) = \mathcal{R}[\eta_+^* \mid \eta_+ \in \mathtt{Hom}_+^{(q)}(T,G), |\operatorname{supp}\eta| \geq p+1],$$

∂^* is the restriction of the differential in $C^*(\mathtt{Hom}_+(T,G);\mathcal{R})$, and $\mathtt{Hom}_+^{(q)}(T,G)$ denotes the q-th skeleton of $\mathtt{Hom}_+(T,G)$. Phrased verbally: $F^{p,q}$ is generated by all elementary cochains corresponding to q-dimensional cells, which are supported in at least $p+1$ vertices of T. Note, that this restriction defines a filtration, since the differential does not decrease the cardinality of the support set.

We have

$$C^q(\mathtt{Hom}_+(T,G);\mathcal{R}) = F^{0,q} \supseteq F^{1,q} \supseteq \cdots \supseteq F^{|V(T)|-1,q} \supseteq F^{|V(T)|,q} = 0,$$

which is the Serre filtration associated to the support map.

4.3.2. The 0th and the 1st Tableaux

By writing the brackets $[-]$ after the name of a cochain complex, we shall mean the index shifting (to the left), that is for the cochain complex $\mathcal{C} = (C^*, d^*)$, the cochain complex $\mathcal{C}[s] = (C^*[s], d^*)$ is defined by $C^i[s] := C^{i+s}$; note that we choose not to change the sign of the differential.

Proposition 4.3.1 ([**BK04**, Proposition 3.4])**.**
For any p,

$$F^p/F^{p+1} = \bigoplus_{\substack{S\subseteq V(T)\\ |S|=p+1}} C^*(\mathtt{Hom}\,(T[S],G);\mathcal{R})[-p]. \tag{4.3.1}$$

Hence, the 0th tableau of the spectral sequence associated to the cochain complex filtration F^* is given by

$$E_0^{p,q} = C^{p+q}(F^p,F^{p+1}) = \bigoplus_{\substack{S\subseteq V(T)\\ |S|=p+1}} C^q(\mathtt{Hom}\,(T[S],G);\mathcal{R}). \tag{4.3.2}$$

Furthermore, using the equation (4.2.5), we obtain the description of the first tableau as well.

$$E_1^{p,q} = H^{p+q}(F^p, F^{p+1}) = \bigoplus_{\substack{S \subseteq V(T) \\ |S|=p+1}} H^q(\mathtt{Hom}\,(T[S], G); \mathcal{R}). \tag{4.3.3}$$

4.3.3. The First Differential

Set now $\mathcal{R} = \mathbb{Z}_2$. According to the formula (4.3.3) the first differential $d_1^{p,q} : E_1^{p,q} \longrightarrow E_1^{p+1,q}$ can be viewed as a map

$$\bigoplus_{\substack{S \subseteq V(T) \\ |S|=p+1}} H^q(\mathtt{Hom}\,(T[S], G); \mathbb{Z}_2) \longrightarrow \bigoplus_{\substack{S \subseteq V(T) \\ |S|=p+2}} H^q(\mathtt{Hom}\,(T[S], G); \mathbb{Z}_2).$$

It is possible to describe this map explicitly.

For $S_2 \subseteq S_1 \subseteq V(T)$, let $i[S_1, S_2] : T[S_2] \hookrightarrow T[S_1]$ be the inclusion graph homomorphism. Since $\mathtt{Hom}\,(-, G)$ is a contravariant functor, we have an induced map $i_G[S_1, S_2] : \mathtt{Hom}\,(T[S_1], G) \to \mathtt{Hom}\,(T[S_2], G)$, and hence, an induced map on the cohomology groups

$$i_G^*[S_1, S_2] : H^*(\mathtt{Hom}\,(T[S_2], G); \mathcal{R}) \to H^*(\mathtt{Hom}\,(T[S_1], G); \mathcal{R}).$$

Let $\sigma \in H^q(\mathtt{Hom}\,(T[S], G); \mathbb{Z}_2)$, for some q, and some $S \subseteq V(G)$. The value of the first differential on σ is given by

$$d_1^{p,q}(\sigma) = \sum_{x \in V(T) \setminus S} i_G^*[S \cup \{x\}, S](\sigma). \tag{4.3.4}$$

In the case of integer coefficients, $\mathcal{R} = \mathbb{Z}$, one needs more work to derive the formula for $d_1^{p,q}(\sigma)$ analogous to (4.3.4), since, additionally, the signs have to be taken into consideration.

LECTURE 5
The Proof of the Lovász Conjecture

5.1. Formulation of the Conjecture and Sketch of the Proof

5.1.1. Formulation and Motivation of the Lovász Conjecture

As mentioned in the Section 2.3, the Lovász Theorem 2.3.2, and the fact that the neighbourhood complex $\mathcal{N}(G)$ is simply homotopy equivalent to $\mathrm{Bip}\,(G) = \mathtt{Hom}\,(K_2, G)$, are suggesting that $\mathtt{Hom}$-complexes in general would provide the right context of formulating and proving further topological obstructions to graph colorings.

Up to now, the most important extension of the original Lovász Theorem has been the one, where the edge K_2 is replaced with an odd cycle C_{2r+1}.

Theorem 5.1.1 (**Lovász Conjecture**).
Let G be a graph, such that the complex $\mathtt{Hom}\,(C_{2r+1}, G)$ is k-connected for some $r, k \in \mathbb{Z}$, $r \geq 1$, $k \geq -1$, then $\chi(G) \geq k + 4$.

Lovász Conjecture has been proved in [**BK03b**], for this reason we have stated it here directly as a theorem.

Remark 5.1.2. *It follows from Theorem 2.5.2, that, once the Lovász Conjecture has been proved, the statement will remain true if C_{2r+1} is replaced by any graph T, such that T can be reduced to C_{2r+1} by a sequence of folds.*

We formulate here a strengthening of the original conjecture.

Conjecture 5.1.3. *Let G be a graph, such that $\varpi_1^k(\mathtt{Hom}\,(C_{2r+1}, G)) \neq 0$, for some $r, k \in \mathbb{Z}$, $r \geq 1$, $k \geq -1$, then $\chi(G) \geq k + 3$.*

5.1.2. The Winding Number and the Proof of the Case $k = 0$

The case $k = 0$ of the Lovász Conjecture can be settled with relatively little machinery, that is why we choose to include for it a separate argument. This is the "toy version" which illustrates our general methods, and we prove the more general Conjecture 5.1.3.

To any continuous map $\varphi : S^1 \to S^1$ one can associate an integer $\mathrm{wind}\,(\varphi)$, called the *winding number* of φ. Intuitively, the absolute value of $\mathrm{wind}\,(\varphi)$ measures, as its name suggests, the number of times φ wraps the source circle around

the target circle, whereas the sign of $\operatorname{wind}(\varphi)$ registers whether the orientation has been changed or not. The usual way to define $\operatorname{wind}(\varphi)$ formally is to notice that φ induces a group homomorphism $\varphi^* : H^1(S^1;\mathbb{Z}) \to H^1(S^1;\mathbb{Z})$. Any group homomorphism from $\mathbb{Z}$ to itself is uniquely determined by the image of 1. This image is exactly the winding number.

As the proof of Theorem 3.3.5 suggests, we need to analyze the complexes $\mathtt{Hom}(C_{2r+1}, K_3)$ in some detail. One can see, by direct inspection, that the connected components of $\mathtt{Hom}(C_{2r+1}, K_3)$ can be indexed by the winding numbers α. All one needs to see is that if two homomorphisms $\varphi, \psi : C_{2r+1} \to K_3$ have the same winding number, then there is a sequence of edges in $\mathtt{Hom}(C_{2r+1}, K_3)$ connecting φ with ψ; and this is fairly straightforward.

We notice however, that these winding numbers cannot be arbitrary. Indeed, if the number of times C_{2r+1} winds around K_3 is α, then $2r+1 = 3\alpha + 2t$, for some nonnegative integer $t \le r$. Hence, $\alpha = (2r - 2t + 1)/3$. It follows, that α must be odd, and that it cannot exceed $(2r+1)/3$. So $\alpha = \pm 1, \pm 3, \dots, \pm(2s+1)$, where $s = \lfloor (r-1)/3 \rfloor$, in particular $s \ge 0$.

Let $\varphi : \mathtt{Hom}(C_{2r+1}, K_3) \to \{\pm 1, \pm 3, \dots, \pm(2s+1)\}$ map each point $x \in \mathtt{Hom}(C_{2r+1}, K_3)$ to the point on the real line, indexing the connected component of x. Clearly, φ is a $\mathbb{Z}_2$-map. Since $\dim(\{\pm 1, \pm 3, \dots, \pm(2s+1)\}/\mathbb{Z}_2) = 0$, we have $H^1(\{\pm 1, \pm 3, \dots, \pm(2s+1)\}/\mathbb{Z}_2; \mathbb{Z}_2) = 0$, and the functoriality of the Stiefel-Whitney classes implies $\varpi_1(\mathtt{Hom}(C_{2r+1}, K_3)) = 0$. The Conjecture 5.1.3 for this case follows now from Theorem 3.3.5.

We have shown the Conjecture 5.1.3 for $k = 0$ using the Stiefel-Whitney classes, but it is equally easy to prove the Lovász Conjecture for this case directly. Indeed, following the lines of the proof of Theorem 3.3.5, we see that, a 3-coloring of G would induce a $\mathbb{Z}_2$-map from $\mathtt{Hom}(C_{2r+1}, G)$ to $\mathtt{Hom}(C_{2r+1}, K_3)$. On the other hand, the first one of these spaces is connected, by the conjecture assumption, whereas the second one is not, and has no connected components preserved by the $\mathbb{Z}_2$-action. Clearly, this yields a contradiction.

5.1.3. Sketch of the Proof of the Lovász Conjecture

Our proof of Lovász Conjecture is based on two fundamental properties of the complexes $\mathtt{Hom}(C_{2r+1}, K_n)$. The first one is the following.

Theorem 5.1.4 ([**BK04**, Theorem 2.3(b)]).
We have $\varpi_1^{n-2}(\mathtt{Hom}(C_{2r+1}, K_n)) = 0$, for all $r \ge 1$, and odd n, such that $n \ge 3$.

If $r' > r$, then there is a $\mathbb{Z}_2$-equivariant graph homomorphism $\varphi : C_{2r'+1} \to C_{2r+1}$, in turn inducing a $\mathbb{Z}_2$-map $\varphi_{K_n} : \mathtt{Hom}(C_{2r+1}, K_n) \to \mathtt{Hom}(C_{2r'+1}, K_n)$. It follows, that if Theorem 5.1.4 is true for r', then it is also true for r. Therefore, if necessary, we can assume that r is taken to be sufficiently large.

Some further details of the proof of Theorem 5.1.4 are given in the Section 5.3.

To formulate the second property, consider one of the two embeddings $\iota : K_2 \hookrightarrow C_{2r+1}$ which maps the edge to the $\mathbb{Z}_2$-invariant edge of C_{2r+1}. Clearly, ι is a $\mathbb{Z}_2$-equivariant graph homomorphism. Since $\mathtt{Hom}(-, H)$ is a contravariant functor, ι induces a map of $\mathbb{Z}_2$-spaces $\iota_{K_n} : \mathtt{Hom}(C_{2r+1}, K_n) \to \mathtt{Hom}(K_2, K_n)$, which in turn induces a $\mathbb{Z}$-algebra homomorphism $\iota_{K_n}^* : H^*(\mathtt{Hom}(K_2, K_n); \mathbb{Z}) \to H^*(\mathtt{Hom}(C_{2r+1}, K_n); \mathbb{Z})$.

Theorem 5.1.5 ([**BK04**, Theorem 2.6]).
Assume n is even, then $2 \cdot \iota^*_{K_n}$ is a 0-map.

Some further details of the proof of Theorem 5.1.5 are given in the Section 5.2.

Sketch of the Proof of Theorem 5.1.1 (Lovász Conjecture).
The case $k = -1$ is trivial, so take $k \geq 0$. Assume first that k is even. By Corollary 3.2.4, we have $\varpi_1^{k+1}(\mathtt{Hom}\,(C_{2r+1}, G)) \neq 0$. By Theorem 5.1.4, we have $\varpi_1^{k+1}(\mathtt{Hom}\,(C_{2r+1}, K_{k+3})) = 0$. Hence, applying Theorem 3.3.5 for $T = C_{2r+1}$ we get $\chi(G) \geq k + 4$.

Assume now that k is odd, and that $\chi(G) \leq k + 3$. Let $\varphi : G \to K_{k+3}$ be a vertex-coloring map. Combining Corollary 3.2.4, the fact that $\mathtt{Hom}\,(C_{2r+1}, -)$ is a covariant functor from loopfree graphs to $\mathbb{Z}_2$-spaces, and the map $\iota : K_2 \hookrightarrow C_{2r+1}$, we get the following diagram of $\mathbb{Z}_2$-spaces and $\mathbb{Z}_2$-maps:

$$S_a^{k+1} \xrightarrow{f} \mathtt{Hom}\,(C_{2r+1}, G) \xrightarrow{\varphi^{C_{2r+1}}} \mathtt{Hom}\,(C_{2r+1}, K_{k+3}) \xrightarrow{\iota_{K_{k+3}}} \xrightarrow{\iota_{K_{k+3}}} \mathtt{Hom}\,(K_2, K_{k+3}) \cong S_a^{k+1}.$$

This gives a homomorphism on the corresponding cohomology groups in dimension $k+1$, $h^* = f^* \circ (\varphi^{C_{2r+1}})^* \circ (\iota_{K_{k+3}})^* : \mathbb{Z} \to \mathbb{Z}$. By Theorem 5.1.5 we have $2 \cdot (i_{K_{k+3}})^* = 0$, hence $2h^* = 0$, and therefore $h^* = 0$. It is well-known, see, e.g., [**Hat02**, Proposition 2B.6, p. 174], that a $\mathbb{Z}_2$-map $S_a^n \to S_a^n$ cannot induce a 0-map on the nth cohomology groups (in fact it must be of odd degree). Hence, we have a contradiction, and so $\chi(G) \geq k + 4$. □

As the reader may have already noticed, we are actually proving a sharper statement than the original Lovász Conjecture. First of all, the condition "$\mathtt{Hom}\,(C_{2r+1}, G)$ is k-connected" can be replaced by a weaker condition "the coindex of $\mathtt{Hom}\,(C_{2r+1}, G)$ is at least $k + 1$". Furthermore, for even k, that condition can be weakened even further to "$\varpi_1^{k+1}(\mathtt{Hom}\,(C_{2r+1}, G)) \neq 0$", i.e., the stronger Conjecture 5.1.3 is proved.

5.2. Completing the Sketch for the Case k is Odd

5.2.1. The First Spectral Sequence and the Independence Complexes of Cycles

The main technical tool is to consider the spectral sequence associated to the Serre filtration induced by the support map supp : $\mathtt{Hom}_+(C_{2r+1}, K_n) \to \Delta^{[2r+1]}$. As we already mentioned in the subsection 4.2.3, the spectral sequence converges to the cohomology groups of $\mathtt{Hom}_+(C_{2r+1}, K_n)$. As it happens, the complex $\mathtt{Hom}_+(C_{2r+1}, K_n)$ is much easier to understand than the complex $\mathtt{Hom}\,(C_{2r+1}, K_n)$.

To start with, for $n = 1$, we are simply dealing with the independence complexes of graphs: $\mathtt{Hom}_+(G, K_1) = \mathrm{Ind}\,(G)$. Fortunately, that complex has already been well-understood for cycles. Some examples are shown on Figure 5.2.1.

Proposition 5.2.1 ([**Ko99**, Proposition 5.2]).
For any $t \geq 2$, we have

$$\mathrm{Ind}\,(C_t) \simeq \begin{cases} S^{k-1} \vee S^{k-1}, & \text{if } t = 3k; \\ S^{k-1}, & \text{if } t = 3k \pm 1. \end{cases}$$

Here the degenerate case $t = 2$ makes sense, if we let C_2 be a graph with two vertices, connected by an edge (or a double edge).

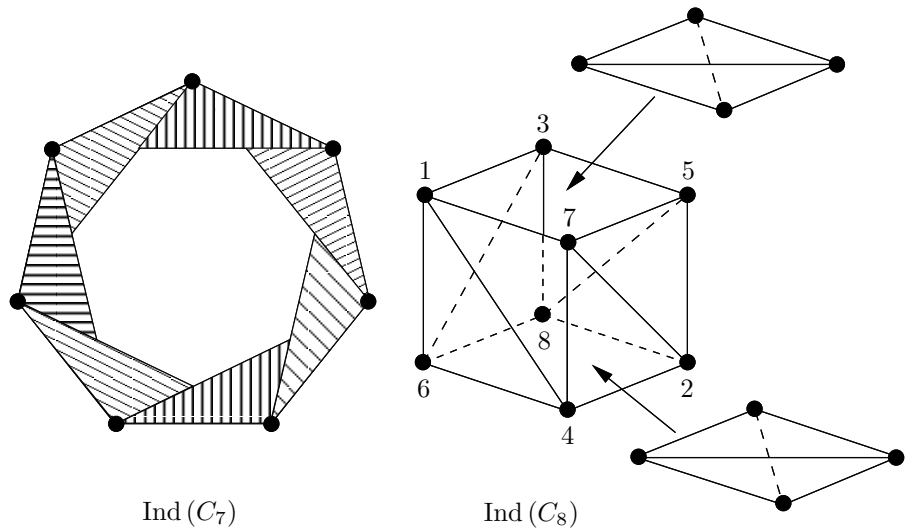


Figure 5.2.1. Examples of independence complexes of cycles. We remark that in the right picture, the 8 triangles on the sides of the cube are filled, while the top and the bottom of the cube are filled with solid tetrahedra.

Now, by Corollary 4.1.8, we have $\mathtt{Hom}_+(C_t, K_n) \simeq \mathrm{Ind}\,(C_t)^{*n}$, hence we derive an explicit description.

Corollary 5.2.2 ([**BK04**, Corollary 4.2])**.**
For any $t \geq 2$, we have

$$\mathtt{Hom}_+(C_t, K_n) \simeq \begin{cases} \bigvee_{2^n \text{ copies}} S^{nk-1}, & \text{if } t = 3k; \\ S^{nk-1}, & \text{if } t = 3k \pm 1. \end{cases}$$

This is a very convenient situation for us, since we know that the spectral sequence converges to something with a single nonzero entry.

5.2.2. The Analysis of the First Tableau

Next, we look at what the first tableau of this spectral sequence is. The general formula (4.3.3) says that the only possibly nonzero entries will be in the columns numbered $0, 1, \ldots, 2r$. Furthermore, the entries in column number p are nothing but the direct sum of the cohomology groups of induced subgraphs with $p + 1$ vertices.

For $p = 2r$ this simply means that this column consists of the cohomology groups of the desired space $\mathtt{Hom}\,(C_{2r+1}, K_n)$ itself. For $p = 0, \ldots, 2r - 1$ we get the cohomology groups of proper induced subgraphs. Fortunately, the proper induced subgraphs of a cycle are very simple: they are disjoint unions of isolated vertices and of strings.

We recall now the formula (2.4.1), which, in this particular case, says that the summands for the entries in the first tableau come from the direct products of $\mathtt{Hom}\,(K_1, K_n)$ and of $\mathtt{Hom}\,(L_m, K_n)$. The first one of these complexes is contractible,

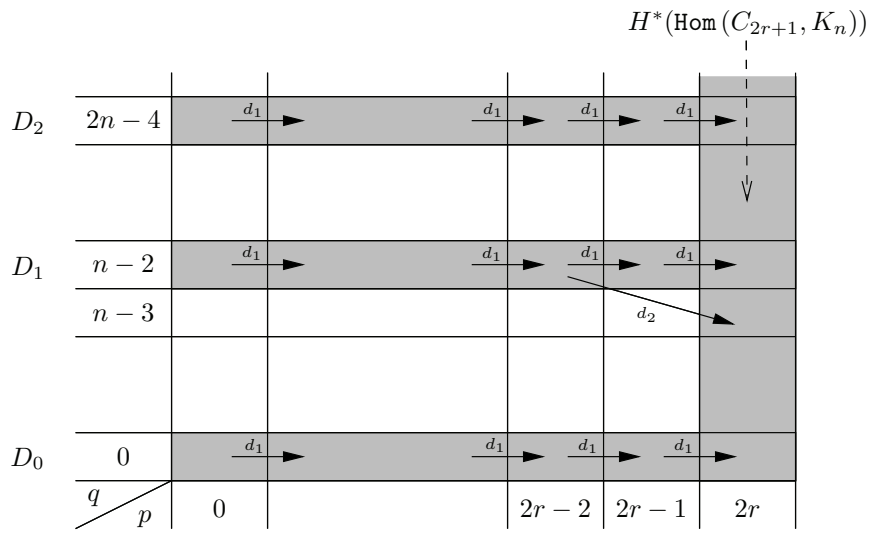


Figure 5.2.2. The $E_1^{*,*}$-tableau, for $E_1^{p,q} \Rightarrow H^{p+q}(\mathtt{Hom}_+(C_{2r+1}, K_n); \mathbb{Z})$.

whereas, L_m can be folded to an edge, hence, by Corollary 2.5.5, $\mathtt{Hom}\,(L_m, K_n)$ is homotopy equivalent to S^{n-2}.

Since the direct products of $(n-2)$-dimensional spheres may only have nontrivial cohomology groups in dimensions which are multiples of $n-2$, we can conclude that the only possibly nontrivial entries of $E_1^{*,*}$ are in rows indexed $t(n-2)$, and in the last column. See Figure 5.2.2 for the schematic summary of these findings; on this figure, the shaded area covers all possibly nontrivial entries of the first tableau.

5.2.3. The Analysis of the Second Tableau

Next, we need to understand what happens in every row once we pass to the second tableau. It is probably possible to perform a complete computation. However, this is a rather tedious task, which is unnecessary if we only care about what happens to the entries $(2r, n-2)$ and $(2r, n-3)$. Instead, we satisfy ourselves with deriving some partial information about $E_2^{*,*}$.

The idea is to introduce some combinatorial encoding for the generators of the entries of the first tableau, and then understand the values of d_1. The first task is not difficult, since the generators of the cohomology groups of direct products of spheres of the same dimension, can be labeled by the subsets of the set of the spheres.

In our case, the spheres correspond to "arcs" on the cycle, and so we can label the generators with induced subgraphs of C_{2r+1}, with certain set of arcs being marked. One can then filter these entries and employ another spectral sequence to compute the cohomology groups with respect to the differential d_1. We refer the reader to [**BK04**, Lemma 4.8] for details.

The main outcome is that the possibly nontrivial entries in $E_2^{*,*}$ will be indexed by such collections of arcs, that the gaps between neighboring arcs will not exceed 2. Clearly, if we are dealing with the row $t(n-2)$, our induced subgraph of C_{2r+1} cannot have fewer than $(2r+1) - 2t$ vertices, since otherwise one of the gaps

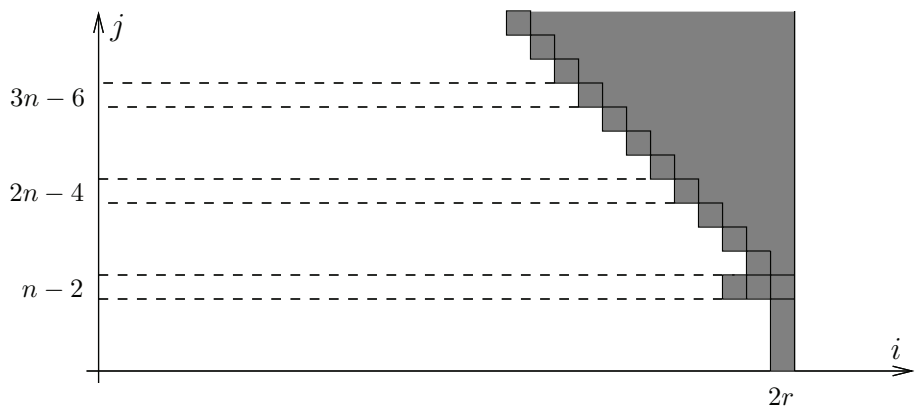


Figure 5.2.3. The possibly nonzero entries in $E_2^{*,*}$-tableau, for $E_2^{p,q} \Rightarrow H^{p+q}(\mathtt{Hom}_+(C_{2r+1}, K_n); \mathbb{Z})$.

would be too large. Almost always this ensures that the entries of $E_2^{*,*}$ outside of the shaded area on Figure 5.2.3 are equal to 0. There are two exceptional cases: $(n,t) = (5,2)$, and $n = 4$. These cases can then be computed "by hand", using rather specific observations, see [**BK04**, Subsections 4.6 and 4.7].

5.2.4. The Conclusion

After a detailed analysis of the entries $E_2^{2r-2,n-2}$ and $E_2^{2r-1,n-2}$ we derive partial information about the cohomology groups (with integer, as well as with $\mathbb{Z}_2$ coefficients) of $\mathtt{Hom}\,(C_{2r+1}, K_n)$, which is summarized in the Table 5.2.1.

(n,r)	R	H^{n-2}	H^{n-3}
$2 \nmid n, n \geq 5, (n,r) \neq (5,3)$	$\mathbb{Z}$	$\mathbb{Z}$	$\mathbb{Z}$
$(n,r) = (5,3)$	$\mathbb{Z}$	$\mathbb{Z}^2$	$\mathbb{Z}$
$2 \mid n, n \geq 6$, or $n = 4, r \leq 3$	$\mathbb{Z}$	$\mathbb{Z}_2$	0
$n = 4, r \geq 4$	$\mathbb{Z}$	$\mathbb{Z} \oplus \mathbb{Z}_2$	0
$n \geq 5, (n,r) \neq (5,3)$, or $n = 4, r \leq 3$	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}_2$
$(n,r) = (5,3)$, or $n = 4, r \geq 4$	$\mathbb{Z}_2$	$\mathbb{Z}_2^2$	$\mathbb{Z}_2$

Table 5.2.1.

We remark here that the results presented in the Table 5.2.1 have been somewhat strengthened recently.

Theorem 5.2.3 ([**CK04b**, Corollary 4.6])**.**
For arbitrary integers $r, n \geq 3$, the complex $\mathtt{Hom}\,(C_r, K_n)$ is $(n-4)$-connected.

Let us now return to Theorem 5.1.5. From the Table 5.2.1, we see that, in most of the cases, $2 \cdot \iota_{K_n}^*$ is a 0-map for a prosaic reason: the target group $H^*(\mathtt{Hom}\,(C_{2r+1}, K_n); \mathbb{Z})$ is isomorphic to $\mathbb{Z}_2$. The only exception is the case $n = 4$,

$r \geq 4$. The validity of the statement of Theorem 5.1.5 in this special case can be verified by the direct analysis of the map $d_1 : E_1^{2r-1,2} \to E_1^{2r,2}$, see [**BK04**, Subsection 4.8] for details.

5.3. Completing the Sketch for the Case k is Even

5.3.1. Topology of the Quotient Space $\mathtt{Hom}_+(C_{2r+1}, K_n)/\mathbb{Z}_2$

To analyze this case we need to extend some of the results of the Section 5.2. As a general guideline for this subsection, we would like to understand the action of $\mathbb{Z}_2$ on $\mathtt{Hom}\,(C_{2r+1}, K_n)$, $\mathtt{Hom}_+(C_{2r+1}, K_n)$, and on their respective cohomology groups somewhat better.

To start with, consider the $\mathbb{Z}_2$-action on $\mathtt{Hom}_+(C_{2r+1}, K_n)$. Fortunately, despite of the fact, that this action is not free, it turns out to be possible to describe the quotient space rather explicitly.

Proposition 5.3.1 ([**BK04**, Proposition 4.4])**.**
For any $r \geq 1$, we have

$$\mathtt{Hom}_+(C_{2r+1}, K_n)/\mathbb{Z}_2 \simeq \begin{cases} \bigvee_{2^{n-1}\ copies} S^{nk-1}, & \text{if } 2r+1 = 3k; \\ S^{kn/2-1} * \mathbb{RP}^{kn/2-1}, & \text{if } 2r+1 = 3k \pm 1. \end{cases}$$

Simple dimension inequalities yield the following corollary.

Corollary 5.3.2 ([**BK04**, Corollary 4.5])**.**
$\widetilde{H}^i(\mathtt{Hom}_+(C_{2r+1}, K_n)/\mathbb{Z}_2) = 0$ for $r \geq 2$, $n \geq 5$, and $i \leq n + r - 2$. Except for the case $r = 3$.

Furthermore, again by the detailed analysis of the differentials in our spectral sequence, but this time, with the $\mathbb{Z}_2$-action in mind, one can prove the following statement.

Proposition 5.3.3 ([**BK04**, Corollary 4.15])**.**
Let n be odd, $n \geq 3$, $r \geq 2$, and assume $(n, r) \neq (5, 3)$. Then, $\mathbb{Z}_2$ acts trivially on $H^{n-2}(\mathtt{Hom}\,(C_{2r+1}, K_n); \mathbb{Z})$, and, it acts as a multiplication by -1, on $H^{n-3}(\mathtt{Hom}\,(C_{2r+1}, K_n); \mathbb{Z}) \simeq \mathbb{Z}$.

We notice at this point that the support map $\mathrm{supp} : \mathtt{Hom}_+(C_{2r+1}, K_n) \to \Delta^{[2r+1]}$ is $\mathbb{Z}_2$-equivariant and hence it induces the quotient map $\mathrm{supp}\,/\mathbb{Z}_2 : \mathtt{Hom}_+(C_{2r+1}, K_n)/\mathbb{Z}_2 \to \Delta^{[2r+1]}/\mathbb{Z}_2$. In order to get simplicial structure on $\Delta^{[2r+1]}/\mathbb{Z}_2$, we subdivide $\Delta^{[2r+1]}$ in a minimal way, so that every simplex preserved by $\mathbb{Z}_2$-action is fixed by this action pointwise.

One can think of this new subdivision as the one obtained by representing simplex $\Delta^{[2r+1]}$ as a topological join of one point and r intervals: $\{c\} * [a_1, b_1] * \cdots * [a_r, b_r]$, inserting an extra vertex c_i into the middle of each of the $[a_i, b_i]$, and then taking the join of $\{c\}$ and the subdivided intervals. We denote the obtained abstract simplicial complex by $\tilde{\Delta}^{[2r+1]}$.

The $\mathbb{Z}_2$-quotient of this simplicial structure gives one on $\Delta^{[2r+1]}/\mathbb{Z}_2$, and we can consider the Serre filtration on $\mathtt{Hom}_+(C_{2r+1}, K_n)/\mathbb{Z}_2$ associated with the map $\mathrm{supp}\,/\mathbb{Z}_2$.

5.3.2. The second spectral sequence

Consider now the spectral sequence associated to this filtration, with the coefficients in $\mathbb{Z}_2$ instead of $\mathbb{Z}$. As before, this time by Proposition 5.3.3, we know precisely what this spectral sequence converges to. The formulae (4.3.1), (4.3.2), and (4.3.3) can be generalized as well, but before we do that we need some additional terminology.

First, we denote the set of the vertices which were added in the subdivision by $\mathcal{C} = \{c, c_1, \dots, c_r\}$. Further, for an arbitrary simplex $\tilde{\sigma} \in \tilde{\Delta}^{[2r+1]}$, we define its support simplex $\vartheta(\tilde{\sigma}) \in \Delta^{[2r+1]}$ by replacing every c_i in $\tilde{\sigma}$ by $\{a_i, b_i\}$, i.e.,

$$\vartheta(\tilde{\sigma}) = (\tilde{\sigma} \setminus \{c_1, \dots, c_r\}) \cup \bigcup_{c_i \in \tilde{\sigma}} \{a_i, b_i\}.$$

We can now state the analog of the formula (4.3.3) for the spectral sequence of the quotient. The analogs of formulae (4.3.1), and (4.3.2), are straightforward, and are omitted for the sake of space, see [**BK04**, Section 6] for further details.

$$\begin{aligned} E_1^{p,q} = &\bigoplus_{\sigma} H^{q-p}(\mathtt{Hom}\,(C_{2r+1}[\vartheta(\sigma)], K_n)/\mathbb{Z}_2; \mathbb{Z}_2) \\ &\bigoplus_{\tau} H^{q-p}(\mathtt{Hom}\,(C_{2r+1}[\vartheta(\tau)], K_n); \mathbb{Z}_2), \end{aligned} \tag{5.3.1}$$

where the first sum is taken over all $\sigma \subseteq \mathcal{C}$, such that $|\sigma| = p+1$, and the second sum is taken over all $\mathbb{Z}_2$-orbits $\langle\tau\rangle$, such that $\tau \subseteq V(\tilde{\Delta}^{[2r+1]})$, $|\tau| = p+1$, and $\tau \setminus \mathcal{C} \neq \emptyset$.

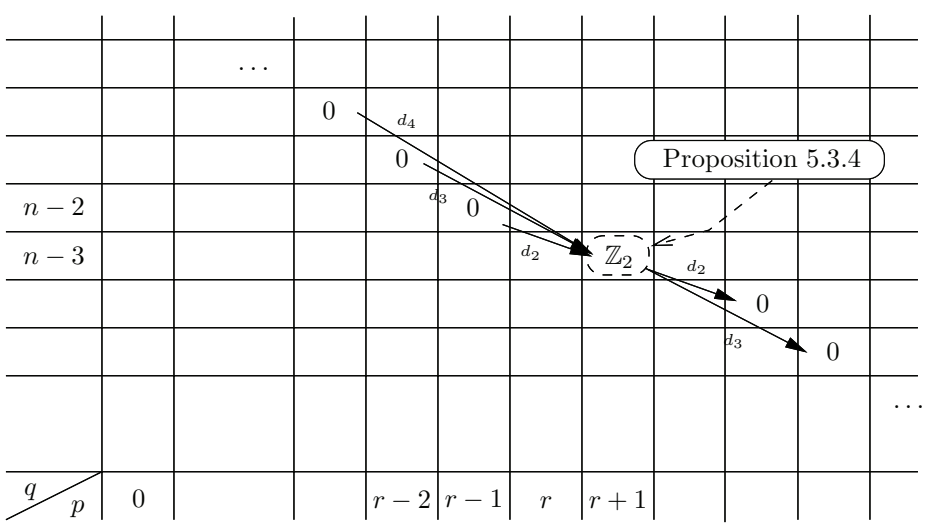


Figure 5.3.1. The $E_2^{*,*}$-tableau, $E_2^{p,q} \Rightarrow H^{p+q}(\mathtt{Hom}_+(C_{2r+1}, K_n)/\mathbb{Z}_2; \mathbb{Z}_2)$.

The next important piece of structure is understanding cohomology map in dimension $n-3$, which is induced by the quotient map $q : \mathtt{Hom}\,(C_{2r+1}, K_n) \to \mathtt{Hom}\,(C_{2r+1}, K_n)/\mathbb{Z}_2$.

Proposition 5.3.4 ([**BK04**, Proposition 6.2])**.**
Let n be odd, $n \geq 3$, $r \geq 2$, and assume $(n, r) \neq (5, 3)$. Then,

$$q^{n-3} : H^{n-3}(\mathtt{Hom}\,(C_{2r+1}, K_n)/\mathbb{Z}_2; \mathbb{Z}_2) \to H^{n-3}(\mathtt{Hom}\,(C_{2r+1}, K_n); \mathbb{Z}_2), \tag{5.3.2}$$

is a 0-map.

The crucial ingredient of the proof is provided by Proposition 5.3.3, see [**BK04**] for a complete argument.

The proof in [**BK04**] proceeds by deriving some partial information about the $E_2^{*,*}$-tableau of the spectral sequence under the consideration. The analysis is somewhat technical and we omit the details. Figure 5.3.1 depicts the values of the entries which are of interest to us.

Let us make two important remarks. First, to derive the value $E_2^{r+1,n-3} = \mathbb{Z}_2$, one needs the result of Proposition 5.3.4, which here ensures that the differential $d_1 : E_2^{r,n-3} \to E_2^{r+1,n-3}$ is a 0-map. Second, the value $E_2^{r-1,n-2} = 0$ is derived under the assumption that $\varpi_1^{n-2}(\mathtt{Hom}\,(C_{2r+1}, K_n)) \neq 0$, which we are trying to disprove.

Finally, we may conclude from Figure 5.3.1, that $E_\infty^{r+1,n-3} = \mathbb{Z}_2$. This contradicts Corollary 5.3.2, proving our original assumption $\varpi_1^{n-2}(\mathtt{Hom}\,(C_{2r+1}, K_n)) \neq 0$ to be wrong.

LECTURE 6
Summary and Outlook

6.1. Homotopy Tests, $\mathbb{Z}_2$-Tests, and Families of Test Graphs

6.1.1. Homotopy Test Graphs

Returning to our ideology of test graphs, it appears natural to give the following definition.

Definition 6.1.1. *A graph T is called a* **homotopy test graph***, if, for an arbitrary graph G, the following equation is satisfied*

$$\chi(G) > \chi(T) + \operatorname{conn} \mathtt{Hom}\,(T, G). \tag{6.1.1}$$

Using the terminology of Definition 6.1.1, Theorems 3.3.6 and 5.1.1 can be interpreted as saying that the complete graphs and the odd cycles are homotopy test graphs. Furthermore, it follows from Theorem 2.5.2(1) that the class of homotopy test graphs is closed under the equivalence relation given by the folds and by their reverses.

More generally, it has been asked by Lovász, [**Lov**], whether every graph is a homotopy test graph. That has been answered in the negative by Hoory & Linial, [**HL04**], whose example HL is presented on Figure 6.1.1. Note that $\chi(HL) = 5$, and set $G = K_5$. It was shown in [**HL04**] that $\mathtt{Hom}\,(HL, K_5)$ is connected, hence the equation (6.1.1) is false for these values of G and T.

The problem of characterizing the homotopy test graphs is a formidable one, with many open questions left to explore, see for e.g., Conjecture 6.2.1.

6.1.2. Stiefel-Whitney Test Graphs

Switching from all spaces to $\mathbb{Z}_2$-spaces, and from homotopy to cohomology, we define a different class of test graphs. First, recall the following standard notion of algebraic topology.

Definition 6.1.2. *Let X be a $\mathbb{Z}_2$-space. The* **height** *of X, denoted $\mathtt{h}(X)$, is the maximal nonnegative integer h, such that $\varpi_1^h(X) \neq 0$.*

It is important to note, that if X and Y are two arbitrary $\mathbb{Z}_2$-spaces, and $\varphi : X \to Y$ is an arbitrary $\mathbb{Z}_2$-map, then, since the Stiefel-Whitney characteristic classes are functorial, we have $(\varphi/\mathbb{Z}_2)^*(\varpi_1(Y)) = \varpi_1(X)$, which in particular implies the inequality $\mathtt{h}(X) \leq \mathtt{h}(Y)$.

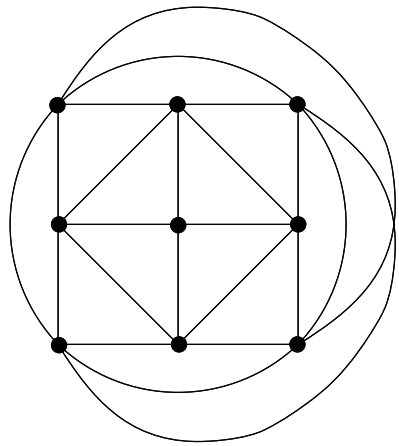


Figure 6.1.1. The Hoory-Linial example of a graph, which is not a homotopy test graph.

We note that, for an arbitrary $\mathbb{Z}_2$-space X, the existence of a $\mathbb{Z}_2$-equivariant map $S^n_a \to X$ implies $n = \mathtt{h}(S^n_a) \leq \mathtt{h}(X)$, whereas the existence of a $\mathbb{Z}_2$-equivariant map $X \to S^m_a$ implies $m = \mathtt{h}(S^m_a) \geq \mathtt{h}(X)$. This can be best summarized with the inequality

$$\operatorname{Coind}(X) \leq \mathtt{h}(X) \leq \operatorname{Ind}(X).$$

Let us now return to graphs.

Definition 6.1.3. *Let T be a graph with a $\mathbb{Z}_2$-action which flips an edge. Then, T is called* **Stiefel-Whitney n-test graph**, *if we have $\mathtt{h}(\mathtt{Hom}(T, K_n)) = n - \chi(T)$. Furthermore, T is called* **Stiefel-Whitney test graph** *if it is Stiefel-Whitney n-test graph for any integer $n \geq \chi(T)$.*

A direct application of Theorem 3.3.5 yields the next corollary, which also serves as an explanation for our terminology.

Corollary 6.1.4. *Assume T is a Stiefel-Whitney test graph, then, for an arbitrary graph G, we have*

$$\chi(G) \geq \chi(T) + \mathtt{h}(\mathtt{Hom}(T, G)). \tag{6.1.2}$$

Note, that by Corollary 3.2.4, we have $\mathtt{h}(X) \geq \operatorname{conn} X + 1$, for an arbitrary $\mathbb{Z}_2$-space X. Therefore, comparing equations (6.1.1) and (6.1.2), we see that if a graph T is a Stiefel-Whitney test graph, then, it is also a homotopy test graph.

Let us stress again that, in analogy to the fact that the *height* is defined for $\mathbb{Z}_2$-spaces, the term *Stiefel-Whitney test graph* actually refers to a pair (T, γ), where T is a graph, and γ is an involution of T, which flips an edge. The following question arises naturally in this context.

Question. *Does there exist a graph T having two different involutions, γ_1 and γ_2, such that (T, γ_1) is a Stiefel-Whitney test graph, whereas (T, γ_2) is not?*

It would be rather surprising, if the answer to this question turned out to be positive.

Next, we describe an important extension property of the class of Stiefel-Whitney test graphs.

Proposition 6.1.5. *Let T be an arbitrary graph, and let A and B be Stiefel-Whitney test graphs, such that $\chi(T) = \chi(A) = \chi(B)$. Assume further that there exist $\mathbb{Z}_2$-equivariant graph homomorphisms $\varphi : A \to T$ and $\psi : T \to B$. Then, T is also a Stiefel-Whitney test graph.*

Proof. Let n be an arbitrary positive integer. By the functoriality of Stiefel-Whitney characteristic classes, we have

$$\mathtt{h}(\mathtt{Hom}\,(A, K_n)) \leq \mathtt{h}(\mathtt{Hom}\,(T, K_n)) \leq \mathtt{h}(\mathtt{Hom}\,(B, K_n)).$$

Hence $n - \chi(A) \leq \mathtt{h}(\mathtt{Hom}\,(T, K_n)) \leq n - \chi(B)$, which, by the assumptions of the proposition, implies $\mathtt{h}(\mathtt{Hom}\,(T, K_n)) = n - \chi(T)$. □

The next corollary describes a simple, but instructive example of the situation in Proposition 6.1.5.

Corollary 6.1.6. *Any connected bipartite graph T with a $\mathbb{Z}_2$-action which flips an edge is a Stiefel-Whitney test graph.*

Indeed, we have $\mathbb{Z}_2$-equivariant graph homomorphisms $K_2 \hookrightarrow T \to K_2$, where the first one is the inclusion of the flipped edge, and the second one is the arbitrary coloring map, see Figure 6.1.2. Since, by Proposition 3.3.1, K_2 is a Stiefel-Whitney test graph, we conclude that T is also a Stiefel-Whitney test graph.

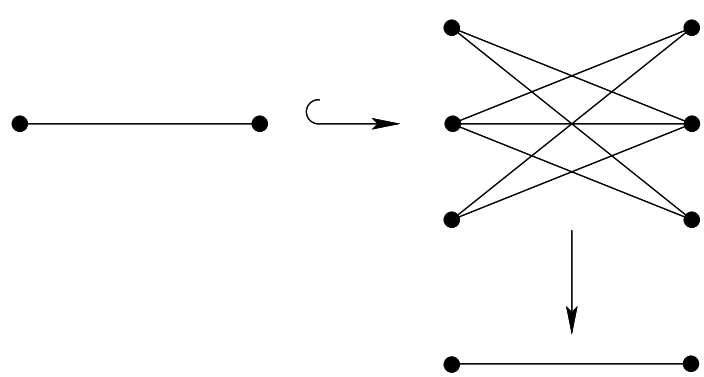


Figure 6.1.2. $\mathbb{Z}_2$-invariant factoring of an edge through a bipartite graph.

In particular, any even cycle with the $\mathbb{Z}_2$-action which flips an edge is a Stiefel-Whitney test graph.

Summary. The class of Stiefel-Whitney test graphs contains complete graphs, connected bipartite graphs (in both cases one can take any involution which flips an edge). Furthermore, it is closed under factorizations, as described in Proposition 6.1.5.

By Theorem 5.1.4, the odd cycles are Stiefel-Whitney n-test graphs, for odd $n \geq 3$. Conjecturally, see Conjecture 6.2.5, odd cycles are Stiefel-Whitney n-test graphs, for all $n \geq 3$.

6.2. Conclusion and Open Problems

It follows from Corollary 6.1.6 that any connected bipartite graph with a $\mathbb{Z}_2$-action which flips an edge is a homotopy test graph. It seems natural to generalize this statement.

Conjecture 6.2.1. *Every connected bipartite graph is a homotopy test graph.*

By what is said above, we know that a connected bipartite graph is a homotopy test graph if there is a sequence of folds and their reverses, reducing it to some connected bipartite graph with a $\mathbb{Z}_2$-action which flips an edge.

Before formulating the next conjecture, we recall that by saying that a topological space is (-1)-connected, we mean that it is nonempty. Clearly, if the maximal valency of G is at most $n-1$, then G can be colored with n colors by means of the greedy procedure. Furthermore, Babson & Kozlov proved in [**BK03b**, Proposition 2.4] that if the maximal valency of G is at most $n-2$, then $\mathtt{Hom}\,(G, K_n)$ is 0-connected. Generalizing this statement to higher dimension, we obtain the following conjecture.

Conjecture 6.2.2 (Babson & Kozlov, [**BK03b**, Conjecture 2.5])**.** [1]
Let G be any graph. If the maximal valency of G is equal to d, then $\mathtt{Hom}\,(G, K_n)$ is k-connected, for all integers $k \geq -1$, $n \geq d+k+2$.

Next, let us recall an important class of manifolds.

Definition 6.2.3. *For an arbitrary positive integer n, the* **Stiefel manifold** *$V_k(\mathbb{R}^n)$ is the set of the orthonormal k-frames in an n-dimensional Euclidean space, topologized as subspace of $(\mathbb{R}^n)^k$.*

Stiefel manifolds are homogeneous spaces and play an important role in the study of characteristic classes, see [**MS74**].

Conjecture 6.2.4 (Csorba, [**Cs04b**])**.**
The complex $\mathtt{Hom}\,(C_5, K_n)$ is homeomorphic to $V_2(\mathbb{R}^{n-1})$, for all $n \geq 1$.

The cases $n = 1, 2$ are tautological, as both spaces are empty. The example on the Figure 2.1.4 verifies the case $n = 3$: $\mathtt{Hom}\,(C_5, K_3) \cong S^1 \coprod S^1$. Several cases, including $n = 4$ have been recently verified by Csorba & Lutz, see [**CL04**].

Returning to the Stiefel-Whitney characteristic classes, we have the following hypothesis.

Conjecture 6.2.5 (Babson & Kozlov, [**BK04**, Conjecture 2.5])**.**
The equation

$$\varpi_1^{n-2}(\mathtt{Hom}\,(C_{2r+1}, K_n)) = 0, \text{ for all } n \geq 2 \tag{6.2.1}$$

is true for an arbitrary positive integer r.

Clearly, the case $n = 2$ is obvious, since $\mathtt{Hom}\,(C_{2r+1}, K_2)) = \emptyset$. The Conjecture 6.2.5 has been proved in [**BK04**] for $r = 1$ and arbitrary $n \geq 2$, as well as for odd n and arbitrary r, see here Theorem 5.1.4. For $r = 2$, $n = 4$, the equation (6.2.1) follows from the fact that $\mathtt{Hom}\,(C_5, K_4) \cong \mathbb{RP}^3$, and the analysis of the corresponding $\mathbb{Z}_2$-action on $\mathbb{RP}^3$.

[1] At the time of the writing of this survey, this conjecture has been proved and is now a theorem, see [**CK04b**].

We remark here that the Conjecture 6.2.5, coupled with Theorem 3.3.5, implies the Conjecture 5.1.3. Note also that, as previously remarked, for a fixed value of n, if the equation (6.2.1) is true for C_{2r+1}, then it is true for any $C_{2\tilde{r}+1}$, if $r \geq \tilde{r}$.

We finish with another conjecture by Lovász. In [**BW04**], Brightwell & Winkler have shown the following result.

Theorem 6.2.6 (Brightwell & Winkler, [**BW04**]).
Let G be an arbitrary graph. If for any graph T, with maximal valency at most d, the graph $\mathtt{Hom}_1(T, G)$ is connected or empty, then $\chi(G) \geq \frac{d}{2} + 2$.

Lovász has suggested that this statement can be strengthened, and that furthermore, a higher dimensional analog is true.

Conjecture 6.2.7 (Lovász). *Let G be an arbitrary graph. If for any graph T, with maximal valency at most d, the complex $\mathtt{Hom}(T, G)$ is k-connected or empty, then $\chi(G) \geq d + k + 2$.*

Acknowledgements. We thank Peter Csorba, Sonja Čukić, Alexander Engström, as well as the anonymous referees, for the helpful comments concerning the presentation in this paper.

BIBLIOGRAPHY

[AM94] A. Adem, J. Milgram, *Cohomology of finite groups*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], **309**, Springer-Verlag, Berlin, 1994.

[Al88] N. Alon, *Some recent combinatorial applications of Borsuk-type theorems*, Algebraic, extremal and metric combinatorics, 1986 (Montreal, PQ, 1986), pp. 1–12, London Math. Soc. Lecture Note Ser., **131**, Cambridge Univ. Press, Cambridge, 1988.

[AFL86] N. Alon, P. Frankl, L. Lovász, *The chromatic number of Kneser hypergraphs*, Trans. Amer. Math. Soc. **298** (1986), no. 1, pp. 359–370.

[AH76] K. Appel, W. Haken, *Every planar map is four colorable*, Bull. Amer. Math. Soc. **82**, (1976), pp. 711–712.

[AH89] K. Appel, W. Haken, *Every planar map is four colorable*, Contemp. Math., vol. 98, Amer. Math. Soc., Providence, RI, 1989.

[Ba05] E. Babson, private communication, 2005.

[BK03a] E. Babson, D.N. Kozlov, *Topological obstructions to graph colorings*, Electron. Res. Announc. Amer. Math. Soc. **9** (2003), pp. 61–68.

[BK03b] E. Babson, D.N. Kozlov, *Complexes of graph homomorphisms*, Israel J. Math., in press.
`arXiv:math.CO/0310056`

[BK04] E. Babson, D.N. Kozlov, *Proof of the Lovász Conjecture*, Annals of Mathematics, in press.
`arXiv:math.CO/0402395`

[Bar78] I. Bárány, *A short proof of Kneser's conjecture*, J. Combin. Theory Ser. A 25 (1978), no. 3, pp. 325–326.

[BSS81] I. Bárány, S.B. Shlosman, A. Szűcs, *On a topological generalization of a theorem of Tverberg*, J. London Math. Soc. (2) **23** (1981), no. 1, pp. 158–164.

[Bj96] A. Björner, *Topological Methods,* Handbook of Combinatorics, vol. 1,2, (eds. R. Graham, M. Grötschel and L. Lovász), Elsevier, Amsterdam, 1995, pp. 1819–1872.

[Bre93] G.E. Bredon, *Topology and geometry*, Graduate texts in mathematics **139**, Springer-Verlag, New York, 1993.

[Bre72] G.E. Bredon, *Introduction to compact transformation groups*, Pure and Applied Mathematics **46**, Academic Press, New York-London, 1972.

[BW04] G.R. Brightwell, P. Winkler, *Graph homomorphisms and long range action*, Graphs, morphisms and statistical physics, DIMACS Ser. Discrete Math. Theoret. Comput. Sci., **63**, Amer. Math. Soc., Providence, RI, 2004, pp. 29–47.

[Cay78] A. Cayley, *On the coloring of maps*, Proc. London Math. Soc. vol. 9, (1878), p.148.

[Co73] M. Cohen, *A course in simple-homotopy theory*, Graduate Texts in Mathematics, Vol. 10, Springer-Verlag, New York-Berlin, 1973.

[Cs04a] P. Csorba, *Homotopy type of the box complexes*, preprint, 11 pages, 2004. `arXiv:math.CO/0406118`

[Cs04b] P. Csorba, private communication, 2004.

[CL04] P. Csorba, F. Lutz, private communication, 2004.

[CK04a] S.Lj. Čukić, D.N. Kozlov, *The homotopy type of the complexes of graph homomorphisms between cycles*, Discrete Comp. Geometry, in press. `arXiv:math.CO/0408015`

[CK04b] S.Lj. Čukić, D.N. Kozlov, *Higher connectivity of graph coloring complexes*, Int. Math. Res. Not. **2005**, no. 25, pp. 1543–1562. `arXiv:math.CO/0410335`

[tD87] T. tom Dieck, *Transformation groups*, de Gruyter Studies in Mathematics, 8. Walter de Gruyter & Co., Berlin, 1987. x+312 pp.

[Dir52] G.A. Dirac, *A property of 4-chromatic graphs and some remarks on critical graphs*, J. London Math. Soc. **27**, (1952), pp. 85–92.

[Dol88] V.L. Dol'nikov, *A combinatorial inequality*, (Russian) Sibirsk. Mat. Zh. **29** (1988), no. 3, pp. 53–58, 219; translation in Siberian Math. J. **29** (1988), no. 3, pp. 375–379.

[FFG86] A.T. Fomenko, D.B. Fuks, V.L. Gutenmacher, *Homotopic topology*, Translated from the Russian by K. Mályusz. Akadémiai Kiadó (Publishing House of the Hungarian Academy of Sciences), Budapest, 1986.

[For98] R. Forman, *Morse theory for cell complexes*, Adv. Math. **134**, (1998), no. 1, pp. 90–145.

[GJ76] M.R. Garey, D.S. Johnson, *The complexity of near-optimal graph coloring*, J. Assoc. Comp. Mach. **23**, (1976), pp. 43–49.

[GJ79] M.R. Garey, D.S. Johnson, *Computers and Intractability*, A guide to the theory of NP-completeness, A Series of Books in the Mathematical Sciences, W.H. Freeman and Co., San Francisco, 1979.

[GM96] S.I. Gelfand, Y.I. Manin, *Methods of homological algebra*, Springer-Verlag, Berlin Heidelberg, 1996.

[GR01] C. Godsil, G. Royle, *Algebraic Graph Theory*, Graduate texts in mathematics **207**, Springer-Verlag, New York, 2001.

[Gr02] J. Greene, *A new short proof of Kneser's conjecture*, Amer. Math. Monthly **109** (2002), no. 10, pp. 918–920.

[Gut80] F. Guthrie, *Note on the colouring of maps*, Proc. Roy. Soc. Edinburgh, vol. 10, (1880), p. 729.

[Had43] H. Hadwiger, *Über eine Klassifikation der Streckenkomplexe*, Vierteljschr. Naturforsch. Ges., Zürich, vol. 88, (1943), pp. 133–142.

[Har69] F. Harary, *Graph Theory*, Addison-Wesley Series in Mathematics, Reading, MA, 1969.

[Hat02] A. Hatcher, *Algebraic topology*, Cambridge University Press, Cambridge, 2002.

[Hea90] P.J. Heawood, *Map-Colour Theorems*, Quart. J. Math., Oxford ser., vol. 24, (1890), pp. 332–338.

[Hee69] H. Heesch, *Untersuchungen zum Vierfarbenproblem*, Bibliog. Institut, AG, Mannheim, 1969.

[HN04] P. Hell, J. Nešetřil, *Graphs and Homomorphisms*, Oxford Lecture Series in Mathematics and Its Applications **28**, Oxford University Press, 2004.

[HL04] S. Hoory, N. Linial, *A counterexample to a conjecture of Lovász on the χ-coloring complex*, preprint, 3 pages, 2004.
`arXiv:math.CO/0405339`

[KS77] P.C. Kainen, T.L. Saaty, *The Four-Color Problem*, McGraw-Hill, New York, 1977.

[Kem79] A.B. Kempe, *On the geographical problem of four colors*, Amer. J. Math. **2**, (1879), pp. 193–204.

[KLS93] S. Khanna, N. Linial, S. Safra, *On the hardness of approximating the chromatic number*, Proc. Israel Symp. Theoretical Computer Science 1993, pp. 250–260.

[Kn55] M. Kneser, Aufgabe 360, *Jber. Deutsch. Math.-Verein.* **58**, (1955/56), 2 Abt., 27.

[Ko99] D.N. Kozlov, *Complexes of directed trees*, J. Comb. Theory Ser. A **88** (1999), pp. 112–122.

[Ko00] D.N. Kozlov, *Collapsibility of $\Delta(\Pi_n)/S_n$ and some related CW complexes*, Proc. Amer. Math. Soc. **128** (2000), no. 8, pp. 2253–2259.

[Ko02] D.N. Kozlov, *Trends in Topological Combinatorics*, Habilitationsschrift, Bern University, 2002.
`http://www.math.kth.se/~kozlov/ps/main.ps`

[Ko04] D.N. Kozlov, *A simple proof for folds on both sides in complexes of graph homomorphisms*, Proc. Amer. Math. Soc., in press.
`arXiv:math.CO/0408262`

[Ko05a] D.N. Kozlov, *Collapsing along monotone poset maps*, preprint, 2005.
`arXiv:math.CO/0503416`

[Ko05b] D.N. Kozlov, *Simple homotopy types of Hom-complexes, neighborhood complexes, Lovász complexes, and atom crosscut complexes*, preprint, 12 pages, 2005.
`arXiv:math.AT/0503613`

[Kr92] I. Kříž, *Equivariant cohomology and lower bounds for chromatic numbers*, Trans. Amer. Math. Soc. **333** (1992), no. 2, pp. 567–577.

[Kr00] I. Kříž, A correction to *Equivariant cohomology and lower bounds for chromatic numbers*, Trans. Amer. Math. Soc. **352** (2000), pp. 1951–1952.

[dL03] M. de Longueville, *25 Jahre Beweis der Kneservermutung der Beginn der topologischen Kombinatorik*, (German) *[25th anniversary of the proof of the Kneser conjecture. The start of topological combinatorics]*, Mitt. Deutsch. Math.-Ver. 2003, no. 4, pp. 8–11.

[Lov78] L. Lovász, *Kneser's conjecture, chromatic number, and homotopy*, J. Combin. Theory Ser. A **25**, (1978), no. 3, pp. 319–324.

[Lov] L. Lovász, private communication.

[McL98] S. MacLane, *Categories for the Working Mathematician*, Second edition, Graduate Texts in Mathematics, **5**, Springer-Verlag, New York, 1998.

[McL95] S. MacLane, *Homology*, Reprint of the 1975 edition, Classics in Mathematics, Springer-Verlag, Berlin, 1995.

[Ma03] J. Matoušek, *Using the Borsuk-Ulam theorem. Lectures on topological methods in combinatorics and geometry*, with A. Björner and G. M. Ziegler, Universitext, Springer-Verlag, Berlin, 2003.

[Ma04] J. Matoušek, *A combinatorial proof of Kneser's conjecture*, Combinatorica **24**, (2004), no. 1, pp. 163–170.

[MZ04] J. Matoušek, G.M. Ziegler, *Topological lower bounds for the chromatic number: A hierarchy*, Jahresbericht der DMV, **106**, 71–90, 2004.

[May99] J.P. May, *A concise course in algebraic topology*, Chicago Lectures in Mathematics, University of Chicago Press, Chicago and London, 1999.

[McC01] J. McCleary, *A user's guide to spectral sequences*, Second edition, Cambridge Studies in advanced mathematics **58**, Cambridge University Press, Cambridge, 2001.

[MSTY] O. Melnikov, V. Sarvanov, R. Tyshkevich, V. Yemelichev, *Lectures on Graph Theory*, BI-Wissenschaftsverlag, Mannheim, 1994. [Transl. by N. Korneenko from Russian original, Moscow, "Science", 1990].

[Mes03] R. Meshulam, *Domination numbers and homology*, J. Combin. Theory Ser. A **102** (2003), no. 2, pp. 321–330.

[MS74] J.W. Milnor, J.D. Stasheff, *Characteristic classes*, Annals of Mathematics Studies **76**, Princeton University Press, Princeton, 1974.

[Ore67] O. Ore, *The Four Color Problem*, Academic Press, New York, 1967.

[Qu73] D. Quillen, *Higher algebraic K-theory* I, Lecture Notes in Mathematics **341**, (1973), pp. 85–148, Springer-Verlag.

[RSST] N. Robertson, D.P. Sanders, P.D. Seymour, R. Thomas, *The four-color theorem*, J. Combin. Theory, Ser. B **70**, (1997), pp. 2–44.

[Sar90] K.S. Sarkaria, *A generalized Kneser conjecture*, J. Combin. Theory Ser. B 49 (1990), no. 2, pp. 236–240.

[Sch78] A. Schrijver, *Vertex-critical subgraphs of Kneser graphs*, Nieuw Arch. Wiskd., III. Ser., (1978), pp. 454–461.

[Sta97] R.P. Stanley, *Enumerative combinatorics*, Vol. 1, 2nd edition, Cambridge Studies in Advanced Mathematics **49**, Cambridge University Press, Cambridge, 1997.

[Ste51] N.E. Steenrod, *The topology of fibre bundles*, Princeton University Press, Princeton, 1951; reprinted in *Princeton landmarks in mathematics and physics*, 1999.

[Th98] R. Thomas, *An update on the Four-Color Theorem*, Notices Amer. Math. Soc., vol. 45, no. 7, August 1998, pp. 848–859.

[Vi88] A. Vince, *Star chromatic number*, J. Graph Theory **12**, (1988), pp. 551-559.

[Wag37] K. Wagner, *Über eine Eigenschaft der ebenen Komplexe*, Math. Ann. **114**, (1937), pp. 570–590.

[Wel84] E. Welzl, *Symmetric graphs and interpretations*, J. Combin. Theory Ser. B, **37**, (1984), pp. 235-244.

[Wh78] G.W. Whitehead, *Elements of homotopy theory*, Graduate Texts in Mathematics **61**, Springer-Verlag, New York, 1978.

[Zhu01] X. Zhu, *Circular chromatic number: a survey*, Combinatorics, graph theory, algorithms and applications, Discrete Math. **229**. (2001), no. 1-3, pp. 371–410.

[Zie02] G.M. Ziegler, *Generalized Kneser coloring theorems with combinatorial proofs*, Invent. Math. **147**, (2002), pp. 671-691.

[Ziv96] R.T. Živaljević, *User's guide to equivariant methods in combinatorics*, Publ. Inst. Math. (Beograd) (N.S.) **59(73)**, (1996), pp. 114–130.

[Ziv97] R.T. Živaljević, *Topological Methods*, Handbook of discrete and computational geometry, CRC Press Ser. Discrete Math. Appl., CRC, Boca Raton, FL, 1997, pp. 209–224.

[Ziv98] R.T. Živaljević, *User's guide to equivariant methods in combinatorics, II*, 50th anniversary of the Math. Inst., Serbian Academy of Sciences and Arts, Publ. Inst. Math. (Beograd) (N.S.) **64(78)**, (1998), pp. 107–132.

[Ziv04] R.T. Živaljević, *WI-posets, graph complexes and $\mathbb{Z}_2$-equivalences*, preprint, 20 pages, 2004.
`arXiv:math.CO/0405419`

Equivariant Invariants and Linear Geometry

Robert MacPherson

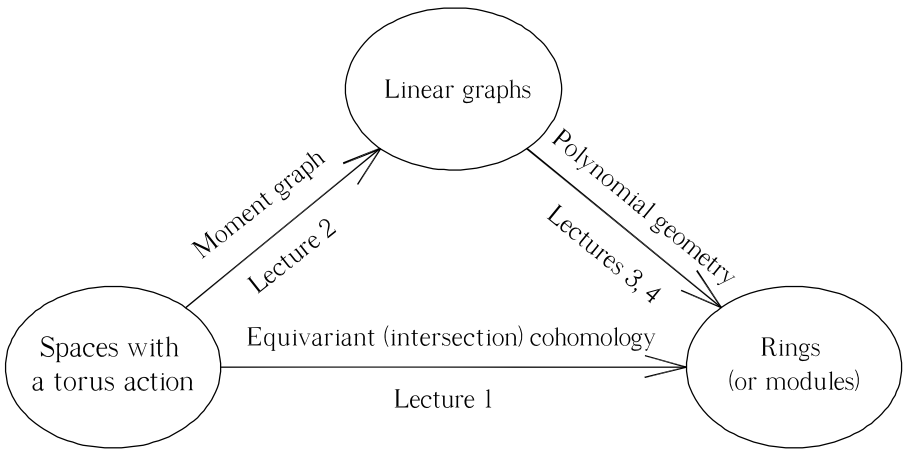
IAS/Park City Mathematics Series
Volume 14, 2004

Equivariant Invariants and Linear Geometry

Robert MacPherson

Introduction

⟨**0.1**⟩ This course will concern the following triangle of ideas.



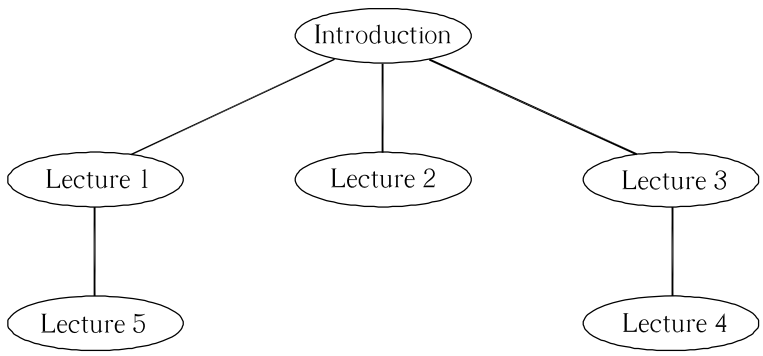
The vertices of this triangle represent mathematical objects. They will be defined in this introduction. The edges from one vertex to another represent mathematical constructions: given an object of the first type, we construct an object of the second type. These constructions will be the subject of the separate Lectures. The main theorem is that the diagram commutes: the construction on the bottom is the same as the composition of the two constructions on the top.

The constructions represented by the three edges all involve geometry, but they are of a completely different character from each other.

[1]Institute for Advanced Study, Princeton NJ 08540.
E-mail address: rdm@ias.edu.

©2007 American Mathematical Society

⟨**0.2**⟩ **Guide to reading.** The Lectures have been made independent of each other as much as possible, so as to allow several different points of entry into the subject. The following is the diagram of dependencies:



The mathematical knowledge required in advance has been kept to a minimum. *Starred sections and exercises are exceptions to this rule. They have mathematical prerequisites that go beyond those of the other sections, and are not needed for the rest of what we will do. The reader is invited to skip the *starred sections on a first reading.

The exercises are designed to be an integral part of the exposition.

⟨**0.3**⟩ **Credit and thanks.** All of my work on this subject has been joint with Bob Kottwitz, Mark Goresky, and Tom Braden. A deep study of moment graphs has been carried out by Victor Guillemin, Tara Holm, and Catalin Zara; Lecture 3 may serve as an introduction to their papers.

I am grateful to Tom Braden and to many participants of PCMI for corrections and improvements to this exposition.

0.1. Spaces with a Torus Action

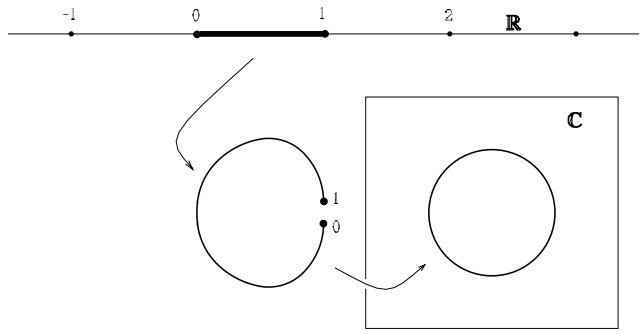
⟨**1.1**⟩ **Definition.** The *n-torus* T is the group

$$T = \mathbb{T}/L = (S^1)^n.$$

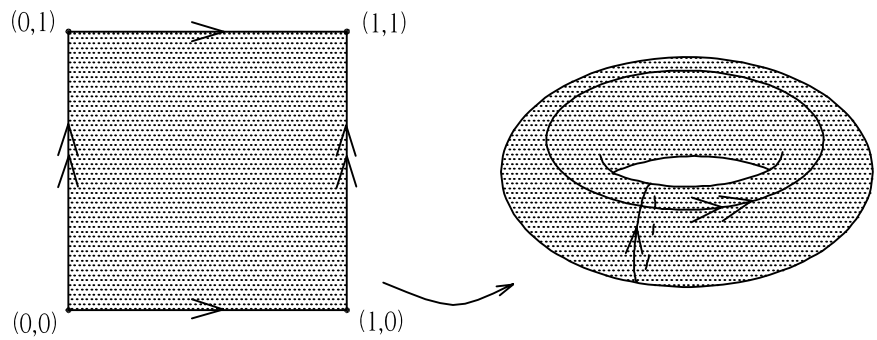
Here $\mathbb{T}$ is an n-dimensional real vector space, which we may take to be $\mathbb{R}^n$. The space $\mathbb{T}$ is a group under vector addition. The subgroup L is a lattice (i.e. a subgroup which is discrete as a topological space, with the property that $\mathbb{T}/L$ is compact). We may take L to be $\mathbb{Z}^n \subset \mathbb{R}^n$, the subgroup consisting of points whose coordinates are integers. The group S^1 is the unit circle group: the elements of norm 1 in the complex plane $\mathbb{C}$ considered as a group under multiplication. We may identify $\mathbb{R}/\mathbb{Z} \cong S^1$ by the map $\mathbb{R} \longrightarrow \mathbb{C}$ that sends x to $e^{2\pi i x}$, whose kernel is $\mathbb{Z}$. From this we get an identification

$$\mathbb{T}/L = \mathbb{R}^n/\mathbb{Z}^n = (\mathbb{R}/\mathbb{Z})^n = (S^1)^n.$$

⟨**1.2**⟩ We can visualize the n-torus as an n-cube $[0,1]^n$ with the opposite faces identified. For example, if $n = 1$, we have $S^1 = [0,1]/\sim$ where $\sim$ identifies 0 and 1.



Or, for example, the 2-torus is the square with the opposite edges identified,

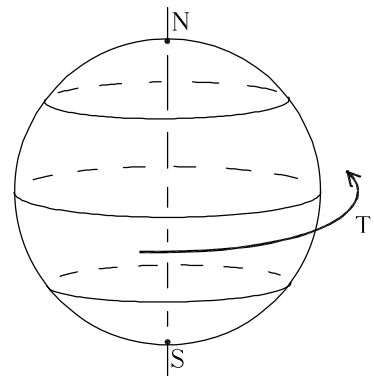


which shows why it's called a torus.

⟨1.3⟩ Exercise. Show in general that an n-torus as an n-cube $[0,1]^n$ with the opposite faces identified. Hint: Show every $\mathbb{Z}^n$ coset in $\mathbb{R}^n$ meets the unit cube $[0,1]^n \subset \mathbb{R}^n$, so $\mathbb{R}^n/\mathbb{Z}^n = [0,1]^n/\sim$ where $x \sim y$ when $x - y \in \mathbb{Z}^n$. Check that $\sim$ identifies opposite faces.

⟨1.4⟩ Exercise*. Let T be the n-torus $\mathbb{R}^n/\mathbb{Z}^n$ and let T' be the k-torus $\mathbb{R}^k/\mathbb{Z}^k$. Every group homomorphism $h : \mathbb{Z}^n \longrightarrow \mathbb{Z}^k$ extends uniquely to a continuous group homomorphism $\tilde{h} : \mathbb{R}^n \longrightarrow \mathbb{R}^k$, and so it passes to a continuous group homomorphism $\bar{h} : T \longrightarrow T'$. Show that the map $\mathrm{Hom}(\mathbb{Z}^n, \mathbb{Z}^k) \longrightarrow \mathrm{Hom}(T, T')$ that sends h to $\bar{h}$ is an isomorphism. Here $\mathrm{Hom}(T, T')$ is the set of continuous homomorphisms from T to T'.

⟨**1.5**⟩ **Definition.** A *space with a torus action* is a (Hausdorff) topological space X together with a self map $X \xrightarrow{t} X$ for every $t \in T$, notated $x \mapsto tx$, such that composition of homeomorphisms corresponds to multiplication in the group $t_1(t_2x) = (t_1 \times t_2)x$, and $(t, x) \mapsto tx$ is jointly continuous in x and t. We symbolize this by $T \circlearrowright X$. A quintessential example will be the circle action on the 2-sphere, where the circle rotates the 2-sphere about an axis. (Think of the action of the 24 hour day on the surface of the Earth.)



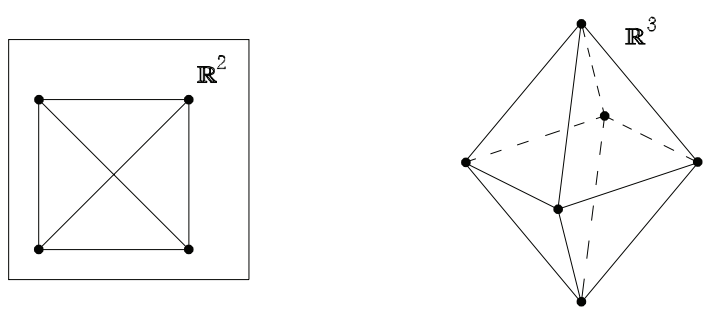
⟨**1.6**⟩ **Exercise.** Suppose that an n-torus T acts on a space X. Show that the orbit Tx of every point x in X is itself homeomorphic to a k-torus for some $k \leq n$. (Note the special case $k = 0$, which occurs at the North Pole N and the South Pole S of the example above.)

⟨**1.7**⟩ Why are we interested in torus actions $T \circlearrowright X$, rather than the actions of more general connected Lie groups $G \circlearrowright X$? In fact, computations for $G \circlearrowright X$ reduce to the computations for $T \circlearrowright X$, as explained in §3.8.11.

0.2. Linear Graphs

⟨**2.1**⟩ **Definition.** A *linear graph* is a finite set of points $\{v_i\}$ in a real vector space $\mathbb{V}$, called *vertices*, and a finite set of line segments $\{e_k\}$ in $\mathbb{V}$, called *edges* such that the two endpoints of each edge are both vertices.

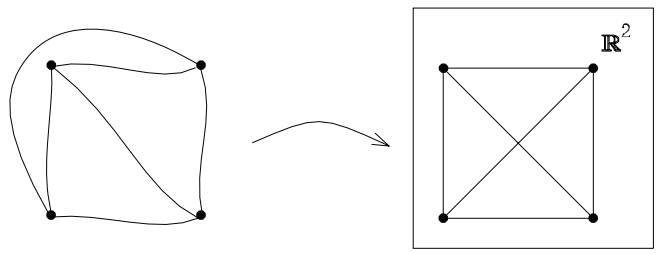
⟨**2.2**⟩ For example, the following are linear graphs:



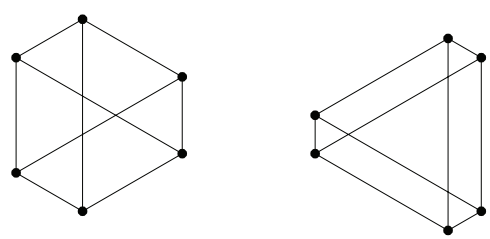
The first one, in $\mathbb{R}^2$, has four vertices and six edges. Note that the edges do not have to be disjoint: In this example, the two diagonals cross each other. The second one has six vertices and twelve edges. It is just the vertices and edges of an octahedron in $\mathbb{R}^3$. Any convex polyhedron gives rise to a linear graph by taking the vertices and the edges.

⟨**2.3**⟩ A topological graph is, of course, defined in a similar way, but without the embedding into a vector space. (For our purposes, a topological graph has at most

one edge between a pair of vertices, and has no edge going from a vertex to itself.) So a linear graph is a graph together with a mapping into $\mathbb{V}$ in such a way that its edges are mapped into straight lines.



⟨**2.4**⟩ **Equivalent linear graphs.** We consider two linear graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ in $\mathbb{V}$ to be *equivalent* if they correspond to the same topological graph Γ, and for each edge of Γ, the corresponding line in $\mathcal{G}_1$ is parallel to the corresponding line in $\mathcal{G}_2$. For example, these two linear graphs are equivalent:



⟨**2.5**⟩ **Directions and direction data.** We define a *direction* in $\mathbb{V}$ to be a parallelism class of lines in $\mathbb{V}$, or equivalently, a line through the origin in $\mathbb{V}$. (To specify a direction, it suffices to give a nonzero vector in $D \in \mathbb{V}$. If $\lambda \in \mathbb{R}$ is nonzero, then λD and D determine the same direction, since they determine the same line through the origin.) To give an equivalence class of linear graphs of graphs in $\mathbb{V}$, it suffices to give a topological graph with *direction data*, i.e. for each edge of the graph, we give a direction D.

⟨**2.6**⟩ **Exercise.** What is the dimension of the space of linear graphs equivalent to the linear graphs pictured above?

⟨**2.7**⟩ **Exercise.** Suppose an abstract graph is embedded in the plane as a linear graph. Can you find a formula for the dimension of its equivalence class?

⟨**2.8**⟩ **Exercise.** Consider the triangle graph with the direction data that assigns to the three edges the following three directions D in $\mathbb{R}^3$: $(1,0,1)$, $(-1,1,1)$, and $(0,-1,1)$. Show that there is no linear graph with this direction data.

0.3. Rings and Modules

⟨**3.1**⟩ Our rings R will all be graded algebras over the real numbers $\mathbb{R}$.

⟨**3.2**⟩ **Definition.** A *graded* $\mathbb{R}$*-algebra* is an $\mathbb{R}$-algebra with a direct sum decomposition

$$R = \bigoplus_{i \geq 0} R^i$$

into $\mathbb{R}$ vector spaces called the *graded pieces*, indexed by the non-negative integers, so that the multiplication is compatible with the grading: If $r \in R^i$ and $r' \in R^j$, then $rr' \in R^{i+j}$. Similarly, a *graded module* over R module M with a direct sum decomposition

$$M = \bigoplus_{i \geq 0} M^i$$

into $\mathbb{R}$ vector spaces, so that if $r \in R^i$ and $m \in M^j$, then $rm \in M^{i+j}$. Ring and module homomorphisms are required to respect the gradings.

All of our graded rings and modules will have the property that the odd numbered graded pieces are all zero, so $R = \bigoplus_{j \in \mathbb{Z}, j \geq 0} R^{2j}$. This perverse factor of 2 comes from the topological side of the story.

⟨**3.3**⟩ **The polynomial ring $\mathcal{O}(\mathbb{T})$.** We denote by $\mathcal{O}(\mathbb{T})$ the ring of real valued polynomial functions on the real vector space $\mathbb{T}$. It is the same as the ring of polynomials with real coefficients in n variables, where n is the dimension of $\mathbb{T}$. This is a graded ring. The $2j$-th graded piece is the space of polynomials of *homogeneous degree* j, i.e. the space spanned by monomials of degree j.

⟨**3.4**⟩ In all our graded rings and modules, the graded pieces are finite dimensional real vector spaces. Their dimensions are encoded in the Hilbert series.

Definition. The *Hilbert series* of R is the power series whose coefficients are the dimensions of the graded pieces of R

$$\mathrm{Hilb}(R) = \sum_{i \geq 0} x^i \dim(R^i).$$

Since all of our graded rings are zero in odd degree, it is conventional to introduce the variable $q = x^2$.

$$\mathrm{Hilb}(R) = \sum_{j \geq 0} (x^2)^j \dim(R^{2j}) = \sum_{j \geq 0} q^j \dim(R^{2j}).$$

⟨**3.5**⟩ **Proposition.** The Hilbert series of the polynomial ring $\mathcal{O}(\mathbb{T})$ is

$$\mathrm{Hilb}(\mathcal{O}(\mathbb{R})) = \left(\frac{1}{1-q}\right)^n$$

where n is the dimension of the vector space $\mathbb{T}$.

⟨**3.6**⟩ **Exercise.** Prove this. Hint: here are two possible strategies:

1) Show directly that the number of monomials in n variables of degree j is the coefficient of q^j in $(q-1)^n$. For example, the number of monomials of degree j in 3 variables is the $(j+1)$-st triangular number: the number of points in a triangular array with $j+1$ points on a side. This is because the monomials of degree j can be arranged in a triangular array.

$$\begin{array}{ccccccc} & & & z^3 & & & \\ & & xz^2 & & yz^2 & & \\ & x^2z & & xyz & & y^2z & \\ x^3 & & x^2y & & xy^2 & & y^3 \end{array}$$

2) Or, calculate the Hilbert series of the polynomial ring $\mathcal{O}(\mathbb{R})$ of polynomials in one variable

$$\mathrm{Hilb}(\mathcal{O}(\mathbb{R})) = 1 + q + q^2 + \cdots = \frac{1}{1-q}$$

then justify the following manipulations:

$$\mathrm{Hilb}(\mathcal{O}(\mathbb{T})) = \mathrm{Hilb}(\mathcal{O}(\underbrace{\mathbb{R} \times \cdots \times \mathbb{R}}_{n \text{ factors}}) = \mathrm{Hilb}(\underbrace{\mathcal{O}(\mathbb{R}) \otimes \cdots \otimes \mathcal{O}(\mathbb{R})}_{n \text{ factors}}) =$$

$$= \underbrace{\mathrm{Hilb}(\mathcal{O}(\mathbb{R})) \cdots \mathrm{Hilb}(\mathcal{O}(\mathbb{R}))}_{n \text{ factors}} = \left(\frac{1}{1-q}\right)^n$$

$\langle$**3.7**$\rangle$ **Exercise.** Let R be the ring of continuous functions on the real line $\mathbb{R}$, whose restriction to the positive reals $\mathbb{R}^{>0}$ and the negative reals $\mathbb{R}^{<0}$ are both polynomial functions. Show that R is a graded ring isomorphic to the polynomial ring in two variables x and y divided by the principal ideal generated by the polynomial xy, i.e. $R = \mathbb{R}[x, y]/(xy)$, and that its Hilbert series is $(1 + q)/(1 - q)$.

LECTURE 1
Equivariant Homology and Intersection Homology

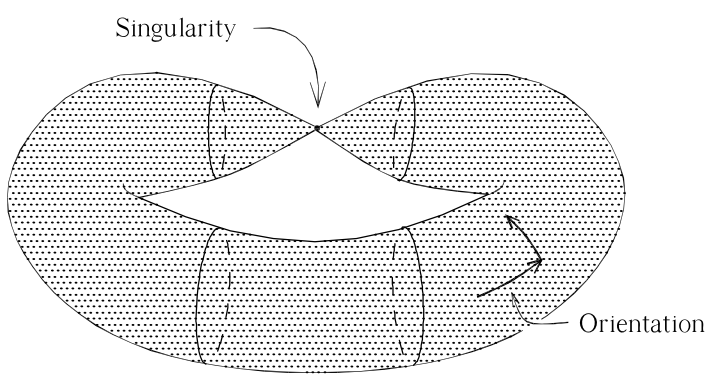
(Geometry of Pseudomanifolds)

1.1. Introduction

⟨**1.1**⟩ In this lecture, we will give a geometric way of defining equivariant homology and equivariant intersection homology. The standard definitions of these homology theories, as found in the literature, are good for proving properties, but are perhaps not so intuitive. In this lecture, we will consider $G \circlearrowright X$: an action of a general Lie group G on a space X, although in the other lectures we are interested mainly in the case that G is a torus T.

⟨**1.2**⟩ The definitions we present are based on the notion of a pseudomanifold. A k-dimensional manifold is a space that looks locally like k-dimensional Euclidean space near every point. A k-dimensional pseudomanifold P is allowed to have *singularities*, i.e. points where it doesn't locally look like Euclidean space. However, it must satisfy two properties:

(1) The part of P where it is a k-manifold is open and dense in P and it must be oriented.
(2) The set of singularities has dimension at most $k-2$ (i.e. codimension at least 2).



A pseudomanifold (the pinched torus)

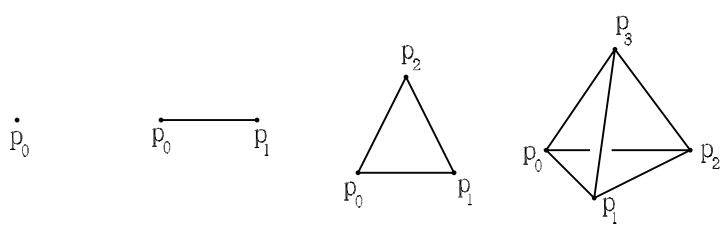
There are several ways to make this intuitive notion of a pseudomanifold rigorous. We will use simplicial complexes, because that is the one most in keeping with the spirit of these notes. Readers who are comfortable with pseudomanifolds can skip directly to §1.4

⟨**1.3**⟩ Equivariant homology theories are difficult to compute directly from the definitions as given in this Lecture. However, the methods of Lectures 3 to 5 provide effective computations in many interesting cases.

1.2. Simplicial Complexes

Readers familiar with simplicial complexes and orientations can skip this section.

⟨**2.1**⟩ A *k-simplex* Δ is the convex hull of $k+1$ points $p_0, \ldots, p_k$ in general position in some Euclidean space. Here general position just means that the points don't all lie in any $(k-1)$-dimensional Euclidean subspace. The k-simplex is a polyhedron. Its faces are themselves simplices; they are the convex hulls of subsets of the points p_i. The points p_i are the vertices of Δ.



0-simplex 1-simplex 2-simplex 3-simplex

⟨**2.2**⟩ **Definition.** A *simplicial complex* is a set S of simplices in some Euclidean space with the properties

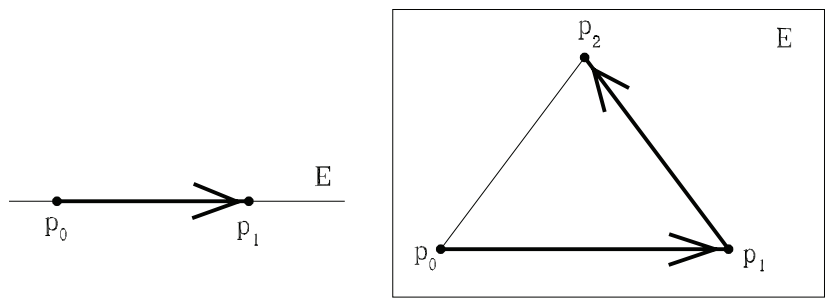
(1) Any two simplices in S are either disjoint or intersect in a set that is a face of each of them.
(2) Any face of a simplex in S is itself in S.

⟨**2.3**⟩ **Spaces of finite type.** We define a *space of finite type* to be a topological space homeomorphic to the difference $S - S'$ where S is a simplicial complex and S' is a sub simplicial complex. We will assume without further mention that all of our spaces are of finite type. Spaces generated by finite operations, such as real algebraic varieties, or their images under algebraic maps, are all of finite type (although proving it takes technology developed over many years). On the other hand, a Cantor set, or $\mathbb{Z}$ are not of finite type.

⟨**2.4**⟩ An *orientation* $\mathbb{O}$ of a simplex Δ is an ordering of the vertices of Δ, two orderings being considered equivalent if one is an even permutation of the other. (This definition doesn't work for a 0-simplex. An orientation of a 0-simplex is simply one of the symbols $+$ or $-$.) Any simplex has exactly two orientations, these two orientations are called opposite orientations of each other.

⟨**2.5**⟩ An orientation of a Euclidean space is an ordered set of basis vectors, two being considered equivalent if one is a continuous deformation of the other. We

can draw an orientation by representing the basis vectors as arrows, and signaling the ordering by placing the tail of each arrow at the head of the previous one. An orientation $\mathbb{O}$ of k-simplex Δ determines an orientation of the k-dimensional Euclidean space E containing Δ as follows: Suppose $\mathbb{O} = \{p_0 < p_1 < \cdots < p_k\}$. Then $\{p_1 - p_0, p_2 - p_1, \ldots, p_k - p_{k-1}\}$ is the ordered basis.



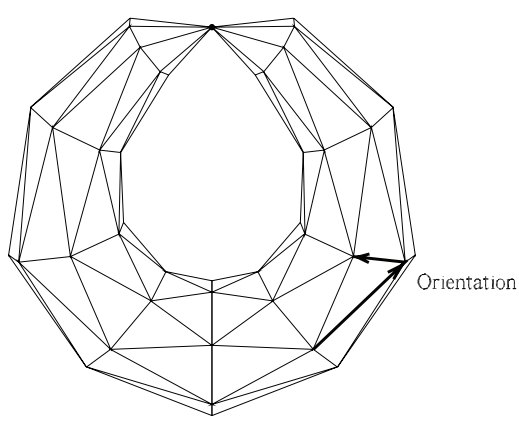
Exercise. Show that two orientations of Δ are equivalent if and only if they determine equivalent ordered bases of E.

$\langle$**2.6**$\rangle$ If Δ is a k-simplex and Δ' is a $(k-1)$-simplex, an orientation $\mathbb{O}$ of Δ induces an orientation $\mathbb{O}'$ of Δ' as follows: Pick an equivalent ordering such that the unique vertex of Δ not in Δ' is the last one of the ordering. Then $\mathbb{O}'$ is the restriction of that ordering to Δ'. (This definition doesn't work if Δ is a 1-simplex. In this case, $\mathbb{O}'$ is $-$ if Δ' is the first vertex of the ordering, and it is $+$ if it is the second one.)

1.3. Pseudomanifolds

$\langle$**3.1**$\rangle$ **Definition.** A *k-dimensional pseudomanifold* is a simplicial complex together with an orientation $\mathbb{O}(\Delta)$ of each of its k-simplices, with the following properties:

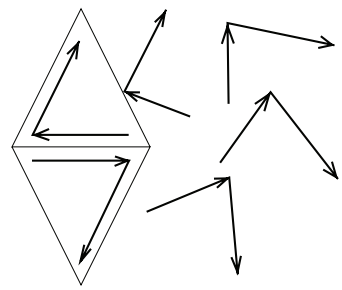
(1) Every simplex is a face of some k-simplex.
(2) Every $(k-1)$-simplex is the face of exactly two k-simplices.
(3) *(The continuity of orientation property)* If Δ' is a $(k-1)$-simplex and Δ and $\tilde{\Delta}$ are the two k-simplices that contain Δ' in their boundary, then the given orientations $\mathbb{O}(\Delta)$ and $\mathbb{O}(\tilde{\Delta})$ induce opposite orientations on Δ'.



A pseudomanifold (the simplicial pinched torus)

⟨**3.2**⟩ The following exercise shows why property 3 is called continuity of orientation.

Exercise. Suppose that Δ and $\tilde{\Delta}$ are two k-simplices in a Euclidean k-space, and that they intersect in a $(k-1)$-simplex Δ'. Show that the orientations $\mathbb{O}(\Delta)$ and $\mathbb{O}(\tilde{\Delta})$ induce opposite orientations on Δ' if and only if the ordered basis for E determined by Δ as in exercise 1.2.5 can be continuously deformed into the ordered basis for E determined by $\mathbb{O}(\tilde{\Delta})$.



A path in the space of ordered bases of the plane

⟨**3.3**⟩ **Definition.** A *k-dimensional pseudomanifold with boundary* is a simplicial complex S, an orientation $\mathbb{O}(\Delta)$ of each of its k-simplices, and a sub simplicial complex B called the *boundary*, with the following properties:

(1) The boundary B is a $(k-1)$-dimensional pseudomanifold
(2) Every simplex of S is a face of some k-simplex.
(3) Every $(k-1)$-simplex Δ' that not in B is the face of exactly two k-simplices, and the continuity of orientation property holds for Δ'.
(4) Every $(k-1)$-simplex Δ' in B is the face of exactly one k-simplex Δ in S. The orientation of Δ' induced from $\mathbb{O}(\Delta)$ coincides with the orientation $\mathbb{O}(\Delta')$ of Δ' from the pseudomanifold structure on B.

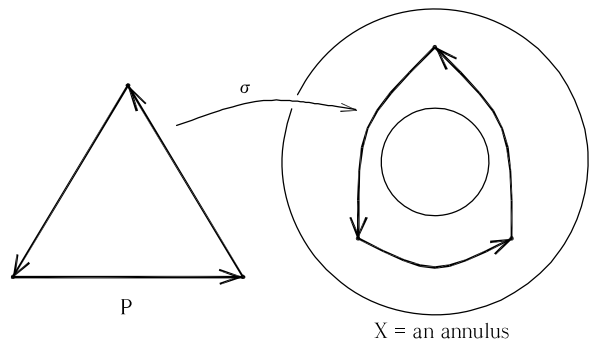
⟨**3.4**⟩ **Exercise.** Show that the continuity of orientation property for the boundary B of a pseudomanifold with boundary follows from the other properties in the definition.

1.4. Ordinary Homology Theory

As a warm up, we will give a definition of ordinary homology theory in the spirit of the definitions of more complicated theories to come. This definition of ordinary homology has roots going back to Poincaré and Veblen and the earliest days of homology theory.

⟨**4.1**⟩ **Definition.** Let X be a topological space. An *i-cycle* is an i-dimensional pseudomanifold P together with a map $\sigma : P \longrightarrow X$.

The idea is that an i-cycle captures the "holes" in a topological space by surrounding them. For example, the following 1-cycle surrounds the hole in the annulus:

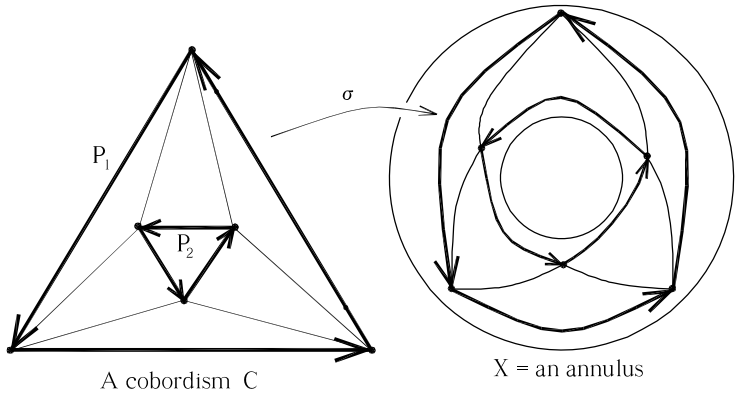


We will refer to the i-cycle $\sigma : P \longrightarrow X$ by the symbol P when there's no confusion about what σ is.

⟨**4.2**⟩ **Definition.** If $\sigma_1 : P_1 \longrightarrow X$ and $\sigma_2 : P_2 \longrightarrow X$ are two i-cycles, then the *sum* $P_1 + P_2$ is their union $\sigma : P_1 \cup P_2 \longrightarrow X$ where $\sigma|P_1 = \sigma_1$ and $\sigma|P_2 = \sigma_2$.

The *negative* $-P$ of an i-cycle is the same i-cycle P with the opposite orientation for every i-simplex. As usual, $P_1 - P_2$ is $P_1 + (-P_2)$.

⟨**4.3**⟩ **Definition.** A *cobordism* between two i-cycles $\sigma_1 : P_1 \longrightarrow X$ and $\sigma_2 : P_2 \longrightarrow X$ is a $(i+1)$-dimensional pseudomanifold with boundary C, and a map $\sigma : C \longrightarrow X$ such that the boundary B of C is $P_1 - P_2$ and the restriction of σ to B coincides with σ_1 and σ_2. Two i-cycles $\sigma_1 : P_1 \longrightarrow X$ and $\sigma_2 : P_2 \longrightarrow X$ are said to be *cobordant* if there is a cobordism between them.



The 1-cycles P_1 and P_2 are cobordant

The idea behind this definition is that if P_1 and P_2 are cobordant, they surround the same holes in the same way. For example, if σ has appropriate differentiability assumptions so that it makes sense, any closed differential i-form will have the same integral on P_1 and on P_2, by Stokes' Theorem.

⟨**4.4**⟩ **Proposition.** Cobordism is an equivalence relation among i-cycles.

Exercise Prove this. For example, if S_1 is a cobordism between P_1 and P_2, and S_2 is a cobordism between P_2 and P_3, then S_1 and S_2 can be glued together to provide a cobordism between P_1 and P_3.

⟨**4.5**⟩ **Definition − Proposition.** The i-th homology, notated $H_i(X)$, is the set of cobordism classes of i-cycles. The operations $+$ and $-$ induce the structure of an Abelian group on this set. The identity element is represented by the empty pseudomanifold.

⟨**4.6**⟩ For example, if X is the annulus, $H_0(X)$ is $\mathbb{Z}$ generated by a point, and $H_1(X)$ is $\mathbb{Z}$ generated by the cycle P_1 or P_2 as in the pictures above.

⟨**4.7**⟩ **Exercise.** Show for any X that $H_0(X)$ is $\mathbb{Z}^k$ where k is the number of path connected components of X.

⟨**4.8**⟩ **Convention.** We write $H_*(X)$ for $\bigoplus_i H_i(X)$. It's a summation convention: wherever a star appears, it means a direct sum over the possible indices i that might appear there.

1.5. Basic Definitions of Equivariant Topology

⟨**5.1**⟩ A *topological group* G is a set that is simultaneously a group and a topological space, with the property that the multiplication operation $G \times G \longrightarrow G$ and the inverse operation $G \longrightarrow G$ are both continuous. G is a Lie group if it is one of our spaces of finite type §1.2.3 (or, what turns out to be the same thing for topological groups, if it's a topological manifold with finitely many connected components.)

⟨**5.2**⟩ A *space with a group action* $G \circlearrowright X$ is a topological space X (which for us will always be of finite type), and an a map $G \times X \longrightarrow X$ that is continuous, such that $(g \cdot g')x = g(g'(x))$ (§0.1.5).

⟨**5.3**⟩ **The equivariant category.** Suppose $G \circlearrowright X$ and $G' \circlearrowright X'$ are two topological spaces with a group action. A morphism $G \circlearrowright X \Longrightarrow G' \circlearrowright X'$ is a continuous group homomorphism $\phi : G \longrightarrow G'$ together with a continuous map $\psi : X \longrightarrow X'$ such that $\psi(gx) = \phi(g)\psi(x)$.

For example, for any $G \circlearrowright X$ there is a canonical morphism $G \circlearrowright X \Longrightarrow 1 \circlearrowright X/G$. Here 1 is the one element group; X/G is the quotient space $X/\sim$ where $\sim$ is the equivalence relation $x \sim x'$ if there is a $g \in G$ such that $gx = x'$; $\phi : G \longrightarrow 1$ is the only thing it could be; and $\phi : X \longrightarrow X/G$ is the quotient map.

⟨**5.4**⟩ G **equivariant maps.** If G is a fixed group, then *the category of G-spaces* is the sub category of the equivariant category where the map ϕ on G is the identity. The maps in this category are called G *equivariant maps.* In other words, if X and X' are both G-spaces, then an equivariant map from X to X' is a continuous map $\psi : X \longrightarrow X'$ such that $\psi(gx) = g\psi(x)$ for all $g \in G$ and all $x \in X$. We consider two G spaces *equivalent* if they are isomorphic in this category. This means that the map ψ is a homeomorphism.

⟨**5.5**⟩ **Free actions.** An action of G on X is *free* if no element of G except the identity fixes any point in X, i.e. $gx = x$ implies g is the identity.

Another commonly used terminology for the same thing is this: The map $\pi : X \longrightarrow X/G$ is called a *principal G-bundle* if and only if the action $G \circlearrowright X$ is free. In this terminology, X/G is called the *base* of the principal bundle; X is called the total space; and π is called the projection.

Yet another popular terminology is to say that X is a *G-torsor* over X/G.

⟨**5.6**⟩ **Exercise.** Show that every orbit Gx of a free action $G \circlearrowright X$ is homeomorphic to G.

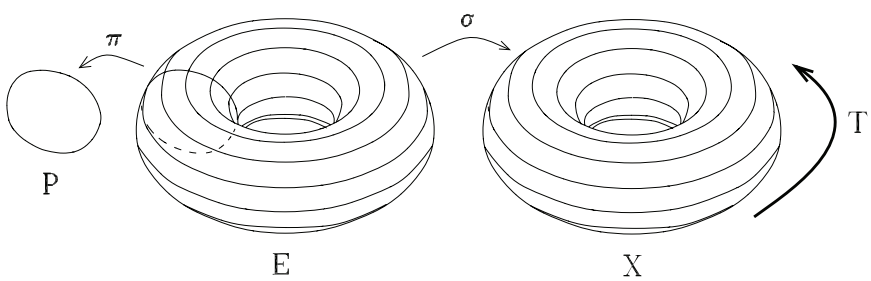
1.6. Equivariant Homology

⟨**6.1**⟩ **Definition.** Let $G \circlearrowright X$ be a G-space. An *equivariant i-cycle* is a diagram

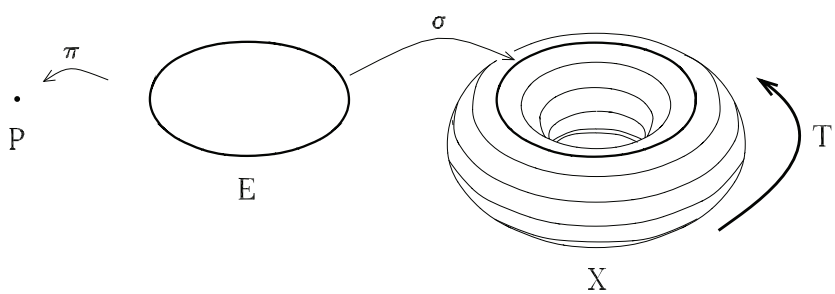
$$P \xleftarrow{\pi} E \xrightarrow{\sigma} X$$

where P is an i-dimensional pseudomanifold, E is a G-space with a free G action, $\pi : E \longrightarrow P$ is the projection on the quotient by the G action (so $P = E/G$), and σ is a G equivariant map.

Another way of describing the same thing (see §1.5.5) is this: A i-cycle is an equivariant map into X of the total space of a G-principal bundle E, whose base space is an i-dimensional pseudomanifold.



An equivariant 1-cycle for the circle T acting on the torus X



An equivariant 0-cycle for the circle T acting on the torus X

⟨**6.2**⟩ When we want to refer to an i-cycle, we use the symbol P for the pseudomanifold, even though it is really a 4-tuple of data. The *sum* $P_1 + P_2$ of two equivariant i-cycles P_1 and P_2 is $P_1 \cup P_2 \longleftarrow E_1 \cup E_2 \longrightarrow X$. The **negative** $-P$ of an equivariant i-cycle is the same i-cycle with the opposite orientation on P.

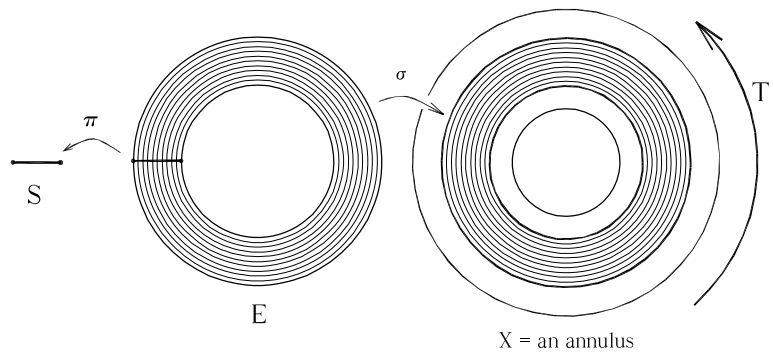
⟨**6.3**⟩ **Definition.** A *cobordism* between two equivariant i-cycles P_1 and P_2 is a commutative diagram

$$\begin{array}{ccccc} C & \xleftarrow{\pi} & E & \xrightarrow{\sigma} & X \\ \text{inclusion as boundary B} \uparrow & & \text{inclusion} \uparrow & & \uparrow = \\ P_1 - P_2 & \xleftarrow{\pi_1,\pi_2} & E_1 \cup E_2 & \xrightarrow{\sigma_1,\sigma_2} & X \end{array}$$

where C is a pseudomanifold with boundary $B = P_1 - P_2$, E is a G principal bundle over C, σ is an equivariant map, and $E_1 \cup E_2$ is $\pi^{-1}(P_1 \cup P_2)$.

Two equivariant i-cycles P_1 and P_2 are said to be *cobordant* if there is a cobordism between them.

For example, the following picture shows a cobordism between two equivariant 0-cycles. Here $T \circlearrowright X$ is a circle rotating the annulus.



⟨**6.4**⟩ **Proposition.** Cobordism is an equivalence relation on equivariant i-cycles.

⟨**6.5**⟩ **Definition – Proposition.** The i-th equivariant homology of $G \circlearrowright X$, which we will notate $H_i(G \circlearrowright X)$, is the set of cobordism classes of equivariant i-cycles. The operations $+$ and $-$ induce the structure of an Abelian group on this set.

The usual notation for the ith equivariant homology group is $H_i^G(X)$, and Borel's original notation was $H_i(X_G)$.

⟨**6.6**⟩ The topological juice of this definition comes in the requirement that the G action on the cycles and cobordisms be free.

Exercise. Show that if we dropped the requirement that the G action be free from the definitions of equivariant cycles and cobordisms, then we would just get the homology of the quotient $H_i(X/G)$.

⟨**6.7**⟩ **Exercise.** Show that if G acts freely on X, then $H_i(G \circlearrowright X)$ is, in fact, the homology of the quotient $H_i(X/G)$.

For example, if T is the 1-torus, X is the 2-torus, and T rotates X as in the pictures above, then the T action is free and $H_i(T \circlearrowright X) = H_i(X/T)$. Since X/T is a circle, we have

$$H_i(T \circlearrowright X) = \begin{cases} \mathbb{R} & \text{if} \quad i = 0 \\ \mathbb{R} & \text{if} \quad i = 1 \\ 0 & \text{otherwise} \end{cases}$$

The two nonzero equivariant homology groups are generated by the equivariant cycles shown in the pictures above.

Similarly, the equivariant homology of the annulus is nonzero only in degree 0, where it is generated by the 0-cycle shown in the picture.

In both of these examples, the equivariant homology is smaller than the ordinary homology. For the cases we are going to study later, the reverse is true.

⟨**6.8**⟩ **Exercise.** Show that $H_0(G \circlearrowright X) = H_0(X)$, the free group generated by the connected components of X.

⟨**6.9**⟩ **Exercise.** Show that $H_i(1 \circlearrowright X) = H_i(X)$, where 1 is the one element group.

1.7. Formal Properties of Equivariant Homology

⟨**7.1**⟩ **Functoriality.** Suppose that we have an equivariant map

$$G \circlearrowright X \Longrightarrow G' \circlearrowright X'.$$

Then there is an induced map on equivariant homology

$$H_i(G \circlearrowright X) \longrightarrow H_i(G' \circlearrowright X').$$

If $G = G'$ and the map $G \longrightarrow G'$ is the identity, then this map can be defined simply by composition of G-equivariant maps. The general case is a little complicated: If

$$P \xleftarrow{\pi} E \xrightarrow{\sigma} X$$

is a cycle for $H_i(G \circlearrowright X)$, then

$$P \xleftarrow{\pi} G' \times_G E \xrightarrow{\sigma'} X'$$

is the corresponding cycle for $H_i(G' \circlearrowright X')$, where $G' \times_G E$ is the "associated bundle" defined as the quotient space $(G' \times E)/G$ where G acts on $G' \times E$ by $g(g', e) = (g' \cdot \phi g^{-1}, ge)$. The group G' acts freely on the associated bundle $G' \times_G E$ by $g'(g'', e) = (g \cdot g'', e)$. The quotient of this G' action is again P. The map $\sigma' : G' \times_G E \longrightarrow X'$ sends (g', e) to $g'(\psi(\sigma(e)))$. One needs to check that this is well defined.

⟨**7.2**⟩ **Coefficients.** Suppose that R is a ring containing the integers $\mathbb{Z}$. We can define the equivariant homology with coefficients in R by

$$H_i(G \circlearrowright X; R) = H_i(G \circlearrowright X) \otimes_{\mathbb{Z}} R$$

Assumption. *We will assume that our coefficient ring is the real numbers* $\mathbb{R}$. (Actually, any other field of characteristic zero would work as well for everything we will do.) With these coefficients, the equivariant homology is a real vector space; we notate it simply by $H_i(G \circlearrowright X)$. There are torsion phenomena that are killed by taking these coefficients, but the most interesting phenomena survive. Many of our theorems are false, or are at least much more complicated to state, for coefficients that are not fields of characteristic zero.

⟨**7.3**⟩ **The Kunneth theorem.** Given two spaces with group actions $G_1 \circlearrowright X_1$ and $G_2 \circlearrowright X_2$, we have the *Kunneth map*

$$\begin{array}{ccccc} H_i(G_1 \circlearrowright X_1) & \times & H_j(G_2 \circlearrowright X_2) & \longrightarrow & H_{i+j}((G_1 \times G_2) \circlearrowright (X_1 \times X_2)) \\ [P_1] & \times & [P_2] & \mapsto & [P_1 \times P_2] \end{array}$$

where $P_1 \times P_2$ is defined to be the equivariant $(i+j)$-cycle

$$P_1 \times P_2 \xleftarrow{\pi_1 \times \pi_2} E_1 \times E_2 \xrightarrow{\sigma_1 \times \sigma_2} X_1 \times X_2$$

Proposition. The Kunneth map induces an isomorphism

$$H_i(G_1 \circlearrowright X_1) \otimes H_j(G_2 \circlearrowright X_2) \xrightarrow{\cong} H_{i+j}((G_1 \times G_2) \circlearrowright (X_1 \times X_2)).$$

⟨**7.4**⟩ **Note on proofs.** From here until the end of the lecture, we will state hard theorems without proofs. What is important is to understand the statements, and to understand why the various objects and maps are well defined. If you invent proofs of any of these statements without knowing a lot of topology, that will be quite an accomplishment.

⟨**7.5**⟩ **The cohomology ring.** The cohomology is the vector space dual of the homology

$$H^i(G \circlearrowright X) = (H_i(G \circlearrowright X))^* = \mathrm{Hom}\,(H_i(G \circlearrowright X), \mathbb{R})\,.$$

Consider the diagram

$$H_i(G \circlearrowright X) \otimes H_j(G \circlearrowright X) \xrightarrow{\cong} H_{i+j}(G \times G \circlearrowright X \times X) \xleftarrow{\Delta} H_{i+j}(G \circlearrowright X)$$

where $\Delta : G \circlearrowright X \Longrightarrow (G \times G) \circlearrowright (X \times X)$ is the diagonal map that sends g to (g, g) and sends x to (x, x).

Now take the vector space dual of the whole diagram

$$H^i(G \circlearrowright X) \otimes H^j(G \circlearrowright X) \xleftarrow{\cong} H^{i+j}(G \times G \circlearrowright X \times X) \xrightarrow{\Delta^*} H^{i+j}(G \circlearrowright X)$$

Proposition. The composed map from the left to the right in this diagram is the multiplication rule of a ring structure on equivariant cohomology.

⟨**7.6**⟩ **Signs.** The product in cohomology satisfies the following sign rule: If $x \in H^i(G \circlearrowright X)$ and $y \in H^j(G \circlearrowright X)$, then $xy = (-1)^{ij} yx$. The reason for this it that if P_1 and P_2 are pseudomanifolds of dimensions i and j, then $P_1 \times P_2 = (-1)^{ij} P_2 \times P_1$. A nice exercise is to figure out how to define precisely the product of two pseudomanifolds, and to show this commutation rule.

⟨**7.7**⟩ **Exercise.** Use the map $1 \circlearrowright X \Longrightarrow G \circlearrowright X$ to show that there is a canonical map $H_i(X) \longrightarrow H_i(G \circlearrowright X)$, or dually $H^i(G \circlearrowright X) \longrightarrow H^i(X)$.

⟨**7.8**⟩ **Exercise.** Let A be the ring $H^*(G \circlearrowright \mathrm{pt})$, where pt is a point. Use the map $G \circlearrowright X \Longrightarrow G \circlearrowright \mathrm{pt}$ to show that there is a map $A \longrightarrow H^*(G \circlearrowright X)$, so $H^*(G \circlearrowright X)$ is an A-module. We will be interested mainly in actions where this map is an injection, unlike the illustrative examples considered in the last section.

⟨**7.9**⟩ **Homology vs. cohomology.** There is a psychological dilemma. An equivariant cohomology class is hard to imagine. It is an element in a dual space — it eats a homology class and gives you a number. But the cohomology is an algebra, which most people find to be an intuitive structure. An equivariant homology class is easy to visualize, but the homology forms a co-algebra, which is hard to think about.

Choosing the demons of dual spaces over the demons of co-algebras, our computations will be in equivariant cohomology. Of course, the computation of equivariant cohomology is mathematically equivalent to the computation of equivariant homology, so the information is the same in the end.

1.8. Torus Equivariant Cohomology of a Point

The ordinary homology of a point p is uninteresting: It is simply a 1-dimensional vector space in homology degree 0, and zero in every other degree. However, for Lie groups of positive dimension, the equivariant homology (or cohomology) or cohomology of a point is quite interesting: It is infinite dimensional.

⟨**8.1**⟩ **Generators of the circle equivariant homology of a point.** Suppose T^1 is a 1-torus and pt is a point. Then for every integer $k \geq 0$, we have the following equivariant $2k$-cycle:

$$\mathbb{CP}^k \xleftarrow{\ \pi\ } S^{2k+1} \xrightarrow{\ \sigma\ } \mathrm{pt}$$

Here S^{2k+1} is the real unit real sphere in complex $(k+1)$-space, given by the equation $|z_0|^2 + |z_1|^2 + \cdots + |z_k|^2 = 1$. The circle $T^1 = S^1 \subset \mathbb{C}$ acts on it freely by scalar multiplication. The quotient space $\mathbb{CP}^k$ is the complex projective k-space (see §2.3.1), a pseudomanifold (indeed a manifold) of real dimension $2k$.

⟨**8.2**⟩ It is useful to think about why this $2k$-cycle is nonzero. 1. The $2k+1$ sphere bounds the $2k+2$ ball $|z_0|^2 + |z_1|^2 + \cdots + |z_k|^2 \leq 1$. The S^1 action extends to the ball. If $k > 0$, the quotient is a pseudomanifold with boundary $\mathbb{CP}^k$. Why isn't this a cobordism to zero? 2. The fibers of the map $S^{2k+1} \longrightarrow \mathbb{CP}^k$ are all circles S^1. There is a "trivial" example of a map to $\mathbb{CP}^k$ whose fibers are all circles: if we had a homeomorphism $S^1 \times \mathbb{CP}^k \longrightarrow \mathbb{CP}^k$ which is cobordant to zero, since we can take $S^1 \times C(\mathbb{CP}^k) \longrightarrow C(\mathbb{CP}^k)$ where $C(\mathbb{CP}^k)$ is the cone on $\mathbb{CP}^k$. So, since our $2k$-cycle is not cobordant to zero, it must not be equivalent to the"trivial" example.

⟨**8.3**⟩ **Exercise*.** Prove that the $2k$-cycle above is not 0 in $H_{2k}(T^1 \circlearrowright \mathrm{pt})$. Hint: Use characteristic classes.

⟨**8.4**⟩ **Proposition.** The equivariant homology of $T^1 \circlearrowright \mathrm{pt}$, where T^1 is the 1-torus and pt is a point, is given by

$$H_i(T^1 \circlearrowright \mathrm{pt}) = \begin{cases} \mathbb{R} \quad \text{generated by } \mathbb{CP}^k & \text{if} \quad i = 2k \\ \qquad 0 & \text{if} \quad i \quad \text{is odd} \end{cases}$$

Dually, the equivariant cohomology ring is

$$H^*(T^1 \circlearrowright \mathrm{pt}) = \{\text{polynomial functions on } \mathbb{T}^1 = \mathbb{R}\} = \mathcal{O}(\mathbb{T}^1)$$

The basis $\{1, t, t^2, \ldots\}$ of $H^*(T^1 \circlearrowright \mathrm{pt})$ is dual to the basis $\{\mathbb{CP}^0, \mathbb{CP}^1, \mathbb{CP}^2, \ldots\}$ of $H_*(T^1 \circlearrowright \mathrm{pt})$.

⟨**8.5**⟩ **The torus equivariant homology of a point.** A general n-torus $T = \mathbb{T}/L$ is a product of n copies of the circle $T^1 = \mathbb{T}^1/L^1$. Therefore, we have the product of spaces with group action

$$T \circlearrowright \mathrm{pt} = \underbrace{T^1 \circlearrowright \mathrm{pt} \times T^1 \circlearrowright \mathrm{pt} \times \cdots \times T^1 \circlearrowright \mathrm{pt}}_{n \text{ factors}}$$

Applying the Kunneth theorem for cohomology, we have

$$\begin{aligned} H^*(T \circlearrowright \text{pt}) &= \underbrace{H^*(T^1 \circlearrowright \text{pt}) \otimes \cdots \otimes H^*(T^1 \circlearrowright \text{pt})}_{n \text{ factors}} \\ &= \left\{ \text{polynomial functions on } \underbrace{\mathbb{T}^1 \times \cdots \times \mathbb{T}^1}_{n \text{ factors}} \right\} \\ &= \{\text{polynomial functions on } \mathbb{T}\} = \mathcal{O}(\mathbb{T}) \end{aligned}$$

where $\mathbb{T} = \mathbb{R}^n$ is the product of n copies of $\mathbb{T}^1$.

1.9. The Equivariant Cohomology of a 2-Sphere

⟨9.1⟩ Homology of the fixed point set $N \cup S$. Suppose that the circle T^1 acts on the 2-sphere $X = S^2$ by rotation as in §0.1.4. There are two fixed points, the North pole N and the South pole S. The space $N \cup S$ is just two points, so its T^1 equivariant homology is just two copies of the equivariant homology of a point:

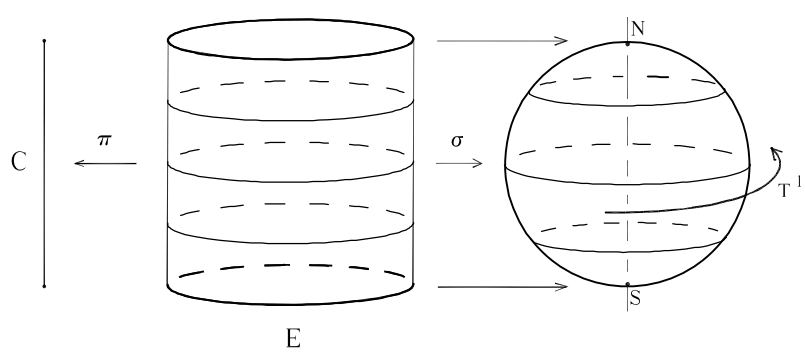
$$H_i(T^1 \circlearrowright (N \cup S)) = \begin{cases} \mathbb{R}[(\mathbb{CP}^k)_N] \oplus \mathbb{R}[(\mathbb{CP}^k)_S] = \mathbb{R}^2 & \text{if } i = 2k \\ 0 & \text{if } i \text{ is odd} \end{cases}$$

The equivariant map $N \cup S \hookrightarrow X$ is T^1 equivariant, so it induces a map on equivariant homology $H_i(T^1 \circlearrowright (N \cup S)) \longrightarrow H_i(T^1 \circlearrowright X)$.

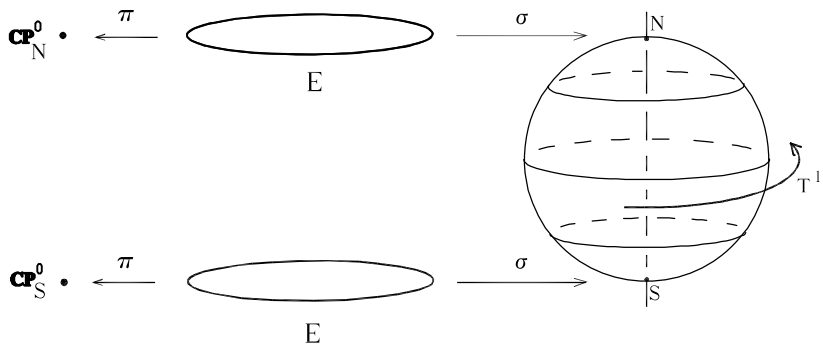
⟨9.2⟩ Circle equivariant homology of the 2-sphere. Proposition. The map

$$H_i(T^1 \circlearrowright (N \cup S)) \longrightarrow H_i(T^1 \circlearrowright X)$$

is an isomorphism for all $i > 0$. For $i = 0$, there is one relation $[\mathbb{CP}^0_N] = [\mathbb{CP}^0_S]$ given by the following cobordism whose boundary is $\mathbb{CP}^0_N - \mathbb{CP}^0_S$.



Cobordism giving the relation in $H_0(T^1 \circlearrowright X)$



The boundary of this cobordism

⟨**9.3**⟩ **Exercise*.** This result says that every equivariant cycle for $T \circlearrowright X$ is cobordant to one that maps into just the two fixed points N and S. The corresponding statement in ordinary homology (i.e. $1 \circlearrowright X$) is false. Can you see geometrically why this is true?

⟨**9.4**⟩ **Translating this calculation to equivariant cohomology.** In summary, the equivariant homology of X is a quotient of the equivariant homology of $N \cup S$; i.e. we have the exact sequence of graded vector spaces

$$0 \longrightarrow \mathbb{R} \xrightarrow{q} H_*(T^1 \circlearrowright (N \cup S)) \longrightarrow H_*(T^1 \circlearrowright X) \longrightarrow 0$$

where the map q sends 1 to $[\mathbb{CP}^0_N] - [\mathbb{CP}^0_S]$. Dualizing, the equivariant cohomology of X is a sub of the equivariant cohomology of $N \cup S$; i.e. we have the exact sequence of rings:

$$0 \longleftarrow \mathbb{R} \xleftarrow{q^*} H^*(T^1 \circlearrowright (N \cup S)) \longleftarrow H^*(T^1 \circlearrowright X) \longleftarrow 0.$$

Here $H^*(T^1 \circlearrowright (N \cup S)) = H^*(T^1 \circlearrowright N) \oplus H^*(T^1 \circlearrowright S)$ which is two copies of the ring $\mathcal{O}(\mathbb{T}^1)$ of polynomials on $\mathbb{T}^1$. The map q^* sends the difference of the identity elements of the two copies of the polynomial ring $1_N - 1_S$ to $1 \in \mathbb{R}$.

In other words, $H^*(T^1 \circlearrowright (N \cup S))$ is the ring of pairs (f_N, f_S) of polynomial functions on $\mathbb{T}^1 = \mathbb{R}$. The ring $H^*(T^1 \circlearrowright X)$ is pairs (f_N, f_S) such that $f_N(0) = f_S(0)$.

⟨**9.5**⟩ **The torus equivariant cohomology of a sphere.** Now suppose that an n torus T acts on the sphere $X = S^2$ by rotating it. By changing coordinates in the torus, we can arrange things so that $T = T^1 \times T^{n-1}$ where the circle T^1 acts on X as in the discussion above and the torus T^{n-1} acts trivially. Therefore, we have the product of spaces with group action

$$T \circlearrowright X = T^1 \circlearrowright X \times T^{n-1} \circlearrowright p.$$

Applying the Kunneth theorem, we get

$$\begin{aligned} H^*(T \circlearrowright X) &= H^*(T^1 \circlearrowright X) \otimes H^*(T^{n-1} \circlearrowright \text{pt}) \\ &= \big\{(f_N, f_S) \in \mathcal{O}(T^1) \oplus \mathcal{O}(T^1) \text{ such that } f_N|0 = f_S|0\big\} \otimes \mathcal{O}(\mathbb{T}^{n-1}) \\ &= \big\{(f_N, f_S) \in \mathcal{O}(T^1 \times \mathbb{T}^{n-1}) \text{ such that } f_N|\mathbb{T}^{n-1} = f_S|\mathbb{T}^{n-1}\big\} \end{aligned}$$

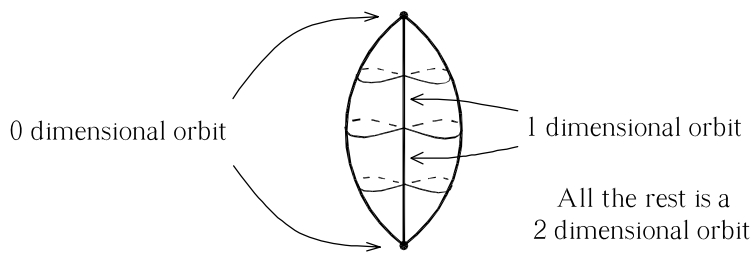
In Lecture 3, this simple calculation will lie at the root of all of our calculations of equivariant cohomology (and ordinary cohomology) of many complicated spaces.

1.10. Equivariant Intersection Cohomology

In this section, we will sketch the construction of intersection cohomology, so the reader can get the flavor.

Intersection cohomology is an invariant of pseudomanifolds. If the pseudomanifold is a manifold, then the equivariant intersection cohomology is the same as the ordinary cohomology. If the pseudomanifold is singular, then often it is the intersection cohomology (equivariant or otherwise) that is important for applications, rather than the ordinary cohomology.

⟨**10.1**⟩ Suppose X is a k-dimensional pseudomanifold of finite type. Consider the group $\mathbf{H} \circlearrowright X$ of self-homeomorphisms of X. The group $\mathbf{H}$ is infinite dimensional, but its orbits in X are of finite type because X is of finite type. The space X will be "uniformly singular" along the orbits of $\mathbf{H} \circlearrowright X$. For example, if X is "the suspension of ∞", we get this picture:



Orbits of the group of homeomorphisms, $\mathbf{H} \circlearrowright X$

Let X_c be the union of all the orbits of codimension c, i.e of dimension $k - c$. The largest orbit X_0 is an open dense k-manifold in X. X_1 is empty because X is a pseudomanifold. *We assume that X_c is empty unless c is even.* This assumption holds for many spaces of interest — particularly for complex algebraic varieties, where X_c is a complex manifold, and therefore of even real dimension.

⟨**10.2**⟩ **Allowable cycles and cobordisms.** Now suppose $G \circlearrowright X$ is a group of finite type acting on X. The group G will necessarily preserve the decomposition $X = \bigcup_c X_c$ into $\mathbf{H} \circlearrowright X$ orbits, since G is a subgroup of $\mathbf{H}$. An *allowable* i-cycle is a diagram

$$P \xleftarrow{\pi} E \xrightarrow{\sigma} X$$

as in the definition of an equivariant i-cycle §1.6.1, that satisfies the *allowability condition*

$$\operatorname{codim} \sigma^{-1}(X_c) < \frac{c}{2}$$

where $\operatorname{codim} \sigma^{-1}(X_c)$ is the codimension of $\sigma^{-1}(X_c)$ in E .

Similarly, an *allowable* cobordism between two allowable equivariant i-cycles P_1 and P_2 is a diagram

$$\begin{array}{ccccc} C & \xleftarrow{\pi} & E & \xrightarrow{\sigma} & X \\ \text{inclusion as boundary B} \uparrow & & \text{inclusion} \uparrow & & \uparrow = \\ P_1 - P_2 & \xleftarrow{\pi_1, \pi_2} & E_1 \cup E_2 & \xrightarrow{\sigma_1, \sigma_2} & X \end{array}$$

as in the definition of a cobordism §1.6.3 satisfying the same allowability condition

$$\operatorname{codim} \sigma^{-1}(X_c) < \frac{c}{2}$$

⟨**10.3**⟩ **Definition.** The *equivariant intersection homology* $IH_i(G \circlearrowright X)$ is the allowable i-cycles modulo allowable cobordism. The ordinary (non-equivariant) intersection homology $IH_i(X)$ is $IH_i(1 \circlearrowright X)$, where 1 is the one element group.

⟨**10.4**⟩ **Remark.** The allowability condition, and particularly the appearance of $\frac{c}{2}$, is unintuitive at first. As usual, the solution is to look at lots of examples.

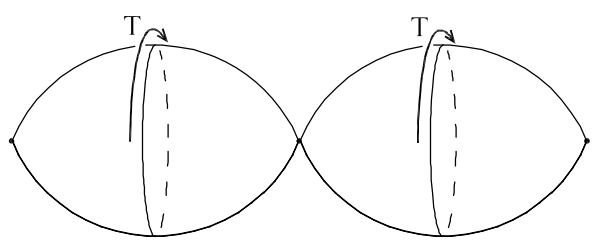
⟨**10.5**⟩ As before, we will take real coefficients by tensoring with $\mathbb{R}$. The intersection cohomology $IH^i(G \circlearrowright X)$ is the vector space dual of the intersection homology. It is no longer a ring, but $IH^*(G \circlearrowright X)$ is a graded module over the graded ring $H^*(G \circlearrowright X)$. We get the ordinary intersection homology by taking the group to be 1: $IH_i(X) = IH_i(1 \circlearrowright X)$.

⟨**10.6**⟩ **Caveat.** This definition should give the right answer ([**10**], [**23**]) but it is unproved at present. What can be proved to give the right answer at the moment differs from this in two respects that are conceptually minor but technically major:

(1) The spaces $S_c \subset X$ are taken as strata in some appropriate stratification theory. (There are various possible choices.) The strata are provably a finer decomposition than the decomposition by orbits of $\mathbf{H} \circlearrowright X$.
(2) The space X and the cycles P are taken to have extra structure like a subanalytic structure or a piecewise linear structure, and the maps preserve this structure. Nevertheless, the resulting groups are provably homeomorphism invariants.

A precise statement, discovered jointly with T. Braden, is in [**13**].

⟨**10.7**⟩ **Exercise.** Consider the example where X is two spheres with the North pole of one glued to the South pole of the other, and T is circle group which rotates both spheres simultaneously.



The various types of homology groups of X that have been defined in this lecture are given in the following table:

Type of homology	i odd	$i = 0$	$i = 2$	i even, $i \geq 4$
$H_i(X)$	0	$\mathbb{R}$	$\mathbb{R}^2$	0
$H_i(T \circlearrowright X)$	0	$\mathbb{R}$	$\mathbb{R}^3$	$\mathbb{R}^3$
$IH_i(X)$	0	$\mathbb{R}^2$	$\mathbb{R}^2$	0
$IH_i(T \circlearrowright X)$	0	$\mathbb{R}^2$	$\mathbb{R}^4$	$\mathbb{R}^4$

Give explicit cycles generating these groups, and give plausibility arguments that these calculations are correct. (The hardest ones are the 4 generators of $IH_i(T \circlearrowright X)$ for $i \geq 2$ and even. Each generator of $IH_2(T \circlearrowright X)$ may be represented as a 3-sphere with a free T action, mapped into X is such a way that the inverse image of a fixed point in X is a single T orbit. This has codimension 2 in the 3-sphere, so it satisfies the allowability condition 1.10.2.)

This example actually comes up. It is a generalized Schubert variety §4.7.3, and it is a Springer variety §5.4.6. The calculation methods of Lectures 3, 4, and 5 all apply to this example.

LECTURE 2
Moment Graphs

(Geometry of Orbits)

⟨**0.1**⟩ In this Lecture we will consider a space X with an action of a torus T satisfying certain conditions. We will associate to $T \circlearrowright X$ a linear graph called its moment graph. (Or more accurately, we will associate to $T \circlearrowright X$ an equivalence class of linear graphs). It turns out that interesting torus actions give rise to beautiful linear graphs. This is perhaps the first indication that the moment graph construction is a natural one to consider. We will construct the moment graphs of several classes of spaces: projective spaces, quadric hypersurfaces, Grassmannians, Lagrangian Grassmannians, flag manifolds, and toric varieties.

⟨**0.2**⟩ **Notation.** Our torus is $\mathbb{T}/L$ where $\mathbb{T}$ is an n-dimensional real vector space and L is a lattice, as in §0.1. We denote by t an element of $\mathbb{T}$ and by $\bar{t}$ its coset in $T = \mathbb{T}/L$. We reserve the symbol $\mathbb{V}$ for the dual vector space to $\mathbb{T}$ so we have an evaluation map

$$\begin{array}{ccccc} \mathbb{T} & \times & \mathbb{V} & \longrightarrow & \mathbb{R} \\ t & \times & v & \mapsto & < t, v > . \end{array}$$

In most of our examples, $\mathbb{T}$ will naturally be $\mathbb{R}^n$ and L will be $\mathbb{Z}^n$. In this case, $\mathbb{V}$ is also naturally $\mathbb{R}^n$. We write $t = (t_1, \ldots, t_n)$ and $v = (v_1, \ldots, v_n)$ so

$$< t, v >= t_1 v_1 + \cdots + t_n v_n$$

We can also think of T as $(S^1)^n$. We will denote an element of $(S_1)^n$ by $z = (z_1, \ldots, z_n)$, so $z_j = e^{2\pi i t_j}$.

2.1. Assumptions on the Action of T on X

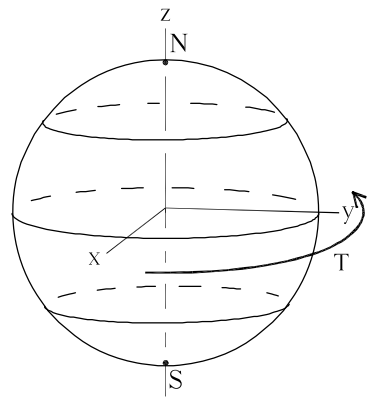
The action of the n-torus T on X decomposes it into orbits Tx each of which has a dimension that is at most n. (In fact, each orbit Tx is a k-torus where $k \leq n$.)

⟨**1.1**⟩ **Definition.** The *k-skeleton* of $T \circlearrowright X$ is the union of all the orbits of X that are of dimension at most k.

For example, the 0-skeleton is the union of the fixed points $\{x \in X \mid tx = x \text{ for all } t \in T\}$. The n-skeleton is X itself. The k-skeleton is preserved as a set by the action of T on X, so the k-skeleton is itself a space with a T action.

⟨**1.2**⟩ **Definition.** A *balloon* $T \circlearrowright B$ is a 2-sphere $B = \{x, y, z \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1\}$ together with a linear function $D(B) : \mathbb{T} \longrightarrow \mathbb{R}$ taking L to $\mathbb{Z}$ such that $T = \mathbb{T}/L$ acts on B as follows: If $t \in \mathbb{T}$, then the projection $\bar{t}$ of t in T rotates the sphere about the z axis by an angle of $2\pi D_B(t)$, i.e.

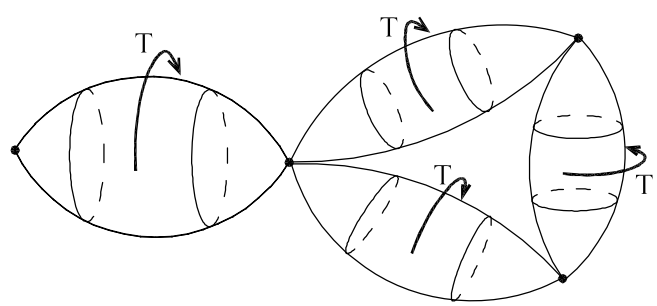
$$\bar{t} = \begin{bmatrix} \cos 2\pi D_B(t) & -\sin 2\pi D_B(t) & 0 \\ \sin 2\pi D_B(t) & \cos 2\pi D_B(t) & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$



⟨**1.3**⟩ **Exercise.** Suppose T acts on S^2 so that the orbits are the North pole $N = (0, 0, 1)$, the South pole $S = (0, 0, -1)$ and the circles of constant latitude $z = c$ where c is a constant between -1 and 1. Show that $T \circlearrowright S^2$ is equivalent to a balloon.

⟨**1.4**⟩ **Exercise*.** Show that any action of a torus T on S^2 is either a balloon or else it's the trivial action, where every point of T leaves every point of S^2 fixed.

⟨**1.5**⟩ **Definition.** A *balloon sculpture* is a space with a torus action such that is a finite union of balloons B_j such that any two balloons are either disjoint or intersect a fixed point of the torus action.



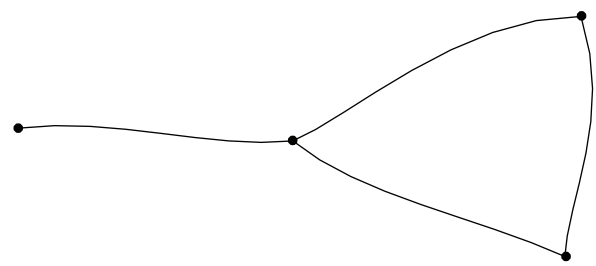
A balloon sculpture Y

⟨**1.6**⟩ **Assumption.** *We will assume that the* 1*-skeleton of* $T \circlearrowright X$ *is a balloon sculpture* (until Lecture 5). ⟨**1.7**⟩ **Exercise.** If X is compact, show that this assumption is equivalent to the assumption that the T acts with finitely many fixed points, and the 1-skeleton with the 0-skeleton deleted is a 2-dimensional manifold.

2.2. The Moment Graph

⟨**2.1**⟩ If $T \circlearrowright Y$ is a balloon sculpture, then the quotient space Y/T is a graph whose vertices correspond to the fixed points of T, and whose edges correspond to the

balloons. The graph Y/T is obtained by collapsing each balloon down to a line segment.



The graph Y/T

$\langle$**2.2**$\rangle$ We want to enhance the graph Y/T to a linear graph in a vector space $\mathbb{V}$, as in §0.2. To do this, we need direction data §0.2.5: To each edge of the graph, we need to give a direction $D \in \mathbb{V}$. The idea is to use D_{B_j} as the direction. It is a vector in the vector space $\mathrm{Hom}(\mathbb{T}, \mathbb{R})$, the space of all of linear maps $\mathbb{T} \longrightarrow \mathbb{R}$, i.e. $\mathbb{V}$ is the dual vector space $\mathbb{T}^*$. We will call D_{B_j} the *direction vector* of the balloon B_j or of the corresponding edge of the graph Y/T

Definition [**13**]. The moment graph of $T \circlearrowright X$ is the linear graph in $\mathbb{V} = \mathrm{Hom}(\mathbb{T}, \mathbb{R})$ obtained from the graph Y/T, where Y is the 1-skeleton of $T \circlearrowright X$, by associating the direction vector D_{B_j} to the edge corresponding to the balloon B_j. We notate the moment graph $\mathcal{G}(T \circlearrowright X)$.

$\langle$**2.3**$\rangle$ **Existence and uniqueness.** The moment graph can be defined only up to equivalence §0.2.4 because we have specified it by direction data.

The direction data for the moment graph is well defined. By changing the identification of B_j with S^2, the actual direction vector D_{B_j} could be changed, but the direction in $\mathbb{V}$ would still be the same, by §2.2.6.

A moment graph will not always exist (§2.2.7). Remarkably, it does exist for the most interesting examples. We will construct it for many examples in this lecture. The general phenomenon of existence of the moment graph will be discussed in §2.9.

$\langle$**2.4**$\rangle$ **Notation in $\mathbb{V}$.** When $\mathbb{T}$ is identified with $\mathbb{R}^n$, we will identify $\mathbb{V} = \mathbb{T}^* = (\mathbb{R}^n)^*$ as well. We denote the standard basis for $\mathbb{V}$ by $e_1, \ldots, e_n$, where e_i is the point where $v_i = 1$ and all of the other v_j are 0. Considered a linear map $\mathbb{T} \longrightarrow \mathbb{R}$, we have $<(t_1, \ldots, t_j, \ldots, t_n), e_j> = t_j$.

$\langle$**2.5**$\rangle$ **Exercise.** Suppose that $T \circlearrowright B$ is a balloon and that $D'_B : \mathbb{T} \longrightarrow \mathbb{R}$ is a nonzero linear map such that if $D'_B(t) = 0$ then $\bar{t}$ fixes every point of B. Then D'_B is some scalar multiple of D_B, so D'_B and D_B determine the same direction in $\mathbb{V}$.

$\langle$**2.6**$\rangle$ **Exercise.** Suppose that B is displayed as a balloon in two different ways, i.e. there are two different homeomorphisms equivariant homeomorphisms from B to a sphere as in the definition of a balloon. Suppose that D_B and D'_B are the corresponding functions from $\mathbb{T} \longrightarrow \mathbb{R}$. Show that D_B and D'_B determine the same direction. (You can use exercise 2.2.5.) In fact, $D'_B = \pm D_B$.

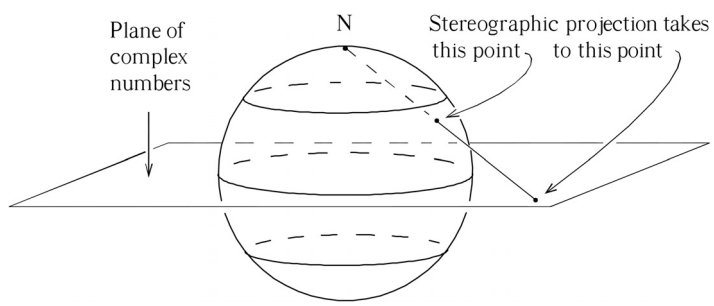
⟨**2.7**⟩ **Exercise.** Construct an example $T \circlearrowright X$ where the moment graph does not exist. (Hint: Take X to be the balloon sculpture whose direction data coincides with that of exercise 0.2.8.)

2.3. Complex Projective Line and the Line Segment

Most of the rest of this Lecture will be devoted to explicit computation of moment graphs for specific torus actions $T \circlearrowright X$.

⟨**3.1**⟩ **Definition.** The *complex projective k-space* $\mathbb{P}^k$ is the quotient space $\mathbb{C}^{k+1} - \{0\}/\mathbb{C}^\times$ where $\mathbb{C}^\times$ is multiplicative group of the complex numbers acting on $\mathbb{C}^n$ by scalar multiplication. A point in $\mathbb{P}^k$ is denoted by *homogeneous coordinates* $(x_1 : x_2 : \cdots : x_{k+1})$ where the x_j are complex numbers, not all of which are zero, and $(\lambda x_1 : \lambda x_2 : \cdots : \lambda x_{k+1})$ represents the same point in $\mathbb{P}^K$ as $(x_1 : x_2 : \cdots : x_{k+1})$ if λ is a nonzero complex number.

⟨**3.2**⟩ **The projective line.** Complex 1-space is called the projective line. Topologically, it is a 2-sphere, called the *Riemann sphere* in complex analysis. We may identify $\mathbb{P}^1 - (0 : 1)$ with the complex plane $\mathbb{C}$ by sending $(x_1 : x_2)$ to x_2/x_1. We may identify the 2-sphere minus the north pole N with the complex plane $\mathbb{C}$ by stereographic projection.



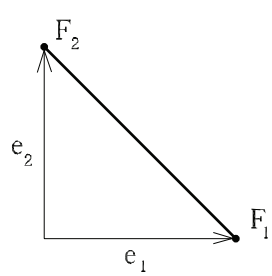
Stereographic projection takes rotation about the z axis to rotation in the complex plane about 0, i.e. to multiplication by a complex number on the unit circle S^1.

Now, suppose that the 2 torus acts on the projective line by the formula

$$z(x_1 : x_2) = (z_1 x_1 : z_2 x_2).$$

Proposition. With this action, $\mathbb{P}^1$ is a balloon B where the direction vector D_B is $e_1 - e_2$ in $\mathbb{V}$. (Here e_i is the standard basis as in §2.2.4.)

This proposition gives us the moment graph of $\mathbb{P}^1$. We must send the two vertices corresponding to the two fixed points $F_1 = (1 : 0)$ and $F_2 = (0 : 1)$ to points p_1 and p_2 in $\mathbb{V} = \mathbb{R}^2$ so that the straight line from p_2 to p_1 is parallel to the direction vector $e_1 - e_2$. An obvious choice is $p_1 = e_1$ and $p_2 = e_2$, so the moment graph is a line segment between e_1 and e_2, as in this picture.



$\mathcal{G}(T \circlearrowright \mathbb{CP}^1)$

Proof. We compute the action of T in $\mathbb{P} - (0:1) = \mathbb{C}$

$$\frac{z_2x_2}{z_1x_1} = \frac{e^{2\pi it_2}x_2}{e^{2\pi it_1}x_1} = e^{2\pi i(t_2-t_1)}\left(\frac{x_2}{x_1}\right) = e^{2\pi i(e_2-e_1)(t_1,t_2)}\left(\frac{x_1}{x_2}\right) = e^{2\pi i(e_2-e_1)t}\left(\frac{x_1}{x_2}\right)$$

which means that $\bar{t}$ gives a rotation of $(e_1 - e_2)(t)$ Alternatively, the proposition can be seen by §2.2.5: If $(e_2 - e_1)(t_1, t_2) = 0$, then $t_1 = t_2$ so $(e^{2\pi it_1}x_1 : e^{2\pi it_2}x_2)$ is the same point as $(x_1; x_2)$ because both homogeneous coordinates are multiplied by the same number.

⟨**3.3**⟩ **Exercise.** More generally, show that if an n-torus T acts on the projective line by $\bar{t}(x_1 : x_2) = (e^{2\pi i\phi_1(t)}x_1 : e^{2\pi i\phi_2(t)}x_2)$ for $\phi_1, \phi_2 : \mathbb{T} \longrightarrow \mathbb{R}$, and $\phi_1 \neq \phi_2$, then $\mathbb{P}^1$ is a balloon with $D_B = \phi_1 - \phi_2$.

⟨**3.4**⟩ Almost all of the balloons in the 1-skeleta of the $T \circlearrowright X$ we will consider in this Lecture are themselves a copy of $\mathbb{P}^1$ embedded in the space X. So the analysis of this section will be used repeatedly in what follows.

2.4. Projective $(n-1)$-Space and the Simplex

We generalize the discussion $\mathbb{P}^1$ above. The "standard" action of the n-torus on $\mathbb{P}^{n-1}$ is

$$z(x_1 : \cdots : x_n) = (z_1x_1 : \cdots : z_nx_n)$$

where $z = (z_1, \ldots, z_n)$ and $z_j \in S^1 \subset \mathbb{C}$.

⟨**4.1**⟩ **The fixed points** are the n points F_i where all the homogeneous coordinates are zero except the i-th one.

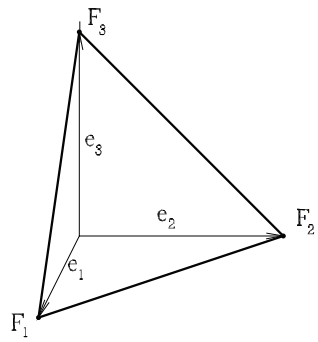
⟨**4.2**⟩ **The balloons.** Let i and j be any pair of distinct indices $1 \leq i, j \leq n$. Then the balloon B_{ij} is where all the homogeneous coordinates are zero except the i-th one or the j-th one. It connects the fixed points F_i and F_j.

⟨**4.3**⟩ **Remark: balloons and $\mathbb{C}^*$ orbits.** The action of $T = S^n_1$ on $\mathbb{P}^{n-1}$ extends to an action of $T_{\mathbb{C}} = (\mathbb{C}^*)^n$ where $\mathbb{C}^*$ is the nonzero complex numbers considered as a group under multiplication. The action of $T_{\mathbb{C}}$ is given by the same formula $z(x_1 : \cdots : x_n) = (z_1x_1 : \cdots : z_nx_n)$ where $z_j \in \mathbb{C}^*$. Each balloon consists of three $T_{\mathbb{C}}$ orbits: the two fixed points and one more, of complex dimension 1. So the classification of balloons is the same as the classification of complex one dimensional orbits of the $T_{\mathbb{C}}$ action.

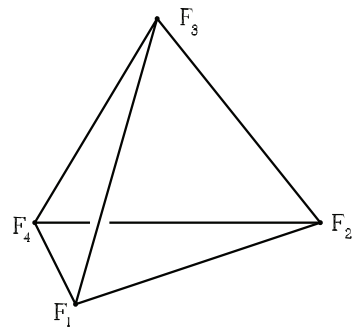
⟨**4.4**⟩ **The direction vector** of $B_{i,j}$ is $e_i - e_j$.

⟨**4.5**⟩ **The moment graph.** If we send F_i to e_i, then the straight line connecting F_i to F_j is parallel to $e_i - e_j$. So the moment graph $\mathcal{G}(T \circlearrowright \mathbb{P}^{n-1})$ of $\mathbb{P}^{n-1}$ is the 1-skeleton of the $(n-1)$-simplex Δ_{n-1}. (The $(n-1)$-simplex Δ_{n-1} is the convex hull of the basis vectors $e_1, \ldots, e_n$, or alternatively

$$\Delta_{n-1} = \{(v_1, \ldots v_n) \mid v_1 + \cdots + v_n = 1 \text{ and } v_j \geq 0\}).$$



The 2 dimensional simplex



The 3 dimensional simplex pictured in a 3 dimensional subspace of the 4 dimensional space V

⟨**4.6**⟩ **The proofs.** The points F_i are fixed by the equivalence relation on homogeneous coordinates. The sets $B_{i,j}$ are projective lines, so they are spheres. The action of T on $B_{i,j}$ is very similar to the action of §2.3.2, so the direction vectors can be computed in a similar way. Alternatively, §2.3.3 can be used directly. The only real challenge is to show that the 1-skeleton is the union of the balloons B_{ij}. This follows from the following exercise.

⟨**4.7**⟩ **Exercise.** Show that if $x \in \mathbb{P}^n$ has k nonzero homogeneous coordinates, then the dimension of the orbit Tx is $k-1$.

2.5. Quadric Hypersurfaces and the Cross-Polytope

⟨**5.1**⟩ **Definition of $T \circlearrowright X$.** The $(2n-2)$-*dimensional quadric hypersurface* Q_{2n-2} is the subset of $\mathbb{P}^{2n-1}$ with homogeneous coordinates $(x_1 : \cdots : x_n : y_1 : \cdots : y_n)$ cut out by the equation $x_1y_1 + x_2y_2 + \cdots + x_ny_n = 0$. (This makes sense because if $(x_1 : \cdots : x_n : y_1 : \cdots : y_n)$ satisfies the equation, then so will $(\lambda x_1 : \cdots : \lambda x_n : \lambda y_1 : \cdots : \lambda y_n)$.)

The n-torus T acts on $X = Q_{2n-2}$ by the formula

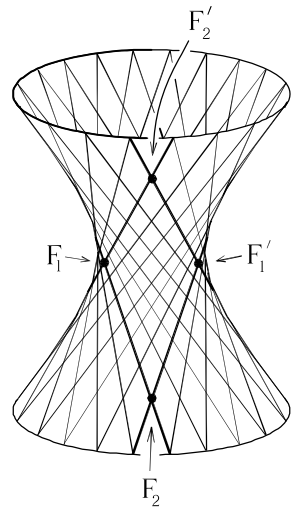
$$z(x_1 : \cdots : x_n : y_1 : \cdots : y_n) = (z_1x_1 : \cdots : z_nx_n : z_1^{-1}y_n : \cdots : z_n^{-1}y_n)$$

You can check that this formula is compatible with the equivalence relation on homogeneous coordinates defining $\mathbb{P}^{2n-1}$ and that it preserves the equation for the hypersurface Q_{2n-2}.

⟨**5.2**⟩ **The fixed points** are the points where exactly one homogeneous coordinate is nonzero. Let's call F_i the point where x_i is nonzero and F_i' the point where y_i is nonzero.

⟨**5.3**⟩ **The balloons.** For every pair of homogeneous coordinates except the n pairs $\{x_i, y_i\}$, the points in X where only that pair of homogeneous coordinates is nonzero is a projective line. These are the balloons. So there is a balloon connecting any pair of fixed points except the n pairs with the same index, F_i and F_i'. So the number of balloons is $\binom{2n}{2} - n = \frac{(2n)(2n-1)}{2} - n$, where $\binom{2n}{2}$ is the number of 2 element subsets of a $2n$ element set.

⟨**5.4**⟩ **Real picture for** $n = 2$**.** We can't draw any interesting complex quadrics, because their dimensions are too large. However, we can draw the real quadric $Q^{\mathbb{R}}_{2n-2}$ for $n = 2$. It is the surface $x_1y_1 + x_2y_2 = 0$ in $\mathbb{RP}^3$. The real projective space $\mathbb{RP}^3$ contains the real affine space $\mathbb{R}^3$ as a dense subspace. The intersection of the quadric with $\mathbb{R}^3$ is pictured at the right. It is doubly ruled surface. The four fixed points F_1, F_2, F_1', F_2' lie in $Q^{\mathbb{R}}_2 \subset Q_2$. Each of the four balloons in Q_2 is a $\mathbb{CP}^1$, it intersects $\mathbb{RP}^3$ in a $\mathbb{RP}^1$, which intersects $\mathbb{R}^3$ in a straight line. These 4 points and 4 lines are shown on the picture. Just as the balloons are the closures of the complex 1-dimensional $T_{\mathbb{C}}$ orbits §2.4.3, these 4 real lines are the closures of the real 1-dimensional $T_{\mathbb{R}}$ orbits, where $T_{\mathbb{R}} = (\mathbb{R}^*)^2$ acts by the same formulas as in the complex case.



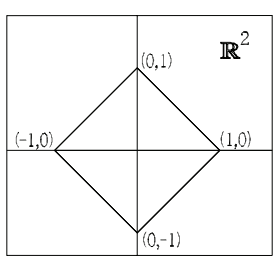
⟨**5.5**⟩ **The direction vectors.** For the balloon B joining F_i and F_j, the direction vector D_B is $e_i - e_j$. For the balloon B joining F_i' and F_j', D_B is $-e_i + e_j$. For the balloon B joining F_i and F_j' for $i \neq j$, D_B is $e_i + e_j$.

⟨**5.6**⟩ **Exercise.** Verify this calculation of direction vectors. Hint: use §2.3.3.

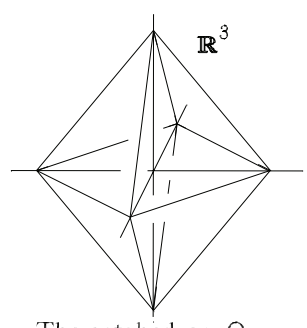
⟨**5.7**⟩ **The** n**-dimensional cross-polytope** O_n is the polyhedron in $\mathbb{V} = \mathbb{R}^n$ defined by the relation that the sum of the absolute values of the coordinates is at most 1.

$$O_n = \{(v_1, \ldots, v_n) \in \mathbb{R}^n \mid |v_1| + \cdots + |v_n| \leq 1\}$$

The cross-polytopes in dimensions 2 and 3 are the square and the octahedron.



The square, O_2



The octahedron, O_3

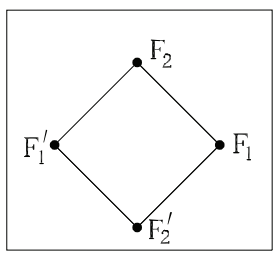
The vertices of the cross-polytope O_n are the $2n$ points $\{e_1, \ldots, e_n; -e_1, \ldots, -e_n\}$ where the e_i are the standard basis vectors for $\mathbb{R}^n$. The convex polyhedron O_n can be defined alternatively as the convex hull of this set of vertices.

⟨**5.8**⟩ **Exercise.** Show that there is an edge between any pair of vertices except for the n pairs $\{e_i, -e_i\}$ so that the number of edges is $\binom{2n}{2} - n = \frac{(2n)(2n-1)}{2} - n$, where $\binom{2n}{2}$ is the number of 2 element subsets of a $2n$ element set.

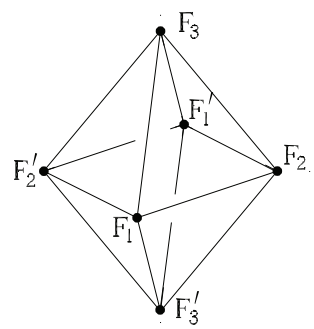
⟨**5.9**⟩ **Exercise.** Show that the number of faces of dimension i in the cross-polytope O_n is the coefficient of q^{i+1} in the polynomial $(1+2q)^n$.

⟨**5.10**⟩ **The moment graph** of the $(2n-2)$-dimensional quadric hypersurface X is the 1-skeleton of the cross-polytope O_n.

More explicitly, we define a map from the set of fixed points to $\mathbb{V} = \mathbb{R}^n$ by sending F_i to e_i and sending F_i' to $-e_i$. Then for every pair of fixed points connected by a balloon, the direction vector of that balloon is parallel to the line connecting the corresponding points in $\mathbb{V}$.

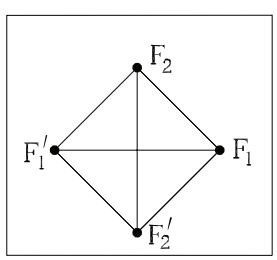


Moment graph $\mathcal{G}(T \circlearrowright Q_2)$

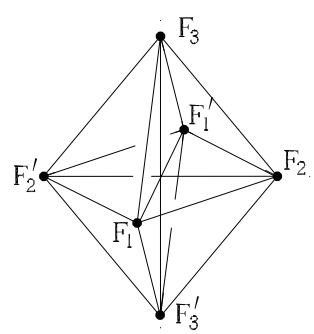


Moment graph $\mathcal{G}(T \circlearrowright Q_4)$

⟨**5.11**⟩ **Odd dimensional quadric hypersurfaces.** The $(2n-1)$-*dimensional quadric hypersurface* Q_{2n-1} is the subset of $\mathbb{P}^{2n}$ with homogeneous coordinates $(w : x_1 : \cdots : x_n : y_1 : \cdots : y_n)$ cut out by the equation $w^2 + x_1y_1 + x_2y_2 + \cdots + x_ny_n = 0$. The n-torus T acts on the x and the y coordinates as before, and it acts trivially on w. So it contains the $(2n-2)$-dimensional quadric hypersurface Q_{2n-2} as the T invariant subspace where $w = 0$. The fixed points of Q_{2n-1} are the same as the fixed points of Q_{2n-2}, but there are n additional balloons: namely, the subspace where only x_i, y_i, and w are nonzero is a balloon connecting F_i and F_i'. The moment graph for Q_{2n-1} is the moment graph for Q_{2n-2} with n additional straight lines connecting F_i to F_i'.



Moment graph $\mathcal{G}(T \circlearrowright Q_3)$



Moment graph $\mathcal{G}(T \circlearrowright Q_5)$

(The moment graph $\mathcal{G}(T \circlearrowright Q_5)$ is the PCMI logo.)

2.6. Grassmannians and Hypersimplices

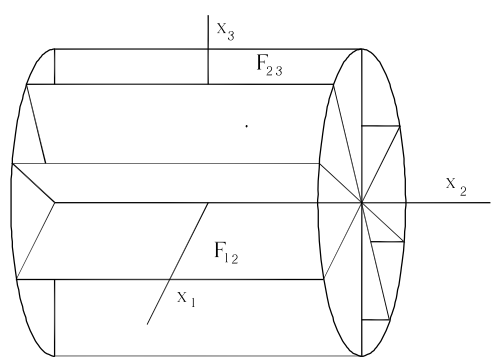
A point in projective space $\mathbb{P}^{n-1}$ represents a line through the origin in the vector space $\mathbb{C}^n$: The points in the line are the different homogeneous coordinates that represent the point in projective space. Similarly, we can make a space whose points represent subspaces of higher dimension in $\mathbb{C}^n$. This leads to various kinds of Grassmannian varieties.

$\langle$**6.1**$\rangle$ **The space** G_i^n is the Grassmannian variety whose points are the i-dimensional subspaces of the n-dimensional complex vector space $\mathbb{C}^n$. The n-torus acts on it through its action on $\mathbb{C}^n$: $z(x_1, \ldots, x_n) = (z_1x_1, \ldots, z_nx_n)$.

$\langle$**6.2**$\rangle$ **The fixed points.** Suppose that S is a subset of $\{1, 2, \ldots, n\}$. Let P_S be the coordinate plane corresponding to S, i.e. P_S is the $|S|$-plane defined by the condition that only the coordinates $\{x_j \mid j \in S\}$ can be nonzero. Here $|S|$ is the number of elements of S. The fixed points in G_i^n are the planes P_S where $|S| = i$. We denote P_S by F_S when thinking of it as a fixed point in G_i^n.

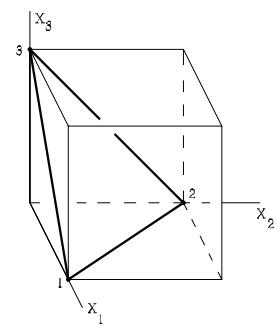
$\langle$**6.3**$\rangle$ **The balloons.** Suppose S' is obtained from S by deleting the number j and adding number k, for $j \neq k$. Then the set of i-dimensional subspaces that contain $P_{S\cap S'}$ and are contained in $P_{S\cup S'}$ is a balloon connecting F_S and $F_{S'}$. The **direction vector** of this balloon is $e_j - e_k$.

$\langle$**6.4**$\rangle$ Here is a picture of the planes in the balloon connecting $F_{\{1,2\}}$, and $F_{\{2,3\}}$ in G_2^3. Since we can't visualize $\mathbb{C}^3$, we're using a real picture, i.e. real planes in the real vector space $\mathbb{R}^3$ instead of complex planes in the complex vector space $\mathbb{C}^3$.

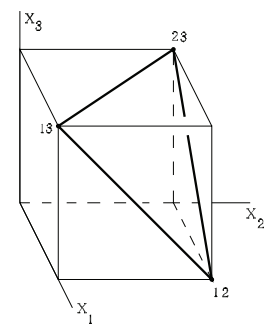


Points in a balloon in G_2^3

$\langle$**6.5**$\rangle$ **The hypersimplex** Δ_i^n is the intersection of the n-cube $[0,1]^n \subset \mathbb{R}^n = \mathbb{V}$ with the plane $v_1 + v_2 + \cdots + v_n = i$. It is a convex polyhedron with vertices $\nu_S = \Sigma_{j\in S} e_j$ where S is an i element subset of $\{1, \ldots, n\}$. The vertices ν_S and $\nu_{S'}$ are connected by an edge if S' is obtained from S by deleting the number i and adding number j, for $i \neq j$. Then the edge is parallel to $e_i - e_j$.



The hypersimplex Δ_1^3

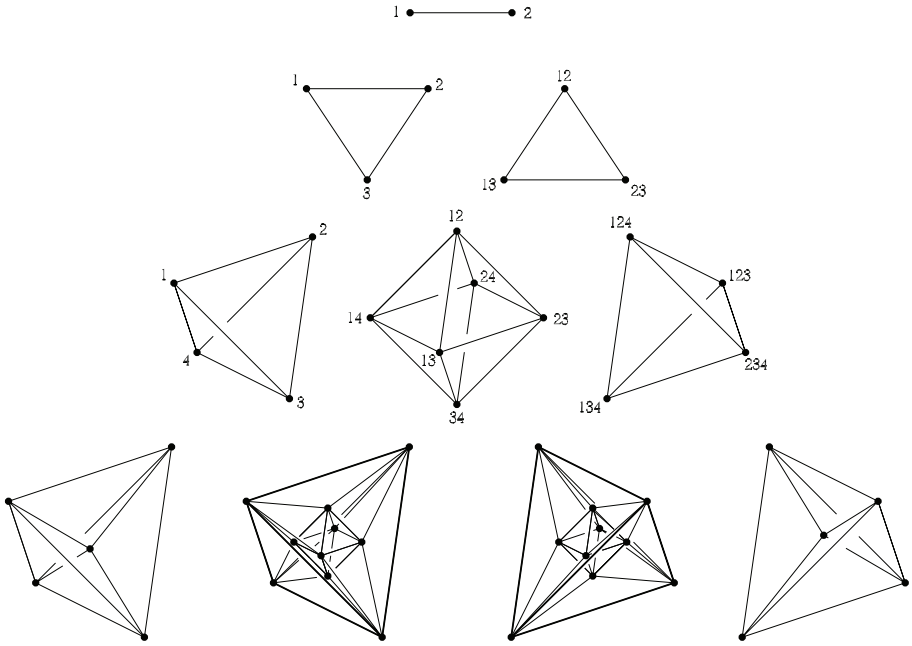


The hypersimplex Δ_2^3

⟨**6.6**⟩ **The moment graph** of the Grassmannian G_i^n is the 1-skeleton of the hypersimplex Δ_i^n. There is a rich theory surrounding hypersimplices and Grassmannians [**12**], [**11**], [**6**].

⟨**6.7**⟩ **Exercise.** Show that the hypersimplices can be arranged in a polyhedral version of Pascal's triangle where the faces of each polyhedron are isomorphic to one of the two polyhedra lying above it.

For dimension up to 4, this is illustrated in the following picture. The labels of vertices show which coordinates are 1 (or equivalently, which coordinate axes are in the corresponding plane representing a T fixed point of the Grassmannian). The figures in last line, representing 4-dimensional hypersimplices, are projections to $\mathbb{R}^3$ called Schlegel diagrams. Note that the polyhedra on the two upper edges of the picture are ordinary simplices.



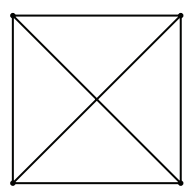
Pascal's triangle of hypersimplices

⟨**6.8**⟩ **The Lagrangian Grassmannian and the cube.** Consider $\mathbb{C}^{2n}$ with coordinates $x_1, \ldots, x_n, y_1, \ldots, y_n$ and the alternating form $\Sigma_i x_i y_i' - x_i' y_i$. The *Lagrangian Grassmannian* L_n is the subvariety of the Grassmannian G_n^{2n} consisting of n planes on which this alternating form vanishes identically. The torus T acts on L_n by through its action on $\mathbb{C}^{2n}$ by the formula

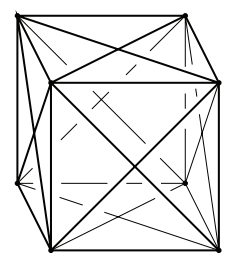
$$z(x_1, \ldots, x_n, y_1, \ldots, y_n) = (z_1 x_1, \ldots, z_n x_n, z_1^{-1} y_n, \ldots, z_n^{-1} y_n).$$

The fixed points F_S are the coordinate planes that lie in L_n. For any subset $S \subset \{1, \ldots, n\}$, F_S is the plane whose nonzero coordinates are the x_i for $i \in S$ and the y_i for $i \notin S$. There are 2^n of them.

Exercise. Show that the vertices of the moment graph of L_n are the vertices of the n-cube $[0,1]^n \subset \mathbb{V}$ and the edges of the moment graph are the edges of the cube together with the diagonals of the 2-dimensional faces.



The moment graph $\mathcal{G}(T \circlearrowright L_2)$



The moment graph $\mathcal{G}(T \circlearrowright L_3)$

2.7. The Flag Manifold and the Permutahedron

⟨**7.1**⟩ **The flag manifold.** Consider $\mathbb{C}^n$ as $\mathbb{R}^{2n}$ in the usual way, with the standard real dot product $\cdot_{\mathbb{R}}$ on it. A point in the *flag manifold* F_n is an ordered set of n mutually orthogonal complex lines through the origin in $\mathbb{C}^n$. Here mutually orthogonal means that if x is in one of the complex lines and y is in another one, then $x \cdot_{\mathbb{R}} y = 0$. (This is the same as their being orthogonal with respect to the standard Hermitian inner product.)

The n-torus T acts on F_n through its standard action on $\mathbb{C}^n$. This action preserves the orthogonality condition.

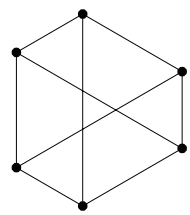
⟨**7.2**⟩ **Fixed points.** A point is fixed if the n mutually orthogonal lines coincide with the complex coordinate axes in $\mathbb{C}^n$. There are $n!$ of them, one for each ordering of the coordinate axes.

⟨**7.3**⟩ **The balloons.** Pick two coordinate axes of $\mathbb{C}^n$, say the x_i axis and the x_j axis. A balloon is the set where all but two of the mutually orthogonal complex lines are required to lie on a coordinate axis that is not the x_i axis or the x_j axis. The remaining two complex lines are free to wander (staying orthogonal to each other) in the 2-dimensional plane spanned by the x_i axis and the x_j axis.

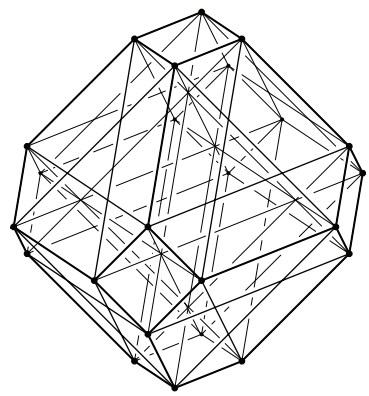
⟨**7.4**⟩ **The permutahedron.** Fix n distinct real numbers $a_1, \ldots, a_n$. The permutahedron is the convex hull in $\mathbb{R}^n$ of the $n!$ points $(a_{\sigma(1)}, \ldots a_{\sigma(n)})$ where σ runs through the $n!$ permutations of the numbers $\{1, \ldots, n\}$. It is an $(n-1)$-dimensional

polytope because it lies in a hyperplane in $\mathbb{R}^n$ where the sum of the coordinates is constant since the sum of the coordinates of the vertices is constant.

⟨7.5⟩ The moment graph of the flag manifold. The vertices of the moment graph for F_n are the vertices of the permutahedron. Two vertices are connected by an edge if one is a reflection of the other in one of the $\binom{n}{2}$ hyperplanes defined by an equation $v_i = v_j$.



Moment graph
$\mathcal{G}(T \circlearrowright F_3)$



Moment graph $\mathcal{G}(T \circlearrowright F_4)$

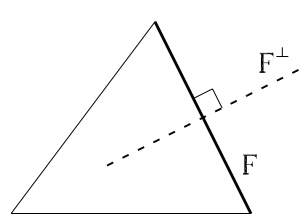
2.8. Toric Varieties and Convex Polyhedra

So far, we have begun with a space with a torus action $T \circlearrowright X$ and we have computed the moment graph $\mathcal{G}(T \circlearrowright X)$. In this section we go the other way. We give ourselves a rational convex polyhedron in a vector space $\mathbb{V}$, and we associate to it a space with a torus action $T \circlearrowright T(P)$ called the toric variety associated to P. The moment graph of $T \circlearrowright T(\mathbb{V})$ is the 1-skeleton of P.

We recall our notational conventions: $\mathbb{V}$ is the real vector space $\mathbb{R}^n$, $\mathbb{T}$ is its dual vector space, also $\mathbb{R}^n$, and L is the lattice $\mathbb{Z}^n \subset \mathbb{T}$.

⟨8.1⟩ Rational polyhedra. A convex n-dimensional polyhedron P in the real vector space $\mathbb{V} = \mathbb{R}^n$ is called *rational* if all of its vertices lie in $\mathbb{Q}^n$, i.e. all the coordinates of its vertices are rational numbers.

⟨8.2⟩ $F^\perp$. Given a face F of the polyhedron $P \subset \mathbb{V}$, we will denote by $F^\perp$ the vector subspace of $\mathbb{T} = \mathbb{V}^*$ consisting of vectors which are perpendicular F, i.e. the set of all $t \in \mathbb{T}$ such that $<t, v - v'> = 0$ for every pair of points $v, v' \in F$. If F is a vertex of P, then $F^\perp = \mathbb{V}$. If F is P itself, then $F^\perp$ is just the zero vector $0 \in \mathbb{T}$.



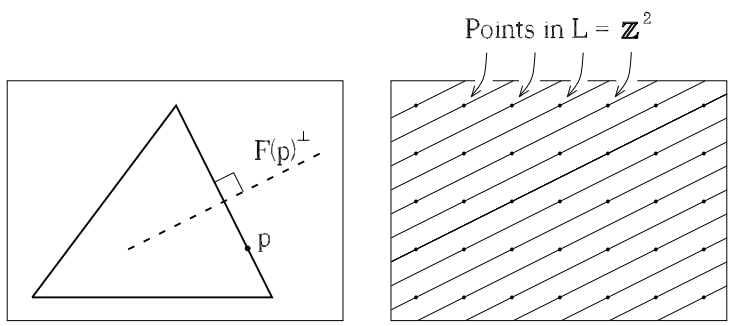
⟨**8.3**⟩ $F(p)$**.** Given a point $p \in P$ of a polyhedron, we write $F(p)$ for the smallest face of P containing p. If p is a vertex, then $F(p)$ is p itself. $F(p) = P$, if and only if p is an interior point of P.

⟨**8.4**⟩ **The toric variety** $\mathbf{T}(P)$ is the quotient space

$$\mathbf{T}(P) = \frac{P \times \mathbb{T}}{\sim}$$

where $\sim$ is the following equivalence relation:

$$(p,t) \sim (p',t') \quad \text{if and only if} \quad p = p' \ \text{ and } \ t \cong t' \bmod \ (F(p)^{\perp} + L)$$



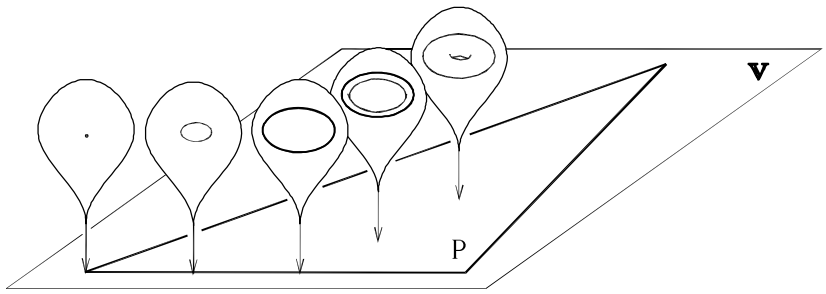
$p \in P \subset \mathbb{V}$ The subgroup $F(p)^{\perp} + L$ in $\mathbb{T}$

⟨**8.5**⟩ **The** T **action.** The torus $T = \mathbb{T}/L$ acts on the toric variety $\mathbf{T}(P)$ as follows: $\mathbb{T}$ acts on $P \times \mathbb{T}$ by vector addition $t(p,t') = (p, t'+t)$, and this action passes to an action of $\mathbb{T}$ on the quotient space $\mathbf{T}(P)$. On the quotient space, L acts trivially, since if $t \in L$, then $t(p,t') \sim (p,t')$ So the quotient group $\mathbb{T}/L$ acts on the quotient space $\mathbf{T}(P)$.

⟨**8.6**⟩ **The moment map.** There is a map $\mu : \mathbf{T}(P) \longrightarrow P$ called the *moment map* which is induced from the projection $(P \times \mathbb{T}) \longrightarrow P$. The reason the projection passes to the quotient $\mathbf{T}(P)$ is that the equivalence relation $\sim$ is compatible with this map — it identifies points only if they lie in the same fiber. In fact, there is an identification $\mathbf{T}(P)/T \approx P$, the moment map $\mathbf{T}(P) \longrightarrow \mathbf{T}(P)/T$ is the quotient map for the group action $T \circlearrowright \mathbf{T}(P)$.

Proposition. The fiber $\mu^{-1}p \subset \mathbf{T}(P)$ over a point $p \in P$ is a torus of the same dimension as the face $F(p)$.

So we can think of the toric variety $\mathbf{T}(P)$ as a family of tori over the polyhedron P whose fiber dimensions decrease as you get to smaller faces. To visualize it, here are some pictures of fibers at various points of P.

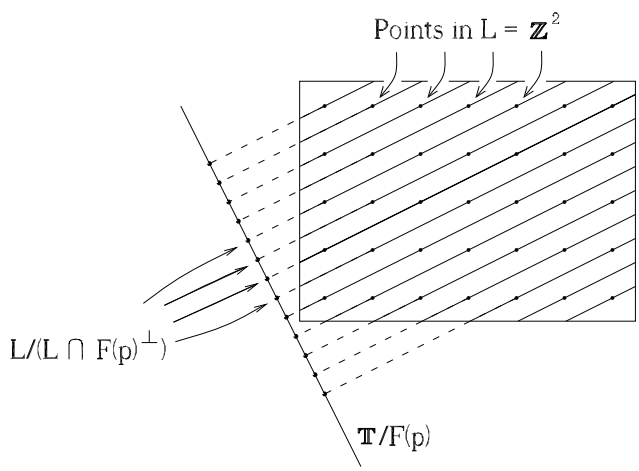


Moment map fibers $\mu^{-1}(p)$ for various $p \in P$

The torus $\mu^{-1}(p)$ over the a point p in the interior of the polyhedron P becomes thinner, looking more like a bicycle tire than a car tire, as p approaches an edge. It collapses into a circle when p reaches the edge. The circle $\mu^{-1}(p)$ over a point p in an edge becomes smaller as p approaches a vertex, and it collapses into a point when p reaches the vertex.

⟨8.7⟩ Proof of the proposition. Why is the fiber $\mu^{-1}(p)$ a torus? It is a single orbit of the action of the torus T, so it must be a torus if it is a Hausdorff space. But why is it Hausdorff? We have

$$\mu^{-1}p = \mathbb{T}/(L + F(p)^{\perp}) = \frac{\mathbb{T}/F(p)}{L/(L \cap F(p)^{\perp})}$$



We must show that $L/(L \cap F(p)^{\perp})$ is a lattice in $\mathbb{T}/F(p)$. Since this quotient space will itself be a torus: it will be the vector space $\mathbb{T}/F(p)^{\perp}$ modulo the lattice $L/(L \cap F(p)^{\perp})$

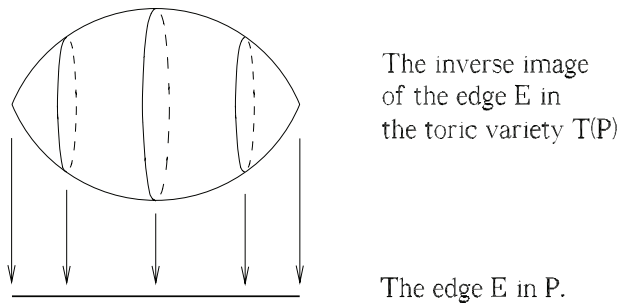
⟨8.8⟩ Exercise. Show that the following conditions are equivalent, and they all hold if the polytope P is rational:

(1) The vector space $F^{\perp}$ is a rational subspace of $\mathbb{T}$ for all faces F.
(2) The vector space $F^{\perp}$ is spanned by $F^{\perp} \cap L$ for all faces F.
(3) The quotient space $\mathbb{T}/(L + F^{\perp})$ is Hausdorff for all faces F.
(4) The subgroup $L/(L \cap F(p)^{\perp})$ is a lattice in the vector space $\mathbb{T}/F(p)$.

(5) The toric variety $(P \times \mathbb{T})/\sim$ is Hausdorff.

⟨**8.9**⟩ **Proposition.** The moment graph of the toric variety $T(P)$ is the 1-skeleton P^1 of P.

Since the dimension of the orbit $\mu^{-1}(p)$ is the dimension of $F(p)$, the 1-skeleton of $T \circlearrowright \mathbf{T}(P)$ is the inverse image of the 1-skeleton of P. The inverse image of an edge of P is a balloon.



It remains to see that the direction vector of this balloon is parallel to the edge. This follows from §2.2.5.

⟨**8.10**⟩ **Exercise.** Show that the projective $(n-1)$-space $\mathbb{P}^{n-1}$ is a toric $\mathbf{T}(P)$ where P is an $(n-1)$-simplex.

⟨**8.11**⟩ **Simple polytopes.** A polytope is *simple* if the edges coming in to every vertex, considered as vectors, are linearly independent. For example, a tetrahedron and a cube are simple, whereas an octahedron is not. All 2-dimensional polyhedra are simple.

Toric varieties of simple polytopes play a special role that will become apparent later (§3.8.1).

2.9.* Moment Maps

This is a * starred section, meaning that its prerequisites go beyond those of the other sections, and its results are not needed for the rest of what we will do. The purpose is to provide an orientation for going further in the subject, and to show how the material ties in to other mathematical ideas.

⟨**9.1**⟩ **The Lie algebra.** Our torus T is a compact Lie group. The vector space $\mathbb{T}$ is its Lie algebra. The map $\mathbb{T} \longrightarrow T$ is the exponential map of Lie theory and the lattice L is its kernel. In general, the exponential map is not a group homomorphism, but it is for the Lie group T, since T is Abelian. If $T \circlearrowright X$ and X is smooth, every $t \in \mathbb{T}$ gives rise to a vector field on X which we notate $x \mapsto t(x)$.

⟨**9.2**⟩ **Complex algebraic varieties.** All of the spaces X we have constructed in this section are complex projective algebraic varieties. Our torus $T = (S^1)^n$ is the maximal compact subgroup of a complex torus $T_{\mathbb{C}} = (\mathbb{C}^*)^n$ which is an algebraic group. The action of T extends to an algebraic action of $T_{\mathbb{C}}$. The fixed points of T are still fixed under $T_{\mathbb{C}}$. The real dimension of a T orbit Tx is the complex dimension of the $T_{\mathbb{C}}$ orbit $T_{\mathbb{C}}x$. If B is one of the balloons and N and S are the two fixed points on it, then $B - N - S$ is a single orbit of $T_{\mathbb{C}}$ of complex dimension

1. These are all the 1 complex dimensional orbits of $T_{\mathbb{C}}$. If we are given a complex algebraic action of $T_{\mathbb{C}}$ on X, then our hypothesis that the 1-skeleton of $T \circlearrowright X$ is a balloon sculpture is equivalent to the hypothesis that $T_{\mathbb{C}}$ has finitely many orbits of complex dimension 0 and 1.

⟨9.3⟩ The moment map. If X is nonsingular and projective, then it has a real symplectic form ω called the Kähler form. By Weyl's trick of averaging over T, we can choose ω to be T invariant. We define a $\mathbb{V}$-valued differential 1-form θ on X as follows: For $t \in \mathbb{T}$, let ξ_t be the corresponding vector field on X. If $\tau \in T_x X$ is a tangent vector to X at x, then $t \mapsto \omega(\tau, \xi_t(x))$ is a linear map $\mathbb{T} \longrightarrow \mathbb{R}$, so it is an element of $\mathbb{V} = \mathbb{T}^*$. That element is $\theta(\tau)$. The moment map $\mu : X \longrightarrow \mathbb{V}$ is defined by the formula

$$\mu(x) = \int_{x_0}^{x} \theta$$

where x_0 is a base point chosen in X. (If X is not connected, we define μ on each connected component separately by this procedure.)

If X is singular, we proceed a little differently. We embed X in a complex projective space in a way that is $T_{\mathbb{C}}$ equivariant. Then we take the moment map on the ambient complex projective space as constructed above, and restrict it to X.

If $T \circlearrowright X$ is a toric variety, then the moment map as defined here will coincide with the moment map from its definition as a toric variety.

⟨9.4⟩ Proposition. If $T_{\mathbb{C}}$ acts algebraically on X with finitely many orbits of dimension 0 and 1, then the moment graph $\mathcal{G}(T \circlearrowright X)$ is $\mu(X^1)$, the moment map image of its 1-skeleton. The set of vertices of the moment graph is $\mu(X^0)$. The image $\mu(X)$ of all of X will be the convex hull of the moment graph. There were several choices in constructing the moment map (choice of a Kähler form, choice of a base point). Different choices will result in different but equivalent linear graphs.

⟨9.5⟩ Exercise*. Suppose X is nonsingular and compact, and that $T_{\mathbb{C}}$ acts algebraically on X with finitely many fixed points F. Suppose further that at each fixed F, the representation $T_{\mathbb{C}} \circlearrowright T_F X$ on the tangent space has no representation of multiplicity greater than 1. Show that $T_{\mathbb{C}}$ acts with finitely many one dimensional orbits, so that the 1-skeleton of $T \circlearrowright X$ is a balloon sculpture.

LECTURE 3
The Cohomology of a Linear Graph

(Polynomial and Linear Geometry)

We will attach a cohomology ring to any linear graph $\mathcal{G}$. Most of this section is a study of this ring and how to compute it. Then section 3.8 contains the main theorem: if the linear graph $\mathcal{G}$ is a moment graph $\mathcal{G}(T \circlearrowright X)$, then the cohomology ring of $\mathcal{G}$ is the equivariant cohomology ring of $T \circlearrowright X$.

3.1. The Definition of the Cohomology of a Linear Graph

$\langle$**1.1**$\rangle$ **Notations.** Suppose $\mathcal{G}$ is a linear graph. We will call the vertices $\nu, \nu', \ldots$ If ν and ν' are connected by an edge, we will call it $\ell_{\nu\nu'}$. The graph is embedded in a real n-dimensional vector space $\mathbb{V}$, whose dual vector space is $\mathbb{T}$. For every edge $\ell_{\nu\nu'}$, let $\ell^{\perp}_{\nu\nu'}$ be the $(n-1)$-dimensional subspace of $\mathbb{T}$ consisting of vectors that are orthogonal to the straight line $\ell_{\nu\nu'}$. Let $\mathcal{O}(\mathbb{T})$ be the ring of real polynomial functions $f : \mathbb{T} \longrightarrow \mathbb{R}$ graded so that the grading degree is twice the degree of the function.

Definition [**13**]**.** Consider the ring

$$\bigoplus_{\text{vertices } \nu \text{ of } \mathcal{G}} \mathcal{O}(\mathbb{T}).$$

An element of this ring is a polynomial function $f_\nu : \mathbb{T} \longrightarrow \mathbb{R}$ for each vertex ν of $\mathcal{G}$. We can notate such an element $\{f_\nu, f_{\nu'}, \ldots\}$ The *cohomology* of $\mathcal{G}$, $\mathcal{H}^*(\mathcal{G})$ is the subring of this cut out by the requirement that for every edge $\ell_{\nu\nu'}$ of $\mathcal{G}$, we have the *restriction condition:*

$$f_\nu|\ell^{\perp}_{\nu\nu'} = f_{\nu'}|\ell^{\perp}_{\nu\nu'}.$$

In other words, the restriction condition requires that if the vertices ν and ν' are connected by an edge $\ell_{\nu\nu'}$, then the polynomial f_ν and the polynomial $f_{\nu'}$ must have the same restriction to the space $\ell^{\perp}_{\nu\nu'}$.

For a useful reformulation of this definition, see §4.3.5.

$\langle$**1.2**$\rangle$ **Exercise.** Show that $\mathcal{H}^*(\mathcal{G})$ is a subring of $\bigoplus_\nu \mathcal{O}(\mathbb{T})$.

$\langle$**1.3**$\rangle$ **Graded structure.** The ring $\mathcal{H}^*(\mathcal{G})$ is a graded ring

$$\mathcal{H}^*(\mathcal{G}) = \bigoplus_{i \geq 0} \mathcal{H}^i(\mathcal{G})$$

where $\mathcal{H}^i(\mathcal{G}) = 0$ if i is odd, and $\mathcal{H}^{2k}(\mathcal{G})$ is the set of elements represented by sets of polynomials $\{f_\nu, \ldots\}$, each of which is homogeneous of degree k (i.e. every term of f_ν is of degree k). If $\alpha \in \mathcal{H}^i(\mathcal{G})$ and $\beta \in \mathcal{H}^j(\mathcal{G})$, then the product $\alpha\beta \in \mathcal{H}^{i+j}(\mathcal{G})$.

⟨**1.4**⟩ **Module structure.** The ring $\mathcal{H}^*(\mathcal{G})$ is a graded module over the graded ring $\mathcal{O}(\mathbb{T})$ of polynomial functions on $\mathbb{T}$. The module action of $g \in \mathcal{O}(\mathbb{T})$ sends $\{f_\nu, f_{\nu'}, \ldots\} \in \mathcal{H}^*(\mathcal{G})$ to $\{gf_\nu, gf_{\nu'}, \ldots\}$.

⟨**1.5**⟩ **Restriction.** Suppose that we have an inclusion of linear graphs $\mathcal{G}' \subset \mathcal{G}$, i.e. $\mathcal{G}'$ has some of the vertices of $\mathcal{G}$ and some of the edges. Then the projection

$$\bigoplus_{\text{vertices } \nu \text{ of } \mathcal{G}} \mathcal{O}(\mathbb{T}) \longrightarrow \bigoplus_{\text{vertices } \nu \text{ of } \mathcal{G}'} \mathcal{O}(\mathbb{T})$$

induces a map

$$\mathcal{H}^*(\mathcal{G}) \longrightarrow \mathcal{H}^*(\mathcal{G}').$$

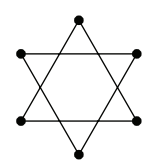
⟨**1.6**⟩ **Exercise.** Show that the graded structure, the module structure, and the restriction, as defined above, make sense – for example that they respect the condition for each edge of $\mathcal{G}$ in the definition of $\mathcal{H}^*(\mathcal{G})$.

⟨**1.7**⟩ Sections 3.2 to 3.7 will be devoted to the study of the cohomology ring of a graph. The definition is simple enough, but it is not immediately clear from the definition how you would compute it or how to think about it. The papers of Guillemin, Holm, and Zara are recommended for further reading [**18**], [**19**], [**20**], [**21**], [**22**].

3.2. Interpreting $\mathcal{H}^i(\mathcal{G})$ for Small i

In this section, we will give interpretations for $i = 0, 2,$ or 4.

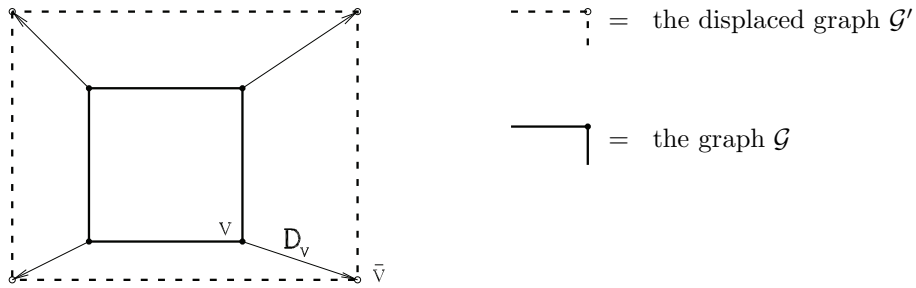
⟨**2.1**⟩ **The degree 0 part of the cohomology.** The dimension of the vector space $\mathcal{H}^0(\mathcal{G})$ is the number of connected components of the topological graph associated to $\mathcal{G}$. (The topological graph of the figure at the right consists of two disjoint triangles.) **Exercise.** Prove this.



$H^0(\mathcal{G}) = \mathbb{R} \oplus \mathbb{R}$

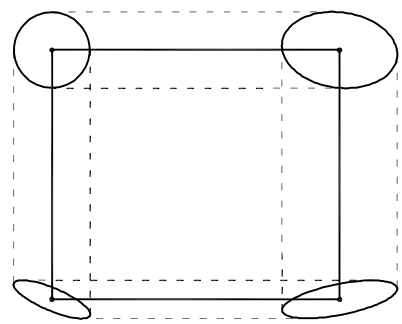
⟨**2.2**⟩ **The degree 2 part of the cohomology.** The dimension of the vector space $\mathcal{H}^2(\mathcal{G})$ is the dimension of the space of graphs in $\mathbb{V}$ that are equivalent to $\mathcal{G}$ (see §0.2.4).

⟨**2.3**⟩ **Proof.** An element of $\bigoplus_{\text{vertices } \nu \text{ of } \mathcal{G}} \mathcal{O}(\mathbb{T})^2$ is a linear function on $\mathbb{T}$ for every vertex ν of the graph $\mathcal{G}$. But a linear function on $\mathbb{T}$ is a vector D_ν in $\mathbb{V}$. Draw the vector D_ν as an arrow, and put its tail at the vertex ν and call its head $\bar{\nu}$. We will consider D_ν a displacement of ν to a new vertex $\bar{\nu}$ of a new graph $\bar{\mathcal{G}}$. For every edge $\ell_{\nu\nu'}$ the restriction condition that $f_\nu|\ell^\perp_{\nu\nu'} = f_{\nu'}|\ell^\perp_{\nu\nu'}$ translates in to the condition that the line $\ell_{\bar{\nu}\bar{\nu'}}$ connecting the head of D_ν to the head of $D_{\nu'}$, is parallel to $\ell_{\nu\nu'}$. But that is exactly the condition that the displaced graph $\bar{\mathcal{G}}$ should be in the same equivalence class of linear graphs as $\mathcal{G}$.



⟨**2.4**⟩ **Exercise.** Determine the dimension of $\mathcal{H}^2(\mathcal{G})$ for all of the linear graphs pictured in Lecture 2.

⟨**2.5**⟩ **The degree 4 part of the cohomology.** The dimension of the vector space $\mathcal{H}^4(\mathcal{G})$ is the dimension of the space $C(\mathcal{G})$ of configurations of the following sort: For each vertex ν of the graph $\mathcal{G}$, we give an ellipsoid E_ν in $\mathbb{V}$ centered at ν. For each edge $\ell_{\nu\nu'}$, we ask that when you take the projection along the direction of $\ell_{\nu\nu'}$ to an $(n-1)$-dimensional quotient space of $\mathbb{V}$, the two ellipsoids E_ν and $E_{\nu'}$ should have the same image. (Recall that an ellipsoid is the zero set of a degree two polynomial that is compact.) I am indebted to Victor Guillemin for this interpretation of $\mathcal{H}^4(\mathcal{G})$.



A configuration of ellipses in $C(\mathcal{G})$

⟨**2.6**⟩ **Exercise.** Prove this statement. More precisely, prove that the tangent space to $C(\mathcal{G})$ at any point is canonically $\mathcal{H}^4(\mathcal{G})$.

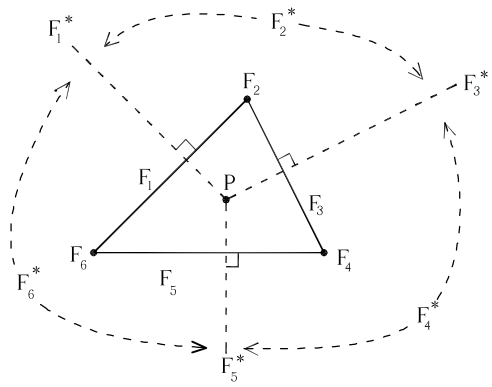
3.3. Piecewise Polynomial Functions

Suppose that $\mathcal{G}$ is the 1-skeleton of a convex polyhedron P. We will give an interpretation of the ring $\mathcal{H}^*(\mathcal{G}(P))$.

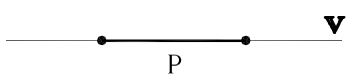
⟨**3.1**⟩ **The dual cone decomposition.** If P is a convex polyhedron in $\mathbb{V}$, then the dual space $\mathbb{T}$ is partitioned into subsets F^* corresponding to faces F of P as follows: If $t \in \mathbb{T}$, suppose that $c \in \mathbb{R}$ is the maximum value that the image $t(P)$ can take. Then $t^{-1}(c)$ will be some face F of P. We say that $t \in F^*$.

For example, $0 \in \mathbb{T}$ is always in P^* (P is a face of itself). If P has the same dimension as $\mathbb{V}$, then $0 = P^*$. The set F^* is an open subset of $\mathbb{T}$ if and only if F is a vertex of P.

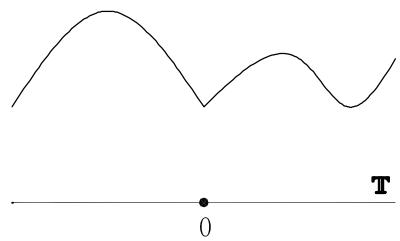
If we identify $\mathbb{V} = \mathbb{T} = \mathbb{R}^2$ and $t(v) =< t, v >$ where $< \cdot, \cdot >$ is the usual inner product, then we can picture the dual cone decomposition like this:



⟨**3.2**⟩ **Piecewise polynomial functions.** A function $f : \mathbb{T} \longrightarrow \mathbb{R}$ is called *piecewise polynomial* with respect to the dual cone decomposition $\mathbb{T} = \bigcup V^*$ if it is continuous and its restriction to each set V^* is given by a polynomial function.



A polyhedron in $V = \mathbb{R}^1$



The graph of a function that is piecewise polynomial with respect to the dual cone decomposition

⟨**3.3**⟩ **Exercise.** Show that a continuous function is piecewise polynomial if its restriction to F^* is given by a polynomial function for every vertex F.

⟨**3.4**⟩ **Interpretation of the cohomology of the** 1**-skeleton of** P**.** If $\mathcal{G}$ is the 1-skeleton of the polyhedron P, then its cohomology ring $\mathcal{H}^*(\mathcal{G})$ is the ring of functions on $\mathbb{T}$ that are piecewise polynomial with respect to the dual cone decomposition.

⟨**3.5**⟩ **Exercise.** Prove this. Use the lemma that a polynomial is entirely determined on its values on any open set.

⟨**3.6**⟩ **Remark.** When the polyhedron is simple, this ring is called the *Reisner Stanley ring* of the dual simplicial polyhedron.

3.4. Morse Theory

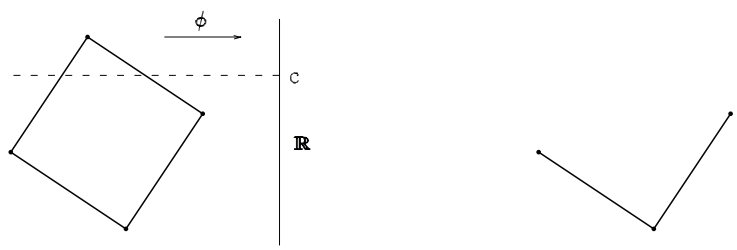
Morse theory is the main tool we have for understanding the cohomology of a graph. The idea of Morse theory is to break the computation of the cohomology down into a series of simpler computations.

⟨**4.1**⟩ **Morse functions.** Suppose we have a linear graph $\mathcal{G}$ in a vector space $\mathbb{V}$. Consider a linear function $\phi : \mathbb{V} \longrightarrow \mathbb{R}$. The values $\phi(\nu)$ where ν is a vertex of $\mathcal{G}$ are called the *critical values* of ϕ. The function ϕ is called a *Morse function* if all of the

critical values are distinct, i.e. for any pair of vertices ν and ν' of $\mathcal{G}$, $\phi(\nu) \neq \phi(\nu')$. It follows that ϕ is not constant on any edge of $\mathcal{G}$.

Morse functions exist for any linear graph. In fact, if you choose a linear function $\phi : \mathbb{V} \longrightarrow \mathbb{R}$ at random, you have to be infinitely unlucky to get one that is not Morse.

⟨4.2⟩ The truncated graph. Now suppose c is a real number, which we call the "cut-off value". We define $\mathcal{G}^{\leq c}$ to be the subgraph of $\mathcal{G}$ consisting of those vertices ν such that the critical value $\phi(\nu) \leq c$, together with all the edges $\ell_{\nu\nu'}$ connecting vertices ν and ν' both of which have critical values $\leq c$.



A Morse function ϕ The truncated graph $\mathcal{G}^{\leq c}$

If we have a Morse function ϕ, we can label the vertices of $\mathcal{G}$ by $\nu_1, \nu_2, \ldots \nu_k$ so that their critical values are increasing $\phi(\nu_1) < \phi(\nu_2) < \cdots < \phi(\nu_k)$. Call c_j the critical value $\phi(\nu_j)$. For $c < c_1$, we have $\mathcal{G}^{\leq c}$ is empty. As the number c increases, $\mathcal{G}^{\leq c}$ grows by jumps every time c reaches a critical value c_j until finally for $c \geq c_k$, $\mathcal{G}^{\leq c} = \mathcal{G}$. The idea of Morse theory is to trace the growth of $\mathcal{H}^*(\mathcal{G}^{\leq c})$ as c increases.

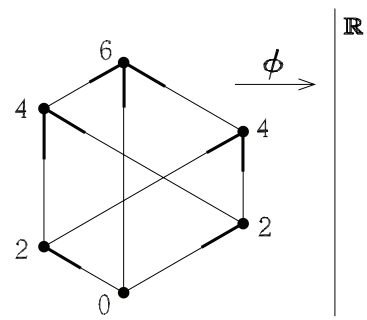
⟨4.3⟩ The Morse module. Suppose that $c_1 < c_2 < \cdots < c_k$ are the critical values of the Morse function ϕ, and c_0 is a real number less than the smallest critical value c_1. Then for all integers $j \in \{1, \ldots, k\}$, we have

$$\mathcal{G}^{\leq c_{j-1}} \overset{i_j}{\subset} \mathcal{G}^{\leq c_j}$$

$$\mathcal{H}^*(\mathcal{G}^{\leq c_{j-1}}) \xleftarrow{i_j^*} \mathcal{H}^*(\mathcal{G}^{\leq c_j}) \longleftarrow M_j \longleftarrow 0$$

Here i_j^* is the map on cohomology induced by the inclusion of graphs i_j and M_j the kernel of the map i_j^*. The kernel M_j is a graded module over $\mathcal{O}(\mathbb{T})$ because it is the kernel of a map of graded modules. It is a graded ideal in $\mathcal{H}^*(\mathcal{G}^{\leq c_j})$, but it is more useful to think of it as a $\mathcal{O}(\mathbb{T})$-module. The module M_j is called *the Morse module* of the vertex ν_j whose critical value is c_j.

⟨**4.4**⟩ **The Morse index.** For any vertex $\nu \in \mathcal{G}$, let $\mathcal{L}(\nu)$ denote the set of edges coming in to ν. The Morse function ϕ splits the edges in $\mathcal{L}(\nu)$ into two types: $\mathcal{L}^-(\nu)$ is the edges going down from ν as measured by ϕ, i.e. the edges connecting ν to vertices ν' with $\phi(\nu') < \phi(\nu)$. The others are in $\mathcal{L}(\nu)^+$, the edges going up from ν. We define the *Morse index* $\mathrm{Index}(\nu)$ to be twice the number of edges in $\mathcal{L}^-(\nu)$.



Morse indices of vertices

⟨**4.5**⟩ **Calculation of the Morse module M_j.** The graph $\mathcal{G}^{\leq c_j}$ has exactly one more vertex than the graph $\mathcal{G}^{\leq c_{j-1}}$, namely ν_j. Therefore for an element $\{\ldots, f_\nu, \ldots\}$ of $M_j \subset \mathcal{H}^*(\mathcal{G}^{\leq c_j})$, all of the f_ν will be zero except for f_{ν_j} corresponding to ν_j. This polynomial $f_{\nu_j} : \mathbb{T} \longrightarrow \mathbb{R}$ will vanish on all of the hyperplanes $\ell^\perp$ for $\ell \in \mathcal{L}^-$. For each $\ell \in \mathcal{L}^-$, let g_ℓ be a nonzero linear function on $\mathbb{T}$ that is zero on $\ell^\perp$.

Proposition. The Morse module M_j is the principal ideal in $\mathcal{O}(\mathbb{T})$ generated by the homogeneous element

$$g_{\nu_j} = \prod_{\ell \in \mathcal{L}^-(\nu_j)} g_\ell.$$

As a module over $\mathcal{O}(\mathbb{T})$, M_j is a free module generated by g_{ν_j}, which lies in the graded piece $\mathcal{O}(\mathbb{T})^{\mathrm{Index}(\nu_j)}$.

⟨**4.6**⟩ **Exercise.** Finish the proof of this proposition.

⟨**4.7**⟩ **Exercise*.** Suppose that $\mathcal{G}$ is the 1-skeleton of a simple polyhedron. Show that the ordering of the vertices given by a Morse function corresponds to a linear shelling of the dual simplicial polytope.

3.5. Perfect Morse Functions

Having a Morse function isn't much help unless the cokernel of the map

$$\mathcal{H}^*(\mathcal{G}^{\leq c_{j-1}}) \xleftarrow{i_j^*} \mathcal{H}^*(\mathcal{G}^{\leq c_j})$$

is zero, because in general it can be very difficult to compute this cokernel. If it is zero, the Morse function is called perfect:

⟨**5.1**⟩ **Definition.** The Morse function φ is called *perfect* if i_j^* is surjective for all j.

⟨**5.2**⟩ **Hilbert series.** One of our goals is to compute the dimensions of the cohomology groups $\mathcal{H}^*(\mathcal{G})$, or equivalently, to compute the Hilbert series of the cohomology of a graph $\mathrm{Hilb}(\mathcal{H}^*(\mathcal{G}))$(see §0.3.4). This determines the isomorphism class of $\mathcal{H}^*(\mathcal{G})$ except for the ring structure and the structure as a module over $\mathcal{O}(\mathbb{T})$.

If ϕ is perfect, then we see by induction that the dimension of the i-th graded piece of $\mathcal{H}^*(\mathcal{G})$ is the sum of the dimensions of the i-th graded pieces of the Morse

modules. Expressed in Hilbert series,

$$\mathrm{Hilb}(\mathcal{H}^*(\mathcal{G})) = \sum_{1\le j\le k} \mathrm{Hilb}(M_j).$$

But since the Morse module M_j is a free $\mathcal{O}(\mathbb{T})$ module on a generator of degree $\mathrm{Index}(\nu_j)$, and the Hilbert series of $\mathcal{O}(\mathbb{T})$ is computed in §0.3.5, we have the following:

⟨**5.3**⟩ **Proposition.** If ϕ is a perfect Morse function, the Hilbert series of the cohomology of the graph is given by

$$\mathrm{Hilb}(\mathcal{H}^*(\mathcal{G})) = \sum_{i=1}^{k} x^{\mathrm{Index}(\nu_i)} \left(\frac{1}{1-x^2}\right)^n = \sum_{i=1}^{k} q^{\mathrm{Index}(\nu_i)/2} \left(\frac{1}{1-q}\right)^n.$$

⟨**5.4**⟩ **The Betti numbers of a graph.** Suppose that $\mathcal{G}$ has a perfect Morse function ϕ. Then we define the *Betti numbers* B_i of $\mathcal{G}$ to be the number of vertices of $\mathcal{G}$ whose Morse index is i. Note that B_i is automatically zero if i is odd.

We define the Poincaré polynomial P to be

$$P(x) = \sum_i B_i x^i.$$

so we have

$$\mathrm{Hilb}(\mathcal{H}^*(\mathcal{G})) = P(x)\left(\frac{1}{1-x^2}\right)^n$$

and

$$P(x) = \mathrm{Hilb}(\mathcal{H}^*(\mathcal{G}))\left(1-x^2\right)^n$$

where the last expression, which is *a priori* an infinite power series, is actually a polynomial.

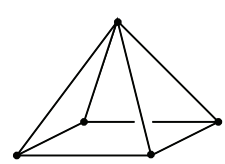
⟨**5.5**⟩ **Exercise.** Show that if the graph $\mathcal{G}$ has more than one different perfect Morse function, the Betti numbers (and the Poincaré polynomials) are independent of the choice of the Morse function.

⟨**5.6**⟩ **Exercise.** Show that the sum of the Betti numbers $\sum_i B_i$ is the number of vertices of the graph $\mathcal{G}$.

⟨**5.7**⟩ **Exercise.** Show that the sum $\sum_i (i/2)B_i$ is the number of edges of the graph $\mathcal{G}$.

⟨**5.8**⟩ **Exercise*.** Show that the homology groups of the topological graph $\mathcal{G}$ are determined by the Betti numbers of $\mathcal{H}^*(\mathcal{G})$.

⟨**5.9**⟩ **Exercise.** Let $\mathcal{G}$ be the 1-skeleton of the Egyptian pyramid in 3-space. Show that not all of the Morse functions on $\mathcal{G}$ are perfect by showing that they would lead to different Betti numbers. Can you identify which ones are not perfect?



⟨**5.10**⟩ **Exercise.** Show that the height function on the linear graph in the plane displayed on the right is not perfect. We call this graph the "inverted V".

⟨**5.11**⟩ **Remark.** It can be difficult to tell whether a given Morse function ϕ is perfect. There is one deep general theorem about this, due to Guillemin and Zara [**18**]. However, as we will see in §3.8.8, most of the cases we are considering have perfect Morse functions for topological reasons.

⟨**5.12**⟩ **Exercise.** It is known that all Morse functions of the graphs pictured in Lecture 2 are perfect. Calculate their Betti numbers.

3.6. Determining $\mathcal{H}^*(\mathcal{G})$ as a $\mathcal{O}(\mathbb{T})$ Module

For many purposes, we want more than the dimensions of the cohomology groups $\mathcal{H}^i(\mathcal{G})$. A Morse function ϕ enables us to determine it as a $\mathcal{O}(\mathbb{T})$-module:

⟨**6.1**⟩ **Proposition.** If $\mathcal{G}$ has a perfect Morse function, its cohomology $\mathcal{H}^*(\mathcal{G})$ is a free graded $\mathcal{O}(\mathbb{T})$-module. The number of free generators of degree $2i$ is B_i.

This proposition can be proved inductively, using §3.4.5.

⟨**6.2**⟩ **Proposition.** If ϕ is perfect, the cohomology $\mathcal{H}^*(\mathcal{G})$ is a free graded module over $\mathcal{O}\mathbb{T}$, i.e.

$$\mathcal{H}^*(\mathcal{G}) = \bigoplus_j \overline{g_j}\mathcal{O}(\mathbb{T})$$

where $\overline{g_j}$ is a lift of g_i to $\mathcal{H}^*(\mathcal{G})$.

⟨**6.3**⟩ **Exercise.** Prove this.

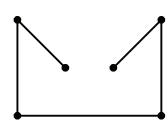
⟨**6.4**⟩ **Definition.** We say that a linear graph has the *free module property* if its cohomology is a free graded module over $\mathcal{O}(\mathbb{T})$.

If a graph has the free module property, we may define its Poincaré polynomial by

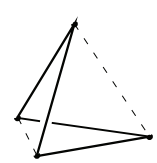
$$P(\mathcal{G}) = \mathrm{Hilb}(\mathcal{H}^*(\mathcal{G}))(1-q)^n$$

which will necessarily be a polynomial. Graphs with a perfect Morse function have the free module property, but the converse isn't true:

⟨**6.5**⟩ **Exercise.** Show that the graph to the right has no perfect Morse function, but has the free module property.



⟨**6.6**⟩ **Exercise.** Show that a nonplanar quadrilateral in 3-space does not have the free module property (example due to T. Braden).



3.7. Poincaré Duality

Suppose that $\mathcal{G}$ is k-valent: it has k edges coming out of every vertex ($\#|\mathcal{L}(\nu)| = k$ for all vertices $\nu \in \mathcal{G}$). The simplest form of Poincaré duality is the numerical statement that the Betti numbers B_j and B_{2k-j} are equal.

This numerical Poincaré duality holds whenever there is a perfect Morse function whose negative is also perfect.

⟨**7.1**⟩ **Exercise.** Suppose that $\mathcal{G}$ is k-valent and that it has a perfect Morse function ϕ such that the Morse function $-\phi$ is also perfect. Show that

$$B_j(\mathcal{G}) = B_{2k-j}(\mathcal{G}).$$

As usual in mathematics, it is better to have a canonical isomorphism or a duality

of vector spaces than an equality of their dimensions. We want something of the kind for Poincaré duality. First, we need some preliminaries on graded rings.

⟨**7.2**⟩ **The canonical filtration of a graded R-module.** Consider a graded module M over a graded ring R. Let $M^{\leq k}$ be the sum of the graded pieces $M^0 \oplus M^1 \oplus \cdots \oplus M^k$. This is not an R-module, but it generates one; call it $F_k M = R \cdot M^{\leq k}$. Then M has an increasing filtration of R-submodules $F_0 M \subseteq F_1 M \subseteq \cdots$.

⟨**7.3**⟩ **Exercise.** If $\mathcal{G}$ has the free module property, then B_i is the dimension of the i graded piece of $F_i\mathcal{H}^*(\mathcal{G})/F_{i-1}\mathcal{H}^*(\mathcal{G})$.

⟨**7.4**⟩ **Internal Hom.** Suppose that M and N are two graded R modules. Then the space $\mathrm{Hom}_R(M,N)$ has the structure of a graded R module. The i-th graded piece is the elements of $\mathrm{Hom}_R(M,N)$ that map each M^j into N^{j+i}.

⟨**7.5**⟩ **Proposition.** *Functorial Poincaré duality.* Now, suppose that $\mathcal{G}$ is connected, k-valent, and that it is universally perfect (i.e. all Morse functions are perfect). Then $\mathcal{H}^*(\mathcal{G})/F_{2k-1}\mathcal{H}^*(\mathcal{G})$ is a free $\mathcal{O}(\mathbb{T})$ module on one generator in degree $2k$. Call it $\mathcal{D}$. The pairing

$$\begin{array}{ccccccc} \mathcal{H}^*(\mathcal{G}) & \otimes_{\mathcal{O}(\mathbb{T})} & \mathcal{H}^*(\mathcal{G}) & \longrightarrow & \mathcal{H}^*(\mathcal{G}) & \longrightarrow & \frac{\mathcal{H}^*(\mathcal{G})}{F_{2k-1}\mathcal{H}^*(\mathcal{G})} = \mathcal{D} \\ x & \times & y & \mapsto & xy & & \end{array}$$

is perfect in the sense that the induced map

$$\mathcal{H}^*(\mathcal{G}) \longrightarrow \mathrm{Hom}(\mathcal{H}^*(\mathcal{G}), \mathcal{D})$$

is an isomorphism of $\mathcal{O}(\mathbb{T})$ modules.

⟨**7.6**⟩ **Exercise.** Show that functorial Poincaré duality implies numerical Poincaré duality.

3.8. The Main Theorems

This section relates the cohomology of a linear graph to torus actions and the moment graph construction.

⟨**8.1**⟩ **Assumptions.** We consider a torus acting on a space $T \circlearrowright X$ such that the moment graph $\mathcal{G}(T \circlearrowright X)$ exists. (This means, in particular, that the 1-skeleton of the action is a balloon sculpture.) *We further assume that X has only even dimensional real cohomology*, i.e. $H_i(X;\mathbb{R}) = 0$ for i odd.

These assumptions hold for complex projective spaces, quadric hypersurfaces, Grassmannians, Lagrangian Grassmannians, flag manifolds and their and toric varieties based on simple polyhedra. In other words, the assumptions hold for all spaces considered in Lecture 2, except for toric varieties of some non-simple polyhedra.

⟨**8.2**⟩ **Theorem** [**13**]. The T equivariant cohomology ring of X is the cohomology ring of the moment graph of X, i.e.

$$H^*(T \circlearrowright X) = \mathcal{H}^*(\mathcal{G}(T \circlearrowright X)).$$

⟨**8.3**⟩ **Theorem.** The ordinary (non-equivariant) cohomology ring of X, calculated from the moment graph by

$$H^*(X) = \frac{\mathcal{H}^*(\mathcal{G}(X))}{\text{the ideal generated by } \mathcal{O}(\mathbb{T})^{>0}}$$

where $\mathcal{O}(\mathbb{T})^{>0}$ is the positive degree part of $\mathcal{O}(\mathbb{T})$.

⟨**8.4**⟩ **Theorem.** The moment graph has the free module property, i.e. $\mathcal{H}^*(\mathcal{G}(X))$ is a free graded module over the polynomial algebra $\mathcal{O}(\mathbb{T})$ and the Poincaré polynomial of the graph $P(\mathcal{G}) = \mathrm{Hilb}(\mathcal{H}^*(\mathcal{G}))(1-q)^n$ is the Poincaré polynomial $\sum_i x^i \dim H_i(X)$ of X.

⟨**8.5**⟩ We can pause to marvel at the statements. The data in moment graph of $T \circlearrowright X$ depends only on a very small part of X – its 1-skeleton. Yet by these theorems, all of the homology and equivariant homology of X is encoded in this data.

The proofs of these three propositions are beyond our ambitions here. The reader is referred to [**13**] and the references given there. However, we have given enough information in our explicit construction of generators and relations for the equivariant cohomology of the 2-sphere, we have to construct the map

$$H^*(T \circlearrowright X) \xleftarrow{\sim} \mathcal{H}^*(\mathcal{G}(T \circlearrowright X))$$

in Theorem 3.8.2.

⟨**8.6**⟩ **Exercise.** Construct a map $H^*(T \circlearrowright X) \mapsto \mathcal{H}^*(\mathcal{G}(T \circlearrowright X))$.

⟨**8.7**⟩ **Exercise.** Show that the free generators $\alpha_1, \alpha_2, \ldots$ for $\mathcal{H}^*(\mathcal{G}(X))$ as a module over $\mathcal{O}(\mathbb{T})$ pass in the quotient to generators of $H^*(X)$ as a vector space, i.e. as a module over $\mathbb{R}$.

⟨**8.8**⟩ **Morse theory and Poincaré duality for our examples.** In Lecture 2, we gave many examples of spaces with a torus action: projective spaces, quadric hypersurfaces, Grassmann manifolds, Lagrangian Grassmannians, flag manifolds, and toric varieties for simple polyhedra. These examples all satisfy the hypotheses of the theorems above. Furthermore, they are all universally perfect (every Morse function is perfect), so they satisfy Poincaré duality. (This may be seen using topological methods.) Many other examples in this favorable class will be mentioned in §4.1.2.

⟨**8.9**⟩*** Morse theory and moment maps.** Suppose that X is a nonsingular algebraic variety, and the action $T \circlearrowright X$ and the moment map $\mu : X \longrightarrow \mathbb{V}$ are as in §2.9. If $\varphi : V \longrightarrow \mathbb{R}$ is a Morse function for the moment graph of $T \circlearrowright X$ in the sense of this Lecture, then $\varphi \circ \mu : X \longrightarrow \mathbb{R}$ is a Morse function in the usual sense of differential topology. In this case, the Morse function will be perfect. In this case, Morse theory we have described is a reflection of the usual topological Morse theory.

⟨8.10⟩* The Schubert basis. Suppose X is a generalized flag manifold, i.e. a projective space, a quadric hypersurface, a Grassmann manifold, a Lagrangian Grassmannian, a flag manifold, or more generally a space of §4.7.2. Then the Morse function $\varphi \circ \mu$ is perfect on ordinary cohomology $H^*(X)$. The basis of cohomology it provides is called the *Schubert basis*, and the study of the properties of this basis in the ring $H^*(X)$ is called *Schubert calculus*, an interesting combinatorial study involving such things as the Littlewood-Richardson rule, Schubert polynomials, etc. By Exercise 3.8.7, the $H^*(X)$ and its Schubert basis is encoded in the moment graph, so in principle questions in Schubert calculus reduce to questions about the moment graph.

⟨8.11⟩* A general Lie group. Here's a brief account. Suppose $G \circlearrowright X$ is an action of a general connected Lie group. Then $H^*(G \circlearrowright X) = H^*(K \circlearrowright X)$ where K is a maximal compact subgroup of G. Then, by a theorem of Borel, $H^*(K \circlearrowright X) = H^*(T \circlearrowright X)^W$ where T is a maximal torus of K and W is the Weyl group of K and the superscript means taking the invariants. Now, suppose the T action satisfies our hypotheses, so it has moment graph in $\mathcal{G}(T \circlearrowright X) \subset \mathbb{V}$. The Weyl group W acts on $\mathbb{V}$ preserving the moment graph, so we can calculate $H^*(G \circlearrowright X) = \mathcal{H}^*(\mathcal{G}(T \circlearrowright X))^W$.

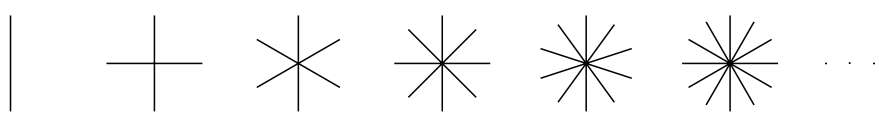
LECTURE 4
Computing Intersection Homology
(Polynomial and Linear Geometry II)

In the last lecture, we saw the value of perfect Morse functions. In this lecture, we consider some linear graphs $\mathcal{G}$ with Morse functions that are not perfect. By changing the cohomology theory, the Morse functions become perfect again. When $\mathcal{G}$ arose as the moment graph of $T \circlearrowright X$, the new cohomology theory turns out to be the equivariant intersection cohomology of $T \circlearrowright X$.

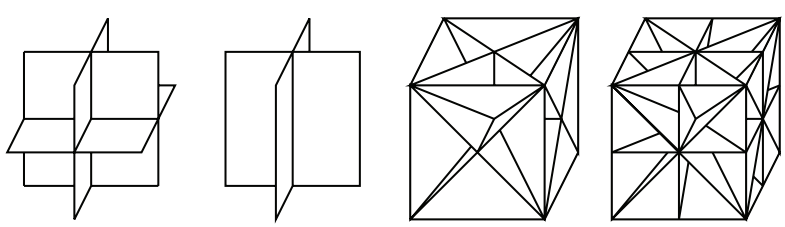
All of the ideas of this Lecture are joint work with Tom Braden.

4.1. Graphs Arising from Reflection Groups

$\langle$**1.1**$\rangle$ **Finite reflection groups.** Consider a finite configuration of hyperplanes $\mathcal{H}$ in $\mathbb{V}$ that pass through the origin. Suppose that reflection in each hyperplane H in $\mathcal{H}$ takes the configuration $\mathcal{H}$ to itself. Then we call $\mathcal{H}$ a set of reflecting hyperplanes. These are all classified. For example here are the sets of reflecting hyperplanes when $\mathbb{V}$ has dimension 2:

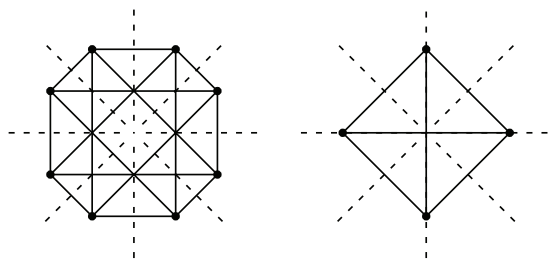


and here are some when $\mathbb{V}$ has dimension 3:



Or, when $\mathbb{V} = \mathbb{R}^n$, $\mathcal{H}$ could be the $\binom{n}{2}$ planes $x_i = x_j$ in $\mathbb{R}^n$ where two coordinates are equal. A *finite reflection group* W is the group of maps of $\mathbb{V}$ to itself generated by reflections in hyperplanes in $\mathcal{H}$.

⟨**1.2**⟩ **The linear graph associated to $\mathcal{H}$.** Choose any point $v \in \mathbb{V}$. We get a linear graph $\mathcal{G}(\mathcal{H}, v)$ as follows: The set of vertices of $\mathcal{G}(\mathcal{H}, v)$ is the orbit Wv of the point v. Two vertices ν and ν' are connected by an edge whenever ν' is the reflection of ν through one of the hyperplanes of $\mathcal{H}$. (So the edge will be perpendicular to the hyperplane.) Here are two of the possible graphs associated to a single $\mathcal{H}$ where $\mathbb{V}$ has dimension 2:



⟨**1.3**⟩ **Exercise.** All of the linear graphs pictured in sections 2.3 to 2.7 arise in this way. Construct the family of hyperplanes $\mathcal{H}$ for each of them.

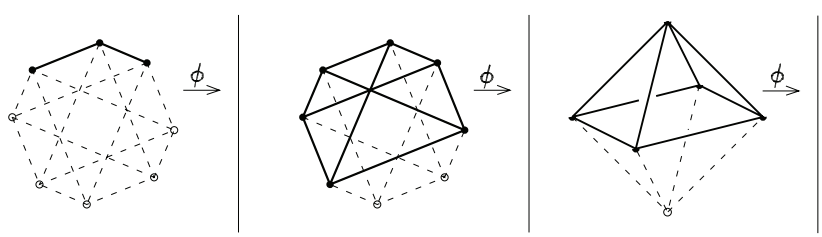
⟨**1.4**⟩ **Crystallographic reflection groups.** If there is some lattice $L \subset \mathbb{V}$ such that reflection in each of the hyperplanes in $\mathcal{H}$ takes this lattice into itself, then $\mathcal{H}$ is called *crystallographic*. This is true for most, but not all of the possible choices for $\mathcal{H}$. If $\mathcal{H}$ is crystallographic, then $\mathcal{G}(\mathcal{H}, v)$ arises as a moment graph, as described in §4.7.2.

⟨**1.5**⟩ **The linear graphs $\mathcal{G}(\mathcal{H}, v)$ are all universally perfect.** In other words, all Morse functions on these graphs are perfect. (This may be seen using a topological argument if $\mathcal{H}$ is crystallographic. In general, it follows from [**18**].) In fact, every graph we have considered so far is universally perfect, with the exception of a few counterexamples and 1-skeleta of non-simple polytopes. We will now construct a large class of examples with non-perfect Morse functions.

4.2. Upward Saturated Subgraphs

⟨**2.1**⟩ Consider a linear graph arising from a finite reflection group $\mathcal{G}(\mathcal{H}, v) \subset \mathbb{V}$ and a Morse function $\phi : \mathbb{V} \longrightarrow \mathbb{R}$ (a linear function that takes distinct values on different vertices of $\mathcal{G}$). Recall (§3.4.4) that if ν is a vertex of $\mathcal{G}$, we define $\mathcal{L}^-(\nu)$ to be the edges going down from ν and $\mathcal{L}^+(\nu)$ to be the edges going up from ν, where "up" and "down" are measured by ϕ.

⟨**2.2**⟩ **Definition.** We call a subgraph $\mathcal{G}'$ of $\mathcal{G}$ *upward saturated* with respect to ϕ if whenever ν is in $\mathcal{G}'$ then every edge in $\mathcal{L}^+(\nu)$ is in $\mathcal{G}'$.



Upward saturated subgraphs

⟨**2.3**⟩ The Morse function ϕ is not usually perfect on upward saturated subgraphs. In fact, for two of the examples above, the inverted V §3.5.10 and the Egyptian pyramid §3.5.9 the function ϕ has already been shown not to be perfect. However,

⟨**2.4**⟩ **Exercise.** Show that $-\phi$ is perfect for an upward saturated subgraph. (Use the fact that $\mathcal{G}(\mathcal{H}, v)$ is universally perfect.)

⟨**2.5**⟩ **Exercise*.** The Morse function ϕ turns the set of vertices of $\mathcal{G}(\mathcal{H}, v)$ into a poset where $\nu \leq \nu'$ if there is a sequence of edges from ν to ν' such that ϕ increases along each edge. The partial order of this poset is called the *Bruhat order*. Show that an upward saturated subgraph can be characterized as a complete subgraph on a set of vertices that is an ideal in this poset.

4.3. Sheaves on Graphs

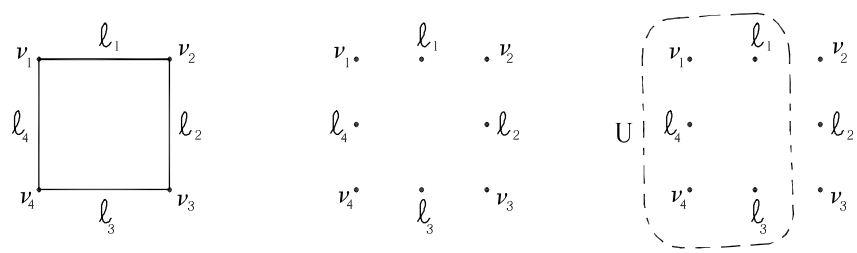
We introduce the notion of a sheaf on a graph. This will give us another interpretation of the cohomology of a linear graph. It will also give us a better understanding of when a Morse function is perfect.

Definition [**7**]**.** Suppose $\mathcal{G}$ is a topological graph. A *sheaf of graded rings* $\underline{S}$ on $\mathcal{G}$ is the following data.

(1) A graded ring S_ν for every vertex ν of $\mathcal{G}$;
(2) A graded ring S_ℓ for ever edge ℓ of $\mathcal{G}$; and
(3) A graded ring homomorphism $s_{\nu\ell} : S_\nu \longrightarrow S_\ell$ whenever ν lies on ℓ.

There is a similar definition replacing "rings" by any other category, such as graded modules over a graded ring.

⟨**3.1**⟩ **Definition.** Consider the set $E(\mathcal{G})$ which is the union of the set of vertices of the graph with the set of edges of the graph. An *open subset* of $E(\mathcal{G})$ is a subset U with the property that if a vertex ν is in U, then all the edges in $\mathcal{L}(\nu)$ are in U (where $\mathcal{L}(\nu)$ is the set of edges containing ν).



Linear graph $\mathcal{G}$ The finite set $E(\mathcal{G})$ An open subset U

⟨**3.2**⟩ **Definition.** Let U be an open subset of $E(\mathcal{G})$. A *section* of a sheaf $\underline{S}$ over U is the choice of an element e_ν for every vertex ν in $E(\mathcal{G})$ and element e_ℓ for every edge ℓ in $E(\mathcal{G})$ such that if ν lies in ℓ, then $s_{\nu\ell}(e_\nu) = e_\ell$. We will notate the set of such sections $\Gamma(\underline{S}, U)$.

If $U' \subset U$, then we have a restriction homomorphism $\Gamma(\underline{S}, U) \longrightarrow \Gamma(\underline{S}, U')$ defined by restricting the data.

⟨**3.3**⟩ **Exercise*.** Show that the definition of open set makes $E(\mathcal{G})$ into a (finite) topological space. Show that the function $U \mapsto \Gamma(\underline{S}, U)$ satisfies the sheaf axioms

$E(\mathcal{G})$. Establish an equivalence between sheaves as usually defined on the finite topological space $E(\mathcal{G})$ and the notion of a sheaf on a graph.

⟨**3.4**⟩ **The sheaf $\underline{A}$.** Now suppose that $\mathcal{G}$ is a linear graph. Then it has a canonical sheaf of graded rings $\underline{A}$ on it defined as follows.

(1) The graded ring A_ν is $\mathcal{O}(\mathbb{T})$, the ring of polynomial functions on $\mathbb{T} = \mathbb{V}^*$ for every vertex ν of $\mathcal{G}$;
(2) The graded ring A_ℓ for the edge ℓ of $\mathcal{G}$ is $\mathcal{O}(\ell^\perp)$, the ring of polynomial functions on $\ell^\perp \subset \mathbb{T}$.
(3) The homomorphism $a_{\nu\ell}$ is the restriction of polynomial functions.

⟨**3.5**⟩ **Proposition.** The cohomology $\mathcal{H}^*(\mathcal{G})$ of the graph $\mathcal{G}$ is the global sections $\Gamma(\underline{A}, E(\mathcal{G}))$ of the sheaf $\underline{A}$.

This is just a slightly disguised presentation of the definition of $\mathcal{H}^*(\mathcal{G})$.

4.4. A Criterion for Perfection

We mix the language of sheaves on graphs with Morse theory.

⟨**4.1**⟩ Recall from §3.5.1 that the criterion for a Morse function ϕ to be perfect is a surjectivity condition for each vertex of the graph. We will focus on this criterion for a single vertex ν. Suppose that ϕ is a Morse function for the linear graph $\mathcal{G}$, ν is the vertex with the largest critical value $\phi(\nu) = c$, and $c' < c$ is the next to the largest critical value. The Morse function ϕ is perfect at ν if the map

$$\mathcal{H}^*(\mathcal{G}) = \mathcal{H}^*(\mathcal{G}^{\leq c}) \longrightarrow \mathcal{H}^*(\mathcal{G}^{\leq c'})$$

is a surjection.

⟨**4.2**⟩ Consider the following open cover of the finite set $E(\mathcal{G})$.

- $E^{<c} = E(\mathcal{G}) - \{\nu\}$
- $\mathcal{S}(\nu) = \{\nu\} \cup \mathcal{L}^-(\nu)$ ($\mathcal{S}$ for star)

so that

- $E^{<c} \cap \mathcal{S}(\nu) = \mathcal{L}^-(\nu)$
- $E^{<c} \cup \mathcal{S}(\nu) = E(\mathcal{G})$

We have a diagram

$$\underline{A}_\nu = \Gamma(\underline{A}, \mathcal{S}(\nu)) \longrightarrow \Gamma(\underline{A}, \mathcal{L}^-(\nu)) \longleftarrow \Gamma(\underline{A}, E^{<c}) = \mathcal{H}^*(\mathcal{G}^{\leq c'})$$

where both maps are restriction maps.

⟨**4.3**⟩ **Proposition.** The image of $\underline{A}_\nu \longrightarrow \Gamma(\underline{A}, \mathcal{L}^-(\nu))$ is contained in the image of $\Gamma(\underline{A}, \mathcal{L}^-(\nu)) \longleftarrow \Gamma(\underline{A}, E^{<c})$. The graph is perfect at ν if and only if these two images coincide, i.e. if and only if the map

$$\underline{A}_\nu \longrightarrow \text{Image}\left[\Gamma(\underline{A}, \mathcal{L}^-(\nu)) \longleftarrow \Gamma(\underline{A}, E^{<c})\right]$$

is a surjection.

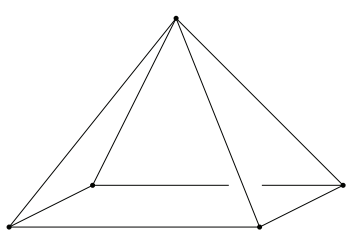
⟨**4.4**⟩ **Exercise.** Prove this. You may want to prove the following lemmas first:
• The map $\Gamma(\underline{A}, E(\mathcal{G})) \longrightarrow \Gamma(\underline{A}, \mathcal{S}(\nu))$ is surjective.
• The maps $\Gamma(\underline{A}, \mathcal{L}^-(\nu)) \longleftarrow \Gamma(\underline{A}, E^{<c})$ and $\Gamma(\underline{A}, E(\mathcal{G})) \longrightarrow \Gamma(\underline{A}, \mathcal{S}(\nu))$ have isomorphic kernels.

⟨**4.5**⟩ If you want to understand what makes a Morse function perfect, it is worthwhile to pause to appreciate this proposition. An element e in $\Gamma(\underline{A}, \mathcal{L}^-(\nu))$ is just a collection of polynomials on the hyperplanes $\ell^\perp \subset \mathbb{T}$ for $\ell \in \mathcal{L}^-(\nu)$. The condition that e be in the image of $\Gamma(\underline{A}, \mathcal{L}^-(\nu)) \longleftarrow \Gamma(\underline{A}, E^{<c})$ is a potentially complicated compatibility condition on these polynomials, coming from the structure of the graph. According to the proposition, it is perfect if and only if a set of polynomials $\ell^\perp$ satisfying this compatibility condition is necessarily the restriction of a single polynomial on $\mathbb{T}$.

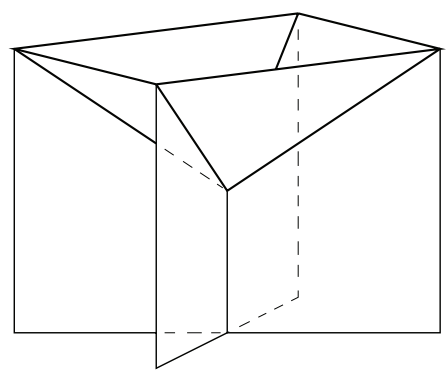
⟨**4.6**⟩ For example, take the Egyptian pyramid. As in §3.3.4, an element of the image of

$$\Gamma(\underline{A}, \mathcal{L}^-(\nu)) \longleftarrow \Gamma(\underline{A}, E^{<c})$$

is a continuous piecewise polynomial function on configuration consisting of the four upper planes on the right below. We are asking whether such a thing is the restriction of a polynomial in 3-space. We can see that it is not, just by a dimension count. For example, there is a 4-dimensional space of linear polynomials on the configuration (it can be anything on each of the 4 lines). But there is only a 3-dimensional space of linear polynomials in 3-space. So the height function is not perfect. (Compare §3.5.9.)



Egyptian pyramid



The dual cone decomposition

4.5. Definition of the Sheaf $\underline{M}$

We define a sheaf of $\mathcal{O}(\mathbb{T})$ modules that repairs the deficiency in perfectness as measured by Proposition 4.4.3.

⟨**5.1**⟩ **Hypotheses on the graph $\mathcal{G}$.** We will assume that the graph $\mathcal{G}$ is an upward saturated subgraph of a graph $\mathcal{G}'$ that arises from a crystallographic reflection group as in §4.1.2. For example, the ambient graph $\mathcal{G}'$ can be any of the graphs of Lecture 2, except for some moment graphs of toric varieties.

⟨**5.2**⟩ **Construction.** [**7**]. We define the sheaf $\underline{M}$ inductively:
• If $\mathcal{L}^-(\nu)$ is empty, then M_ν is a free module over $\mathcal{O}(\mathbb{T})$ generated by a generator g_ν in degree $-|\mathcal{L}^+(\nu)|$, minus the number of edges going up from ν.
• If M_ν has already been constructed, and $\ell \in \mathcal{L}^+(\nu)$ is an edge going up from M_ν then

$$M_\ell = M_\nu \otimes_{\mathcal{O}(\mathbb{T})} \mathcal{O}(\ell^\perp).$$

• If ν is a vertex and $\underline{M}$ has already been constructed on all of the vertices and edges of $\mathcal{G}^{<\nu}$, then M_ν is the free cover of

$$\text{Image}\left[\Gamma(\underline{M}, \mathcal{L}^-(\nu)) \longleftarrow \Gamma(\underline{M}, E^{<\nu}(\mathcal{G}))\right]$$

and the maps $m_{\nu\ell}$ are determined by

$$M_\nu \xrightarrow{F} \text{Image}\left[\Gamma(\underline{M}, \mathcal{L}^-(\nu)) \longleftarrow \Gamma(\underline{M}, E^{<\nu}(\mathcal{G}'))\right] \subset \bigoplus_{\ell \in \mathcal{L}^-(\nu)} M_\ell$$

where F is the free cover map.

⟨**5.3**⟩ **Remark.** As constructed, M_ν is the smallest free module mapping surjectively to $\text{Image}\,[\Gamma(\underline{M}, \mathcal{L}^-(\nu)) \longleftarrow \Gamma(\underline{M}, E^{<\nu}(\mathcal{G}'))]$. By the discussion above, this surjectivity is what we need for ϕ to be a perfect Morse function.

⟨**5.4**⟩ **Proposition.** $\text{Image}\,[\Gamma(\underline{M}, \mathcal{L}^-(\nu)) \longleftarrow \Gamma(\underline{M}, E^{<\nu}(\mathcal{G}'))]$ has the canonical free cover property, as defined in §4.5.6 below, so the construction of M_ν makes sense.

⟨**5.5**⟩ **Exercise.** Show that the induction is possible, i.e. that these three rules determine the sheaf $\underline{M}$ everywhere on $\mathcal{G}$.

⟨**5.6**⟩ **Free covers.** Suppose that R is a graded ring and $\mathcal{M}$ is a finitely generated R-module. A *free cover* of $\mathcal{M}$ is a surjection $\mathcal{F} \xrightarrow{F} \mathcal{M}$ where $\mathcal{F}$ is free a set of generators, and such a surjection can't be found with fewer generators.

Free covers are unique in the following sense: If $\mathcal{F} \xrightarrow{F} \mathcal{M}$ and $\mathcal{F}' \xrightarrow{F'} \mathcal{M}$ are two free covers, then there is always a commutative diagram

$$\begin{array}{ccc} \mathcal{F} & \xrightarrow{i} & \mathcal{F}' \\ & \text{F}\searrow \quad \swarrow\text{F'} & \\ & \mathcal{M} & \end{array}$$

where i is an isomorphism. In general, however, free covers are not functorial since i is not uniquely determined by the commutativity of this diagram.

Definition. The module $\mathcal{M}$ has *the canonical free cover property* if there is exactly one map i making the diagram above commutative.

⟨**5.7**⟩ **Exercise.** Consider the 2-dimensional module $\mathcal{M}$ over the polynomial ring in one variable $\mathbb{R}[t]$ generated by elements g_1 and g_2, both of degree zero, with the relations $tg_i = 0$. Show that its free cover is the module freely generated by g_1 and g_2, and has the canonical free cover property. On the other hand, consider the module generated by g_1 in degree zero and g_2 in degree 2 with the same relation $tg_i = 0$. Show that its free cover is the free module generated by g_0 and g_2, but that it does not have the canonical free covers property because the map determined by $g_1 \mapsto g_1$ and $g_2 \mapsto g_2 + at^2 g_1$ commutes with the map to $\mathcal{M}$ for any real number a.

⟨**5.8**⟩ **Exercise.** Suppose that $\mathcal{M}$ is generated by homogeneous generators $g_1, g_2, \ldots$ with homogeneous relations $r_1, r_2, \ldots$ and that the degree of each of the r_i is greater than the degree of any of the g_i. Show that $\mathcal{M}$ has the canonical free cover property.

This is the reason for Proposition 4.5.4. In the second example of the last exercise, the relation $tg_1 = 0$ has degree 1, while the generator g_2 has degree 2.

⟨**5.9**⟩ **Exercise.** Carry out the inductive construction for the inverted V graph and the Egyptian pyramid. In both cases, $\underline{M}$ will coincide with $\underline{A}$ until the top vertex. At the top vertex ν, for the inverted V graph, M_ν will be as in the first example of Exercise 4.5.8 above. For the Egyptian pyramid, it will have a generator in degree 2 reflecting the phenomenon for linear functions explained in §4.4.6.

4.6. The Main Results

⟨**6.1**⟩ **Theorem.** [**7**]. Suppose that $T \circlearrowright Y$ is a generalized flag manifold as in §4.7.1 and that $T \circlearrowright X \subset T \circlearrowright Y$ is a generalized Schubert variety as in §4.7.3, and $\mathcal{G}(T \circlearrowright X)$ is its moment graph. Then the equivariant intersection cohomology of $T \circlearrowright X$ is the $\mathcal{O}(\mathbb{T})$ module of global sections of the sheaf $\underline{M}$ on its moment graph

$$IH^*(T \circlearrowright X) = \Gamma(\underline{M}, \mathcal{G}(T \circlearrowright X))$$

and the intersection homology of X is given by

$$IH^*(X) = IH^*(T \circlearrowright X) \otimes_{\mathcal{O}(\mathbb{T})} \mathbb{C} = \Gamma(\underline{M}, \mathcal{G}(T \circlearrowright X)) \otimes_{\mathcal{O}(\mathbb{T})} \mathbb{C}$$

⟨**6.2**⟩ **Theorem.** All Morse functions on $\mathcal{G}(\mathcal{H}, v)$ are perfect for the sheaf $\underline{M}$.

⟨**6.3**⟩ **Theorem.** $\Gamma(\underline{M}, \mathcal{G}(\mathcal{H}, v))$ is a free module over $\mathcal{O}(\mathbb{T})$.

Let IB_i be the number of free generators in degree i.

⟨**6.4**⟩ **Theorem.** $\Gamma(\underline{M}, \mathcal{G}(\mathcal{H}, v))$ satisfies Poincaré duality: $IB_i = IB_{-i}$. Moreover, the canonical pairing

$$\Gamma(\underline{M}, \mathcal{G}(\mathcal{H}, v)) \otimes_{\mathcal{O}(\mathbb{T})} \Gamma(\underline{M}, \mathcal{G}(\mathcal{H}, v)) \quad \longrightarrow \quad \mathcal{O}(\mathbb{T})$$

is perfect in the sense that the induced map

$$\Gamma(\underline{M}, \mathcal{G}(\mathcal{H}, v)) \longrightarrow \mathrm{Hom}(\Gamma(\underline{M}, \mathcal{G}(\mathcal{H}, v)), \mathcal{O}(\mathbb{T}))$$

is an isomorphism of $\mathcal{O}(\mathbb{T})$ modules.

4.7.* Flag Varieties and Generalized Schubert Varieties

⟨**7.1**⟩ **Generalized flag manifolds.** Suppose that G is a connected compact Lie group, T is its maximal torus, and H is a connected compact subgroup of G with the same maximal torus T. Then G/H is a *generalized flag manifold*.

⟨**7.2**⟩ This relates to the linear graph $\mathcal{G}(\mathcal{H}, v)$ constructed in §4.1.2 as follows: If $\mathcal{H}$ is crystallographic §4.1.4, then we can construct a compact group G such whose coroots correspond to $\mathcal{H}$ and whose Weyl group is the reflection group. We can construct a subgroup H corresponding to the choice of $v \in \mathbb{V}$. (It depends only on which hyperplanes in $\mathcal{H}$ the element v lies on.) Then the moment graph of G/H will be $\mathcal{G}(\mathcal{H}, v)$. In the case where v lies on no hyperplane in $\mathcal{H}$, P is the Borel subgroup, the abstract graph corresponding to $\mathcal{G}(\mathcal{H}, v)$ is called the Bruhat graph.

⟨**7.3**⟩ **Generalized Schubert varieties.** The generalized flag manifold G/H also has a description as $G_{\mathbb{C}}/P$ where $G_{\mathbb{C}}$ is the complexification of G and P is a parabolic subgroup.

Suppose we have a Morse function $\phi : \mathbb{V} \longrightarrow \mathbb{R}$ and an upward saturated subgraph $\mathcal{G}'$ of $\mathcal{G}(\mathcal{H}, v)$. We can approximate the linear function by a rational one,

without changing the fact that $\mathcal{G}'$ is upward saturated. Then ϕ is a coweight, so it corresponds to a map $\chi : \mathbb{C}^* \longrightarrow T_{\mathbb{C}}$. Now, the generalized Schubert variety whose moment graph is $\mathcal{G}'$ will be those $p \in G_{\mathbb{C}}/P$ such that $\lim_{\lambda \to 0} \chi(\lambda)p$ is a fixed point of $T \circlearrowright (G_{\mathbb{C}}/P)$ corresponding to one of the vertices in $\mathcal{G}'$.

There is an enormous literature on the manifolds $G_{\mathbb{C}}/P$ and on Schubert varieties and their singularities. For introductions with a combinatorial slant, see [**5**] and [**8**].

⟨**7.4**⟩ **A mystery.** What about non-crystallographic configurations of reflecting hyperplanes $\mathcal{H}$? For example, k lines at equal angles in $\mathbb{V} = \mathbb{R}^2$ where $k \neq 1, 2, 3, 4$, or 6 is non-crystallographic. Other examples are constructed out of the icosahedron. All purely graph theoretic theorems we have stated work just as well for these examples. But there is no known topological space corresponding to them. It is reasonable to conjecture that all of the results of this Lecture (in particular the crucial Proposition 4.5.4) hold for the graphs $\mathcal{G}(\mathcal{H}, v)$ associated to non-crystallographic $\mathcal{H}$. Better yet, is there some sort of topological object whose "moment graph" is $\mathcal{G}(\mathcal{H}, v)$?

Another problem is to find general sufficient conditions on a linear graph $\mathcal{G}$ that hold for graphs of the form $\mathcal{G}(\mathcal{H}, v)$, so that the results of this Lecture work for $\mathcal{G}$.

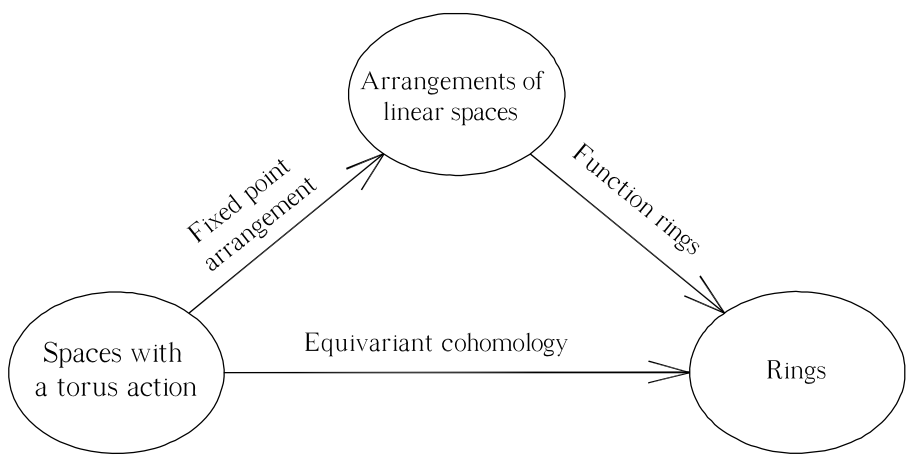
⟨**7.5**⟩ If P is a polyhedron that is rational but not simple, then there is a construction of a sheaf on its 1-skeleton that has a similar relation to equivariant intersection cohomology to the one discussed in this lecture [**4**],[**24**],[**1**],[**2**],[**3**]. A similar mystery applies to 1-skeleta of non-rational polyhedra. This has been the subject of a lot of recent work, e.g. [**9**].

LECTURE 5

Cohomology as Functions on a Variety

(Geometry of Subspace Arrangements)

In this lecture, we will describe another paradigm for computing equivariant cohomology via geometry. This paradigm can be described by the following directed graph:



One advantage is that this paradigm treats certain spaces whose 1-skeleton is not a balloon sculpture, as was required up until now. But the main point is that the arrangements of linear spaces that arise in this way seem interesting in themselves.

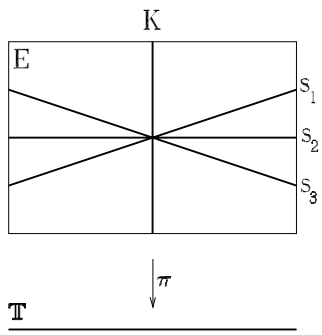
5.1. The Fixed Point Arrangement

⟨1.1⟩ Arrangements of sections. An arrangement of sections is a diagram of vector spaces

$$\begin{tikzcd} E \arrow[d, shift right, "\pi"'] \\ \mathbb{T} \arrow[u, shift right, bend right, "s_1"'] \arrow[u, shift right, bend right=60, "s_2"'] \arrow[u, shift right, bend right=80, "\cdots"'] \end{tikzcd}$$

where $\pi : E \longrightarrow \mathbb{T}$ is a surjection and $s_j : \mathbb{T} \longrightarrow E$ are sections, i.e. maps satisfying the section relations: $\pi(s_j(t)) = t$ for all $t \in \mathbb{T}$, i.e. $\pi \circ s_j$ is the identity on $\mathbb{T}$.

⟨**1.2**⟩ Suppose we have fixed $E \xrightarrow{\pi} \mathbb{T}$. To give a section s_i is the same thing as to give its image $s_i(\mathbb{T})$ as a subspace of E. So an arrangement of sections is equivalent to an arrangement of subspaces of E, each of which is transverse to K, the kernel of π.



⟨**1.3**⟩ **A diagram of spaces.** Now suppose we have a space with a torus action $T \circlearrowright X$ with finitely many fixed points $F_1, F_2, \ldots$ Then we have the following diagram of spaces with a T action and T equivariant maps:

$$\begin{tikzcd} X \arrow[d, shift right, "p"'] \\ \mathrm{pt} \arrow[u, shift right, bend right, "i_1"'] \arrow[u, bend right=60, "i_2 \cdots"'] \end{tikzcd}$$

here pt is a point (with a trivial T action), $p : X \longrightarrow \mathrm{pt}$ is the only thing it could be, and $i_j : \mathrm{pt} \longrightarrow X$ sends pt to the fixed point F_j.

⟨**1.4**⟩ **The fixed point arrangement.** We apply second equivariant homology functor $H_2(\cdot)$ to this diagram.

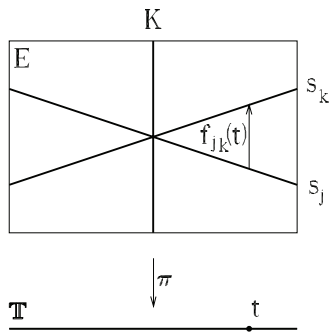
Definition. The *fixed point arrangement* of X is the arrangement of sections $E = H_2(T \circlearrowright X)$, $\mathbb{T} = H_2(T \circlearrowright \mathrm{pt})$, $\pi = p_* : H_2(T \circlearrowright X) \longrightarrow H_2(T \circlearrowright \mathrm{pt})$ and $s_j = (i_j)_* : H_2(T \circlearrowright \mathrm{pt}) \longrightarrow H_2(T \circlearrowright X)$.

5.2. How to Compute the Fixed Point Arrangement

The definition of the fixed point arrangement of $T \circlearrowright X$ is difficult to work with, because we don't usually have a good handle on the second equivariant homology group. In this section, we give a way to compute the fixed point arrangement without using equivariant homology.

⟨**2.1**⟩ **The spaces K and $\mathbb{T}$.** The space $H_2(T \circlearrowright \mathrm{pt})$ is what we've been calling $\mathbb{T}$ all along: the vector space such that $\mathbb{T}/L = T$ or, otherwise put, the Lie algebra of T. If X has only even dimensional homology, then the space K is $H_2(X)$, the ordinary second homology group of X. (This follows from equivariant formality, see [**13**].)

⟨**2.2**⟩ **How to specify an arrangement of sections.** If we have an arrangement of sections, for every ordered pair of sections s_j and s_k, we have a linear map $f_{jk} : \mathbb{T} \longrightarrow K$ which associates to $t \in \mathbb{T}$ the vector $s_k(t) - s_k(t)$ as in this picture:



Conversely, if we know the linear functions f_{jk}, that determines the configuration up to automorphisms of E commuting with π, which is enough for our purposes. In fact, since $f_{jk} + f_{kl} = f_{kl}$, if we're really efficient about it, we only need to know $m - 1$ of them, where m is the number of fixed points.

⟨**2.3**⟩ **Determining f_{kj}.** Suppose that F_j and F_k lie in a balloon B: a 2-sphere that is taken into itself by T. Since there are only finitely many fixed points, T the orbits of T on B will all be circles, except for F_j and F_k. Chose such a circle, S. There will be a linear map $g : T \longrightarrow \mathbb{R}$ and an identification of S with $\mathbb{R}/\mathbb{Z}$ such that the action of $\bar{t}$ corresponds to addition of $g(t)$. The trouble is that there are two such functions g which are negatives of each other, corresponding to opposite identifications of S with $\mathbb{R}/\mathbb{Z}$. If we had an orientation of S, we could specify that the orientation should correspond to the natural orientation of $\mathbb{R}$.

Choose an orientation $\mathbb{O}$ for B. That does two things for us. First, it makes B into a cycle, so it gives us a class $[B]$ in homology. Second, it restricts to an orientation of the disk bounded by S containing F_j. This induces an orientation on its boundary, S, solving the problem above. With these conventions, we get

$$f_{jk} = g[B]$$

⟨**2.4**⟩ **Exercise.** Show that this definition is independent of the orientation $\mathbb{O}$ of B chosen. Show directly from this definition that $f_{jk} = -f_{kj}$.

5.3. The Main Result

⟨**3.1**⟩ **The function ring of the configuration.** Let A be the union of the linear subspaces in the fixed point configuration. In other words, $a \in A$ if and only if $a = s_j(t)$ for some section s_j in our arrangement and some $t \in \mathbb{T}$. The set A is a real algebraic variety - it has a function ring $\mathcal{O}(A)$, which is the polynomial functions on E, two being considered equivalent if they take the same values on every point of A. In other words,

$$\mathcal{O}(A) = \frac{\mathcal{O}(E)}{I(A)}$$

where $I(A)$ is the ideal of polynomials vanishing on the set A.

⟨**3.2**⟩ The map $\pi : E \longrightarrow \mathbb{T}$ provides an inclusion $\mathcal{O}(\mathbb{T}) \subset \mathcal{O}(A)$. An element f of $\mathcal{O}(\mathbb{T})$ is a polynomial function $\mathbb{T} \xrightarrow{f} \mathbb{R}$ so the composition $A \xrightarrow{\pi} \mathbb{T} \xrightarrow{f} \mathbb{R}$ is in $\mathcal{O}(A)$.

⟨**3.3**⟩ **Theorem.** [**17**] Suppose that $H^*(X)$ is generated as a ring by $H^2(X)$. Then the equivariant cohomology of X is the function ring of the fixed point configuration of X

$$H^*(T \circlearrowright X) = \mathcal{O}(A).$$

The ring $\mathcal{O}(A)$ is a free graded module over the polynomial algebra $\mathcal{O}(\mathbb{T})$. As in §3.8.3,

$$H^*(X) = \frac{\mathcal{O}(A)}{\text{the ideal generated by } \mathcal{O}(\mathbb{T})^{>0}}$$

where $\mathcal{O}(\mathbb{T})^{>0}$ is the positive degree part of $\mathcal{O}(\mathbb{T})$.

⟨**3.4**⟩ **Remark.** The fact that $\mathcal{O}(A)$ is a free module over $\mathcal{O}(\mathbb{T})$ is an interesting and mysterious property of the configuration. Here it follows from theorems in topology. We don't know any geometric characterization of which configurations of sections have this property.

5.4. Springer Varieties

Suppose that $k = k_1 + \cdots + k_j$ is a partition of the integer k. Let γ be the $k \times k$ matrix whose Jordan blocks have size $k_1, \ldots, k_j$. For example if the partition is $6 = 3 + 2 + 1$, then γ is given by

$$\gamma = \begin{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} & & \\ & \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} & \\ & & [0] \end{bmatrix}$$

⟨**4.1**⟩ **Definition.** The *Springer variety* X_γ of γ is the set of complete flags $0 \subset F^1 \subset F^2 \subset \cdots \subset F^{k-1} \subset \mathbb{C}^k$, where F^i is an i-dimensional subspace of $\mathbb{C}^k$, with the property that γ takes each space F^k into itself.

⟨**4.2**⟩ **The torus action on** X_γ. There is a torus T that acts on X_γ. This is the set of diagonal matrices a with the following properties: the determinant of a is 1; the entries of a are complex numbers with absolute value 1; and furthermore the first k_1 diagonal entries are equal to each other; the next k_2 diagonal entries are equal to each other, and so on. For example, for our partition $6 = 3 + 2 + 1$, T consists of the matrices

$$a = \begin{bmatrix} \begin{bmatrix} a_1 & 0 & 0 \\ 0 & a_1 & 0 \\ 0 & 0 & a_1 \end{bmatrix} & & \\ & \begin{bmatrix} a_2 & 0 \\ 0 & a_2 \end{bmatrix} & \\ & & [a_3] \end{bmatrix} \quad \text{such that} \quad |a_i| = 1 \quad \text{and} \quad \det a = 1.$$

⟨**4.3**⟩ **Exercise.** Prove that the torus T preserves the Springer variety X_γ. Use the fact that the matrix t commutes with the matrix γ.

⟨**4.4**⟩ **The arrangement of subspaces.** The action $T \circlearrowright X_\gamma$ of the torus T on the Springer variety satisfies the hypotheses of Theorem 5.3.3 (see [**17**]). (In general, the 1-skeleton of $T \circlearrowright X_\gamma$ is too complicated to satisfy the hypotheses §3.8.1 of Theorem 3.8.2, so the methods of Lecture 3 don't apply to Springer varieties.)

The subspace arrangement A for the Springer variety is quite beautiful configuration that depends, of course, only on the given partition.

The Lie algebra $\mathbb{T}$ of T is the set of $k \times k$ diagonal real matrices with trace 0 such that the first k_1 diagonal entries are equal to each other, the next k_2 diagonal entries are equal to each other, and so on. For $6 = 3 + 2 + 1$, the elements t of $\mathbb{T}$ are

$$t = \begin{bmatrix} \begin{bmatrix} t_1 & 0 & 0 \\ 0 & t_1 & 0 \\ 0 & 0 & t_1 \end{bmatrix} & & \\ & \begin{bmatrix} t_2 & 0 \\ 0 & t_2 \end{bmatrix} & \\ & & [t_3] \end{bmatrix} \quad \text{such that} \quad t_i \quad \text{is real and} \quad \operatorname{trace} t = 0$$

The space $K = H_2(X_\gamma)$ is $(k-1)$-dimensional. It can be identified with the set of all $k \times k$ real matrices with trace 0. (This holds for all partitions except for the trivial partition $k = k$. For the trivial partition, the Springer variety X_γ is just a point, so there isn't much to study.)

There are $k!/(k_1! \cdots k_j!)$ fixed points of the action of $\mathbb{T}$ on X_γ, so the arrangement A consists of $k!/(k_1! \cdots k_j!)$ linear subspaces of $E = \mathbb{T} \oplus K$. Each of these is the graph of a linear map m from $\mathbb{T}$ to K. The linear maps m are constructed as follows: Write a diagonal matrix so that t_1 occurs k_1 times, t_2 occurs k_2 times, etc. There are $k!/(k_1! \cdots k_j!)$ ways to do this. For our partition $6 = 3 + 2 + 1$, here's one of them:

$$\begin{bmatrix} t_1 & & & & & \\ & t_3 & & & & \\ & & t_1 & & & \\ & & & t_2 & & \\ & & & & t_1 & \\ & & & & & t_2 \end{bmatrix}$$

This provides our map m from $\mathbb{T}$ to K.

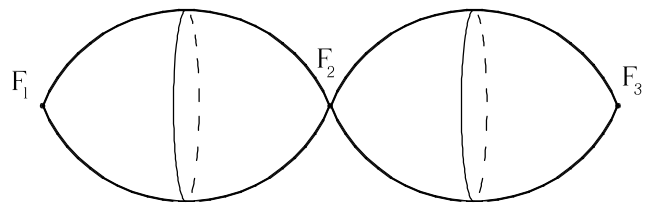
⟨**4.5**⟩ **The Springer action.** The symmetric group S_k on k letters acts on the configuration A of linear subspaces (by permuting the diagonal entries in the last matrix). It follows that S_k acts on the equivariant cohomology of X_γ ([**17**]). The S_k action on A preserves the fibers of the map π to $\mathbb{T}$. Therefore, it passes to an action of S_k on the ordinary cohomology of X_γ. This is the usual Springer action.

A similar equivariant cohomology construction for Springer actions on the cohomology of Springer varieties for loop groups is found in [**14**]. Historically, the effort to solve the problem addressed in [**14**] was what originally led to the whole body of material in this lecture series.

⟨**4.6**⟩ **Example.** Suppose we start with the partition $3 = 2+1$ so γ is the nilpotent matrix

$$\gamma = \begin{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} & \\ & [0] \end{bmatrix}.$$

The Springer variety X_γ consists of two 2-spheres joined together. The T action has 3 fixed points F_1, F_2, and F_3. The picture looks like this:



The Springer variety X_γ

(In this case, the 1 skeleton is a balloon sculpture, so Theorem 3.8.2 applies as well.)

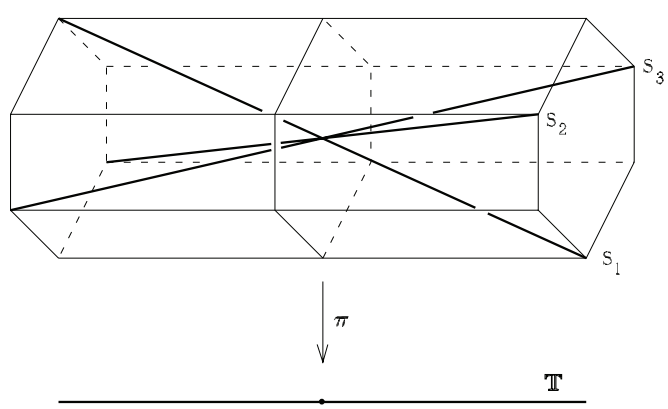
The Lie algebra $\mathbb{T}$ is 1-dimensional. It is the set of real matrices

$$\begin{bmatrix} \begin{bmatrix} t_1 & 0 \\ 0 & t_1 \end{bmatrix} & \\ & [t_2] \end{bmatrix} \quad \text{such that} \quad 2t_1 + t_2 = 0$$

The space K is 2-dimensional. It is the space of diagonal matrices of trace 0. The three maps from $\mathbb{T}$ to K corresponding to the three fixed points are given by

$$\begin{bmatrix} t_1 & & \\ & t_1 & \\ & & t_2 \end{bmatrix}, \begin{bmatrix} t_1 & & \\ & t_2 & \\ & & t_1 \end{bmatrix}, \text{ and } \begin{bmatrix} t_2 & & \\ & t_1 & \\ & & t_1 \end{bmatrix}$$

The resulting configuration of sections looks like this:



Note that this configuration has an obvious action of S_3, the symmetric group which permutes the 3 sections. Thus we get an action of S_3 on the equivariant cohomology of X_γ and also the ordinary cohomology. This is not induced by an action of S_3 on X_γ itself: X_γ is less symmetric than its fixed point configuration.

⟨**4.7**⟩ **Exercise.** Verify the statements in the last section about the Springer variety for the partition $3 = 2 + 1$.

5.5. Relation with Lecture 3

Suppose $T \circlearrowright X$ satisfies both the hypotheses for Theorem 5.3.3 above and for Theorem 3.8.2. This happens, for example, with flag manifolds and toric varieties for simple polytopes. What is the relation between the two pictures?

Given the moment graph $\mathcal{G}(X \circlearrowright T)$ we will construct the configuration of sections.

Suppose the graph has k vertices $\nu_1, \ldots, \nu_k$. The space of all graphs equivalent to $\mathcal{G}(X \circlearrowright T)$ embeds in $\mathbb{V}^k$ by the position of the k vertices. The space $E^* = H^2(T \circlearrowright X)$ is the closure. Also the projections on the factors give k maps η_j from E^* to $\mathbb{V}$. The diagonal gives a map $\Delta : \mathbb{V} \longrightarrow E^*$. As we have assumed from the beginning, $\mathbb{V} = \mathbb{T}^*$. In summary, the moment graph $\mathcal{G}(X \circlearrowright T)$ gives us a diagram

$$\begin{tikzcd} E^* \arrow[d, shift left, "\eta_1"] \arrow[d, shift right] \arrow[d, bend left=60, "\eta_2"] \arrow[d, bend right=60] \cdots \\ \mathbb{T}^* \arrow[u, "\Delta"] \end{tikzcd}$$

If we dualize this diagram, we get the arrangement of sections of §5.1.1.

⟨**5.1**⟩ **Exercise.** In this situation, construct a map

$$\mathcal{O}(A) \longrightarrow \mathcal{H}^*(\mathcal{G}(T \circlearrowright X)).$$

Assuming that $\mathcal{H}^*(\mathcal{G}(T \circlearrowright X))$ is generated by $\mathcal{H}^2(\mathcal{G}(T \circlearrowright X))$, show that this map is an isomorphism.

BIBLIOGRAPHY

1. G. Barthel, J.-P. Brasselet, K.-H. Fieseler, L. Kaup, *Combinatorial Intersection Cohomology for Fans*, Tohoku Math. J. 54 (2002), 1-41.
2. G. Barthel, J.-P. Brasselet, K.-H. Fieseler, L. Kaup, *Equivariant Intersection Cohomology of Toric Varieties*, Contemp. Math. 241, Amer. Math. Soc. (1999), 45-68.
3. G. Barthel, J.-P. Brasselet, K.-H. Fieseler, L. Kaup, *Invariante Divisoren und Schnitthomologie von Torischen Varietäten*, Parameter Spaces (Warsaw, 1994), 9-223, Banach Center Pub. 36.
4. J. Bernstein and V. Lunts, Equivariant Sheaves and Functors, Springer Lect. Notes in Math. 1578 (1994).
5. S. Billey and Lakshmibai, *Singular Loci of Schubert Varieties*, Progress in Math. vol.182 Birkhauser Boston (2000).
6. T. Braden, *Perverse Sheaves on Grassmannians*, Canad. J. Math. 54 (2002) 493-532.
7. T. Braden and R. MacPherson, *From Moment Graphs to Intersection Cohomology*, Math. Ann. 321 (2001), 533-551.
8. F. Brenti, *Kazhdan-Lusztig Polynomials: History, Problems, and Combinatorial Invariance*, Seminaire Lotharingien de Combinatoire 49 (2003), 613-627.
9. P. Bressler and V. Lunts, *Intersection Cohomology on Nonrational Polytopes*, Compositio Math. 135 (2003), 245-278.
10. J.-L. Brylinski, *Equivariant Intersection Cohomology*, Contemp. Math. 139, Amer. Math. Soc., (1992) 5-32.
11. I. Gelfand, M. Goresky, R. MacPherson, V. Serganova, *Combinatorial Geometries, Grassmannians, and the Moment Map*, Advances in Math. 63 (1987) 301-316.
12. I. Gelfand and R. MacPherson *Geometry in Grassmannians and a Generalization of the Dilogarithm*, Advances in Math. 44(1982) 279-312.
13. M. Goresky, R. Kottwitz, and R. MacPherson, *Equivariant Cohomology, Koszul Duality, and the Localization Theorem*, Invent. Math. 131 (1998) 25-83.
14. M. Goresky, R. Kottwitz, and R. MacPherson, *Homology of Affine Springer Fibers in the Unramified Case*, Duke Math. J. 121(2004), 509-561.
15. M. Goresky and R. MacPherson, *Intersection Homology Theory*, Topology 49 (1980), 135-162.

16. M. Goresky and R. MacPherson, *On the topology of Algebraic Torus Actions*, Springer Lect. Notes in Math 1271 (1987) 73-90.
17. M. Goresky and R. MacPherson, *On the spectrum of the equivariant cohomology ring*, to appear.
18. V. Guillemin and K. Zara, *The Existence of Generating Families for the Cohomology Ring of a Graph,* Advances in Math. 174 (2003) 115-173.
19. V. Guillemin and K. Zara, *G-Actions on Graphs*, Int. Math. Res. Notes (2001),519-542.
20. V. Guillemin and K. Zara, 1-*Skeletons, Betti Numbers, and Equivariant Cohomology*, Duke Math. J. 107 (2001), 283-349.
21. V. Guillemin and K. Zara, *Equivariant de Rham Theory and Graphs*, Asian J. of Math. 3 (1999) 49-76.
22. T. Holm, Equivariant Cohomology, Homogeneous Spaces, and Graphs, Thesis, MIT, 2002.
23. R. Joshua, *Equivariant Intersection Cohomology — A Survey*, Contemp. Math. 88, Amer. Math. Soc., (1989), 25-31.
24. V. Lunts, *Equivariant Sheaves on Toric Varieties*, Compositio Math. 96 (1995) 63-83.

An Introduction to Hyperplane Arrangements

Richard P. Stanley

IAS/Park City Mathematics Series
Volume 14, 2004

An Introduction to Hyperplane Arrangements

Richard P. Stanley

LECTURE 1
Basic Definitions, the Intersection Poset and the Characteristic Polynomial

1.1. Basic Definitions

The following notation is used throughout for certain sets of numbers:

$\mathbb{N}$	nonnegative integers
$\mathbb{P}$	positive integers
$\mathbb{Z}$	integers
$\mathbb{Q}$	rational numbers
$\mathbb{R}$	real numbers
$\mathbb{R}_+$	positive real numbers
$\mathbb{C}$	complex numbers
$[m]$	the set $\{1, 2, \ldots, m\}$ when $m \in \mathbb{N}$

We also write $[t^k]\chi(t)$ for the coefficient of t^k in the polynomial or power series $\chi(t)$. For instance, $[t^2](1+t)^4 = 6$.

A *finite hyperplane arrangement* $\mathcal{A}$ is a finite set of affine hyperplanes in some vector space $V \cong K^n$, where K is a field. We will not consider infinite hyperplane arrangements or arrangements of general subspaces or other objects (though they have many interesting properties), so we will simply use the term *arrangement* for a finite hyperplane arrangement. Most often we will take $K = \mathbb{R}$, but as we will see

[1]Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.
E-mail address: rstan@math.mit.edu.
The author was supported in part by NSF grant DMS-9988459. He is grateful to Lauren Williams for her careful reading of the original manuscript and many helpful suggestions, and to Hélène Barcelo and Guangfeng Jiang for a number of of additional suggestions. Students in 18.315, taught at MIT during fall 2004, also made some helpful contributions.

©2007 American Mathematical Society

even if we're only interested in this case it is useful to consider other fields as well. To make sure that the definition of a hyperplane arrangement is clear, we define a *linear hyperplane* to be an $(n-1)$-dimensional subspace H of V, i.e.,

$$H = \{v \in V \,:\, \alpha \cdot v = 0\},$$

where α is a fixed nonzero vector in V and $\alpha \cdot v$ is the usual dot product:

$$(\alpha_1, \ldots, \alpha_n) \cdot (v_1, \ldots, v_n) = \sum \alpha_i v_i.$$

An *affine hyperplane* is a translate J of a linear hyperplane, i.e.,

$$J = \{v \in V \,:\, \alpha \cdot v = a\},$$

where α is a fixed nonzero vector in V and $a \in K$.

If the equations of the hyperplanes of $\mathcal{A}$ are given by $L_1(x) = a_1, \ldots, L_m(x) = a_m$, where $x = (x_1, \ldots, x_n)$ and each $L_i(x)$ is a homogeneous linear form, then we call the polynomial

$$Q_{\mathcal{A}}(x) = (L_1(x) - a_1) \cdots (L_m(x) - a_m)$$

the *defining polynomial* of $\mathcal{A}$. It is often convenient to specify an arrangement by its defining polynomial. For instance, the arrangement $\mathcal{A}$ consisting of the n coordinate hyperplanes has $Q_{\mathcal{A}}(x) = x_1 x_2 \cdots x_n$.

Let $\mathcal{A}$ be an arrangement in the vector space V. The *dimension* $\dim(\mathcal{A})$ of $\mathcal{A}$ is defined to be $\dim(V)$ $(= n)$, while the *rank* $\operatorname{rank}(\mathcal{A})$ of $\mathcal{A}$ is the dimension of the space spanned by the normals to the hyperplanes in $\mathcal{A}$. We say that $\mathcal{A}$ is *essential* if $\operatorname{rank}(\mathcal{A}) = \dim(\mathcal{A})$. Suppose that $\operatorname{rank}(\mathcal{A}) = r$, and take $V = K^n$. Let Y be a complementary space in K^n to the subspace X spanned by the normals to hyperplanes in $\mathcal{A}$. Define

$$W = \{v \in V \,:\, v \cdot y = 0 \ \forall y \in Y\}.$$

If $\operatorname{char}(K) = 0$ then we can simply take $W = X$. By elementary linear algebra we have

$$\operatorname{codim}_W(H \cap W) = 1 \tag{1}$$

for all $H \in \mathcal{A}$. In other words, $H \cap W$ is a hyperplane of W, so the set $\mathcal{A}_W := \{H \cap W \,:\, H \in \mathcal{A}\}$ is an essential arrangement in W. Moreover, the arrangements $\mathcal{A}$ and $\mathcal{A}_W$ are "essentially the same," meaning in particular that they have the same intersection poset (as defined in Definition 1.1). Let us call $\mathcal{A}_W$ the *essentialization* of $\mathcal{A}$, denoted $\operatorname{ess}(\mathcal{A})$. When $K = \mathbb{R}$ and we take $W = X$, then the arrangement $\mathcal{A}$ is obtained from $\mathcal{A}_W$ by "stretching" the hyperplane $H \cap W \in \mathcal{A}_W$ orthogonally to W. Thus if $W^\perp$ denotes the orthogonal complement to W in V, then $H' \in \mathcal{A}_W$ if and only if $H' \oplus W^\perp \in \mathcal{A}$. Note that in characteristic p this type of reasoning fails since the orthogonal complement of a subspace W can intersect W in a subspace of dimension greater than 0.

Example 1.1. Let $\mathcal{A}$ consist of the lines $x = a_1, \ldots, x = a_k$ in K^2 (with coordinates x and y). Then we can take W to be the x-axis, and $\operatorname{ess}(\mathcal{A})$ consists of the points $x = a_1, \ldots, x = a_k$ in K.

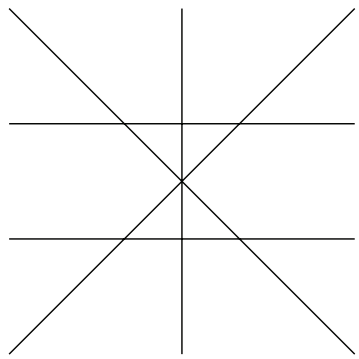
Now let $K = \mathbb{R}$. A *region* of an arrangement $\mathcal{A}$ is a connected component of the complement X of the hyperplanes:

$$X = \mathbb{R}^n - \bigcup_{H \in \mathcal{A}} H.$$

Let $\mathcal{R}(\mathcal{A})$ denote the set of regions of $\mathcal{A}$, and let

$$r(\mathcal{A}) = \#\mathcal{R}(\mathcal{A}),$$

the number of regions. For instance, the arrangement $\mathcal{A}$ shown below has $r(\mathcal{A}) = 14$.



It is a simple exercise to show that every region $R \in \mathcal{R}(\mathcal{A})$ is open and convex (continuing to assume $K = \mathbb{R}$), and hence homeomorphic to the interior of an n-dimensional ball $\mathbb{B}^n$ (Exercise 1). Note that if W is the subspace of V spanned by the normals to the hyperplanes in $\mathcal{A}$, then $R \in \mathcal{R}(\mathcal{A})$ if and only if $R \cap W \in \mathcal{R}(\mathcal{A}_W)$. We say that a region $R \in \mathcal{R}(\mathcal{A})$ is *relatively bounded* if $R \cap W$ is bounded. If $\mathcal{A}$ is essential, then relatively bounded is the same as bounded. We write $b(\mathcal{A})$ for the number of relatively bounded regions of $\mathcal{A}$. For instance, in Example 1.1 take $K = \mathbb{R}$ and $a_1 < a_2 < \cdots < a_k$. Then the relatively bounded regions are the regions $a_i < x < a_{i+1}$, $1 \le i \le k-1$. In $\mathrm{ess}(\mathcal{A})$ they become the (bounded) open intervals (a_i, a_{i+1}). There are also two regions of $\mathcal{A}$ that are not relatively bounded, viz., $x < a_1$ and $x > a_k$.

A (closed) *half-space* is a set $\{x \in \mathbb{R}^n : x \cdot \alpha \ge c\}$ for some $\alpha \in \mathbb{R}^n$, $c \in \mathbb{R}$. If H is a hyperplane in $\mathbb{R}^n$, then the complement $\mathbb{R}^n - H$ has two (open) components whose closures are half-spaces. It follows that the closure $\bar{R}$ of a region R of $\mathcal{A}$ is a finite intersection of half-spaces, i.e., a (convex) *polyhedron* (of dimension n). A bounded polyhedron is called a (convex) *polytope*. Thus if R (or $\bar{R}$) is bounded, then $\bar{R}$ is a polytope (of dimension n).

An arrangement $\mathcal{A}$ is in *general position* if

$$\begin{aligned} \{H_1, \ldots, H_p\} \subseteq \mathcal{A},\ p \le n \quad &\Rightarrow \quad \dim(H_1 \cap \cdots \cap H_p) = n - p \\ \{H_1, \ldots, H_p\} \subseteq \mathcal{A},\ p > n \quad &\Rightarrow \quad H_1 \cap \cdots \cap H_p = \emptyset. \end{aligned}$$

For instance, if $n = 2$ then a set of lines is in general position if no two are parallel and no three meet at a point.

Let us consider some interesting examples of arrangements that will anticipate some later material.

Example 1.2. Let $\mathcal{A}_m$ consist of m lines in general position in $\mathbb{R}^2$. We can compute $r(\mathcal{A}_m)$ using the *sweep hyperplane* method. Add a L line to $\mathcal{A}_k$ (with $\mathcal{A}_K \cup \{L\}$ in general position). When we travel along L from one end (at infinity) to the other, every time we intersect a line in $\mathcal{A}_k$ we create a new region, and we create one new

region at the end. Before we add any lines we have one region (all of $\mathbb{R}^2$). Hence

$$\begin{aligned} r(\mathcal{A}_m) &= \#\text{intersections} + \#\text{lines} + 1 \\ &= \binom{m}{2} + m + 1. \end{aligned}$$

Example 1.3. The *braid arrangement* $\mathcal{B}_n$ in K^n consists of the hyperplanes

$$\mathcal{B}_n : \quad x_i - x_j = 0, \quad 1 \le i < j \le n.$$

Thus $\mathcal{B}_n$ has $\binom{n}{2}$ hyperplanes. To count the number of regions when $K = \mathbb{R}$, note that specifying which side of the hyperplane $x_i - x_j = 0$ a point $(a_1, \ldots, a_n)$ lies on is equivalent to specifying whether $a_i < a_j$ or $a_i > a_j$. Hence the number of regions is the number of ways that we can specify whether $a_i < a_j$ or $a_i > a_j$ for $1 \le i < j \le n$. Such a specification is given by imposing a linear order on the a_i's. In other words, for each permutation $w \in \mathfrak{S}_n$ (the symmetric group of all permutations of $1, 2, \ldots, n$), there corresponds a region R_w of $\mathcal{B}_n$ given by

$$R_w = \{(a_1, \ldots, a_n) \in \mathbb{R}^n \,:\, a_{w(1)} > a_{w(2)} > \cdots > a_{w(n)}\}.$$

Hence $r(\mathcal{B}_n) = n!$. Rarely is it so easy to compute the number of regions!

Note that the braid arrangement $\mathcal{B}_n$ is not essential; indeed, $\operatorname{rank}(\mathcal{B}_n) = n-1$. When $\operatorname{char}(K)$ does not divide n the space $W \subseteq K^n$ of equation (1) can be taken to be

$$W = \{(a_1, \ldots, a_n) \in K^n \,:\, a_1 + \cdots + a_n = 0\}.$$

The braid arrangement has a number of "deformations" of considerable interest. We will just define some of them now and discuss them further later. All these arrangements lie in K^n, and in all of them we take $1 \le i < j \le n$. The reader who likes a challenge can try to compute their number of regions when $K = \mathbb{R}$. (Some are much easier than others.)

- *generic braid arrangement*: $x_i - x_j = a_{ij}$, where the a_{ij}'s are "generic" (e.g., linearly independent over the prime field, so K has to be "sufficiently large"). The precise definition of "generic" will be given later. (The *prime field* of K is its smallest subfield, isomorphic to either $\mathbb{Q}$ or $\mathbb{Z}/p\mathbb{Z}$ for some prime p.)
- *semigeneric braid arrangement*: $x_i - x_j = a_i$, where the a_i's are "generic."
- *Shi arrangement*: $x_i - x_j = 0, 1$ (so $n(n-1)$ hyperplanes in all).
- *Linial arrangement*: $x_i - x_j = 1$.
- *Catalan arrangement*: $x_i - x_j = -1, 0, 1$.
- *semiorder arrangement*: $x_i - x_j = -1, 1$.
- *threshold arrangement*: $x_i + x_j = 0$ (not really a deformation of the braid arrangement, but closely related).

An arrangement $\mathcal{A}$ is *central* if $\bigcap_{H \in \mathcal{A}} H \neq \emptyset$. Equivalently, $\mathcal{A}$ is a translate of a *linear arrangement* (an arrangement of linear hyperplanes, i.e., hyperplanes passing through the origin). Many other writers call an arrangement central, rather than linear, if $0 \in \bigcap_{H \in \mathcal{A}} H$. If $\mathcal{A}$ is central with $X = \bigcap_{H \in \mathcal{A}} H$, then $\operatorname{rank}(\mathcal{A}) = \operatorname{codim}(X)$. If $\mathcal{A}$ is central, then note also that $b(\mathcal{A}) = 0$ [why?].

There are two useful arrangements closely related to a given arrangement $\mathcal{A}$. If $\mathcal{A}$ is a linear arrangement in K^n, then *projectivize* $\mathcal{A}$ by choosing some $H \in \mathcal{A}$

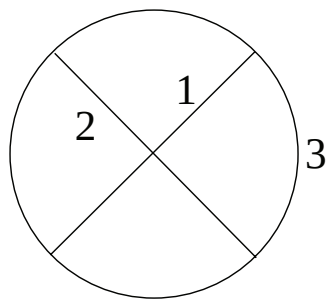
to be the hyperplane at infinity in projective space P_K^{n-1}. Thus if we regard

$$P_K^{n-1} = \{(x_1, \ldots, x_n) : x_i \in K, \text{ not all } x_i = 0\}/\sim,$$

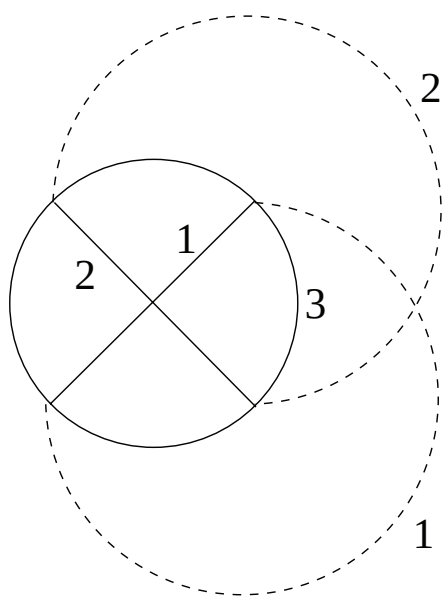
where $u \sim v$ if $u = \alpha v$ for some $0 \neq \alpha \in K$, then

$$H = (\{(x_1, \ldots, x_{n-1}, 0) : x_i \in K, \text{ not all } x_i = 0\}/\sim) \cong P_K^{n-2}.$$

The remaining hyperplanes in $\mathcal{A}$ then correspond to "finite" (i.e., not at infinity) projective hyperplanes in P_K^{n-1}. This gives an arrangement $\mathrm{proj}(\mathcal{A})$ of hyperplanes in P_K^{n-1}. When $K = \mathbb{R}$, the two regions R and $-R$ of $\mathcal{A}$ become identified in $\mathrm{proj}(\mathcal{A})$. Hence $r(\mathrm{proj}(\mathcal{A})) = \frac{1}{2}r(\mathcal{A})$. When $n = 3$, we can draw $P_{\mathbb{R}}^2$ as a disk with antipodal boundary points identified. The circumference of the disk represents the hyperplane at infinity. This provides a good way to visualize three-dimensional real linear arrangements. For instance, if $\mathcal{A}$ consists of the three coordinate hyperplanes $x_1 = 0$, $x_2 = 0$, and $x_3 = 0$, then a projective drawing is given by



The line labelled i is the projectivization of the hyperplane $x_i = 0$. The hyperplane at infinity is $x_3 = 0$. There are four regions, so $r(\mathcal{A}) = 8$. To draw the incidences among all eight regions of $\mathcal{A}$, simply "reflect" the interior of the disk to the exterior:



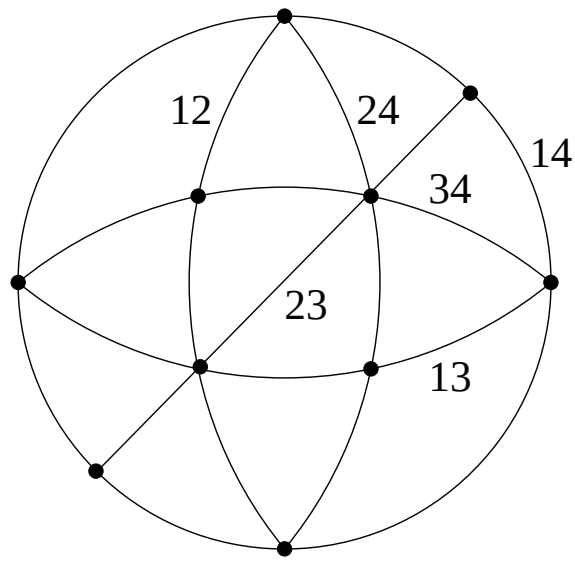


Figure 1. A projectivization of the braid arrangement $\mathcal{B}_4$

Regarding this diagram as a planar graph, the dual graph is the 3-cube (i.e., the vertices and edges of a three-dimensional cube) [why?].

For a more complicated example of projectivization, Figure 1 shows $\mathrm{proj}(\mathcal{B}_4)$ (where we regard $\mathcal{B}_4$ as a three-dimensional arrangement contained in the hyperplane $x_1 + x_2 + x_3 + x_4 = 0$ of $\mathbb{R}^4$), with the hyperplane $x_i = x_j$ labelled ij, and with $x_1 = x_4$ as the hyperplane at infinity.

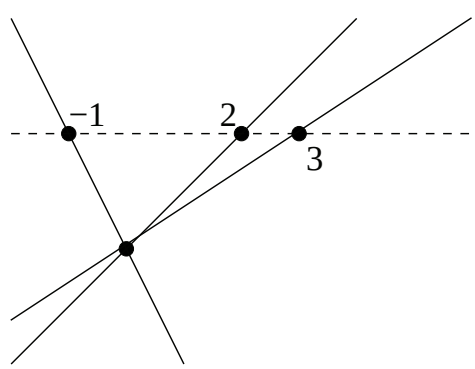
We now define an operation which is "inverse" to projectivization. Let $\mathcal{A}$ be an (affine) arrangement in K^n, given by the equations

$$L_1(x) = a_1, \quad \ldots, \quad L_m(x) = a_m.$$

Introduce a new coordinate y, and define a central arrangement $c\mathcal{A}$ (the *cone* over $\mathcal{A}$) in $K^n \times K = K^{n+1}$ by the equations

$$L_1(x) = a_1 y, \quad \ldots, \quad L_m(x) = a_m y, \quad y = 0.$$

For instance, let $\mathcal{A}$ be the arrangement in $\mathbb{R}^1$ given by $x = -1$, $x = 2$, and $x = 3$. The following figure should explain why $c\mathcal{A}$ is called a cone.



It is easy to see that when $K = \mathbb{R}$, we have $r(c\mathcal{A}) = 2r(\mathcal{A})$. In general, $c\mathcal{A}$ has the "same combinatorics as $\mathcal{A}$, times 2." See Exercise 2.1.

1.2. The Intersection Poset

Recall that a *poset* (short for partially ordered set) is a set P and a relation $\leq$ satisfying the following axioms (for all $x, y, z \in P$):

(P1) (reflexivity) $x \leq x$
(P2) (antisymmetry) If $x \leq y$ and $y \leq x$, then $x = y$.
(P3) (transitivity) If $x \leq y$ and $y \leq z$, then $x \leq z$.

Obvious notation such as $x < y$ for $x \leq y$ and $x \neq y$, and $y \geq x$ for $x \leq y$ will be used throughout. If $x \leq y$ in P, then the (closed) *interval* $[x, y]$ is defined by

$$[x, y] = \{z \in P : x \leq z \leq y\}.$$

Note that the empty set $\emptyset$ is *not* a closed interval. For basic information on posets not covered here, see [**31**].

Definition 1.1. Let $\mathcal{A}$ be an arrangement in V, and let $L(\mathcal{A})$ be the set of all *nonempty* intersections of hyperplanes in $\mathcal{A}$, including V itself as the intersection over the empty set. Define $x \leq y$ in $L(\mathcal{A})$ if $x \supseteq y$ (as subsets of V). In other words, $L(\mathcal{A})$ is partially ordered by *reverse* inclusion. We call $L(\mathcal{A})$ the *intersection poset* of $\mathcal{A}$.

NOTE. The primary reason for ordering intersections by reverse inclusion rather than ordinary inclusion is Proposition 3.8. We don't want to alter the well-established definition of a geometric lattice or to refer constantly to "dual geometric lattices."

The element $V \in L(\mathcal{A})$ satisfies $x \geq V$ for all $x \in L(\mathcal{A})$. In general, if P is a poset then we denote by $\hat{0}$ an element (necessarily unique) such that $x \geq \hat{0}$ for all $x \in P$. We say that y *covers* x in a poset P, denoted $x \lessdot y$, if $x < y$ and no $z \in P$ satisfies $x < z < y$. Every finite poset is determined by its cover relations. The (Hasse) *diagram* of a finite poset is obtained by drawing the elements of P as dots, with x drawn lower than y if $x < y$, and with an edge between x and y if $x \lessdot y$. Figure 2 illustrates four arrangements $\mathcal{A}$ in $\mathbb{R}^2$, with (the diagram of) $L(\mathcal{A})$ drawn below $\mathcal{A}$.

A *chain of length* k in a poset P is a set $x_0 < x_1 < \cdots < x_k$ of elements of P. The chain is *saturated* if $x_0 \lessdot x_1 \lessdot \cdots \lessdot x_k$. We say that P is *graded* of rank n if every maximal chain of P has length n. In this case P has a *rank function* $\mathrm{rk} : P \to \mathbb{N}$ defined by:

- $\mathrm{rk}(x) = 0$ if x is a minimal element of P.
- $\mathrm{rk}(y) = \mathrm{rk}(x) + 1$ if $x \lessdot y$ in P.

If $x < y$ in a graded poset P then we write $\mathrm{rk}(x, y) = \mathrm{rk}(y) - \mathrm{rk}(x)$, the *length* of the interval $[x, y]$. Note that we use the notation $\mathrm{rank}(\mathcal{A})$ for the rank of an arrangement $\mathcal{A}$ but rk for the rank function of a graded poset.

Proposition 1.1. *Let $\mathcal{A}$ be an arrangement in a vector space $V \cong K^n$. Then the intersection poset $L(\mathcal{A})$ is graded of rank equal to rank($\mathcal{A}$). The rank function of $L(\mathcal{A})$ is given by*

$$\mathrm{rk}(x) = \mathrm{codim}(x) = n - \dim(x),$$

where dim(x) is the dimension of x as an affine subspace of V.

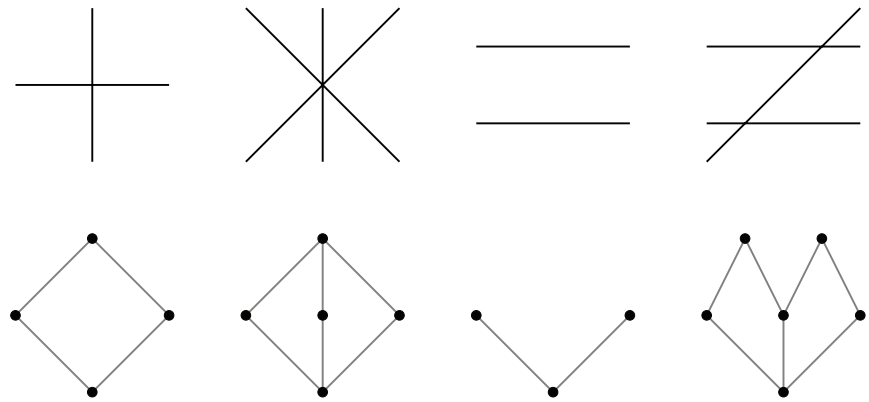


Figure 2. Examples of intersection posets

Proof. Since $L(\mathcal{A})$ has a unique minimal element $\hat{0} = V$, it suffices to show that (a) if $x \lessdot y$ in $L(\mathcal{A})$ then $\dim(x) - \dim(y) = 1$, and (b) all maximal elements of $L(\mathcal{A})$ have dimension $n - \operatorname{rank}(\mathcal{A})$. By linear algebra, if H is a hyperplane and x an affine subspace, then $H \cap x = x$ or $\dim(x) - \dim(H \cap x) = 1$, so (a) follows. Now suppose that x has the largest codimension of any element of $L(\mathcal{A})$, say $\operatorname{codim}(x) = d$. Thus x is an intersection of d linearly independent hyperplanes (i.e., their normals are linearly independent) $H_1, \ldots, H_d$ in $\mathcal{A}$. Let $y \in L(\mathcal{A})$ with $e = \operatorname{codim}(y) < d$. Thus y is an intersection of e hyperplanes, so some H_i $(1 \le i \le d)$ is linearly independent from them. Then $y \cap H_i \neq \emptyset$ and $\operatorname{codim}(y \cap H_i) > \operatorname{codim}(y)$. Hence y is not a maximal element of $L(\mathcal{A})$, proving (b). $\square$

1.3. The Characteristic Polynomial

A poset P is *locally finite* if every interval $[x, y]$ is finite. Let $\operatorname{Int}(P)$ denote the set of all closed intervals of P. For a function $f : \operatorname{Int}(P) \to \mathbb{Z}$, write $f(x, y)$ for $f([x, y])$. We now come to a fundamental invariant of locally finite posets.

Definition 1.2. Let P be a locally finite poset. Define a function $\mu = \mu_P : \operatorname{Int}(P) \to \mathbb{Z}$, called the *Möbius function* of P, by the conditions:

$$\begin{aligned} \mu(x,x) &= 1, \text{ for all } x \in P \\ \mu(x,y) &= -\sum_{x \le z < y} \mu(x,z), \text{ for all } x < y \text{ in } P. \end{aligned} \tag{2}$$

This second condition can also be written

$$\sum_{x \le z \le y} \mu(x,z) = 0, \text{ for all } x < y \text{ in } P.$$

If P has a $\hat{0}$, then we write $\mu(x) = \mu(\hat{0}, x)$. Figure 3 shows the intersection poset L of the arrangement $\mathcal{A}$ in K^3 (for any field K) defined by $Q_{\mathcal{A}}(x) = xyz(x+y)$, together with the value $\mu(x)$ for all $x \in L$.

A important application of the Möbius function is the *Möbius inversion formula*. The best way to understand this result (though it does have a simple direct proof) requires the machinery of incidence algebras. Let $\mathcal{I}(P) = \mathcal{I}(P, K)$ denote

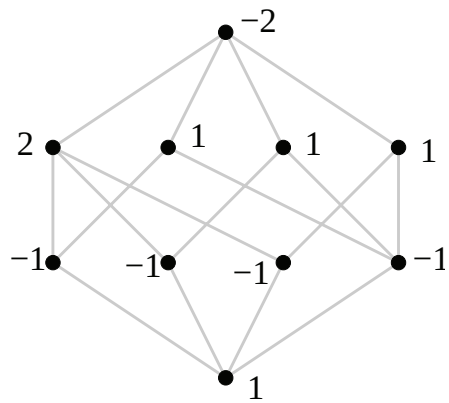


Figure 3. An intersection poset and Möbius function values

the vector space of all functions $f : \mathrm{Int}(P) \to K$. Write $f(x, y)$ for $f([x, y])$. For $f, g \in \mathcal{I}(P)$, define the product $fg \in \mathcal{I}(P)$ by

$$fg(x, y) = \sum_{x \le z \le y} f(x, z)g(z, y).$$

It is easy to see that this product makes $\mathcal{I}(P)$ an associative $\mathbb{Q}$-algebra, with multiplicative identity δ given by

$$\delta(x, y) = \begin{cases} 1, & x = y \\ 0, & x < y. \end{cases}$$

Define the *zeta function* $\zeta \in \mathcal{I}(P)$ of P by $\zeta(x, y) = 1$ for all $x \le y$ in P. Note that the Möbius function μ is an element of $\mathcal{I}(P)$. The definition of μ (Definition 1.2) is equivalent to the relation $\mu\zeta = \delta$ in $\mathcal{I}(P)$. In any finite-dimensional algebra over a field, one-sided inverses are two-sided inverses, so $\mu = \zeta^{-1}$ in $\mathcal{I}(P)$.

Theorem 1.1. *Let P be a finite poset with Möbius function μ, and let $f, g : P \to K$. Then the following two conditions are equivalent:*

$$\begin{aligned} f(x) &= \sum_{y \ge x} g(y), \text{ for all } x \in P \\ g(x) &= \sum_{y \ge x} \mu(x, y) f(y), \text{ for all } x \in P. \end{aligned}$$

Proof. The set K^P of all functions $P \to K$ forms a vector space on which $\mathcal{I}(P)$ acts (on the left) as an algebra of linear transformations by

$$(\xi f)(x) = \sum_{y \ge x} \xi(x, y) f(y),$$

where $f \in K^P$ and $\xi \in \mathcal{I}(P)$. The Möbius inversion formula is then nothing but the statement

$$\zeta f = g \Leftrightarrow f = \mu g.$$

□

We now come to the main concept of this section.

Definition 1.3. The *characteristic polynomial* $\chi_{\mathcal{A}}(t)$ of the arrangement $\mathcal{A}$ is defined by

$$\chi_{\mathcal{A}}(t) = \sum_{x \in L(\mathcal{A})} \mu(x) t^{\dim(x)}. \tag{3}$$

For instance, if $\mathcal{A}$ is the arrangement of Figure 3, then

$$\chi_{\mathcal{A}}(t) = t^3 - 4t^2 + 5t - 2 = (t-1)^2(t-2).$$

Note that we have immediately from the definition of $\chi_{\mathcal{A}}(t)$, where $\mathcal{A}$ is in K^n, that

$$\chi_{\mathcal{A}}(t) = t^n - (\#\mathcal{A})t^{n-1} + \cdots .$$

Example 1.4. Consider the coordinate hyperplane arrangement $\mathcal{A}$ with defining polynomial $Q_{\mathcal{A}}(x) = x_1 x_2 \cdots x_n$. Every subset of the hyperplanes in $\mathcal{A}$ has a different nonempty intersection, so $L(\mathcal{A})$ is isomorphic to the *boolean algebra* B_n of all subsets of $[n] = \{1, 2, \ldots, n\}$, ordered by inclusion.

Proposition 1.2. *Let $\mathcal{A}$ be given by the above example. Then $\chi_{\mathcal{A}}(t) = (t-1)^n$.*

Proof. The computation of the Möbius function of a boolean algebra is a standard result in enumerative combinatorics with many proofs. We will give here a naive proof from first principles. Let $y \in L(\mathcal{A})$, $r(y) = k$. We claim that

$$\mu(y) = (-1)^k. \tag{4}$$

The assertion is clearly true for $\mathrm{rk}(y) = 0$, when $y = \hat{0}$. Now let $y > \hat{0}$. We need to show that

$$\sum_{x \le y} (-1)^{\mathrm{rk}(x)} = 0. \tag{5}$$

The number of x such that $x \le y$ and $\mathrm{rk}(x) = i$ is $\binom{k}{i}$, so (5) is equivalent to the well-known identity $\sum_{i=0}^{k} (-1)^i \binom{k}{i} = 0$ for $k > 0$ (easily proved by substituting $q = -1$ in the binomial expansion of $(q+1)^k$). □

Exercises

We will (subjectively) indicate the difficulty level of each problem as follows:

[1] easy: most students should be able to solve it
[2] moderately difficult: many students should be able to solve it
[3] difficult: a few students should be able to solve it
[4] horrendous: no students should be able to solve it (without already knowing how)
[5] unsolved.

Further gradations are indicated by + and −. Thus a [3–] problem is about the most difficult that makes a reasonable homework exercise, and a [5–] problem is an unsolved problem that has received little attention and may not be too difficult.

NOTE. Unless explicitly stated otherwise, all graphs, posets, lattices, etc., are assumed to be *finite*.

(1) [2] Show that every region R of an arrangement $\mathcal{A}$ in $\mathbb{R}^n$ is an open convex set. Deduce that R is homeomorphic to the interior of an n-dimensional ball.

(2) [1+] Let $\mathcal{A}$ be an arrangement and $\mathrm{ess}(\mathcal{A})$ its essentialization. Show that
$$t^{\dim(\mathrm{ess}(\mathcal{A}))}\chi_{\mathcal{A}}(t) = t^{\dim(\mathcal{A})}\chi_{\mathrm{ess}(\mathcal{A})}(t).$$

(3) [2+] Let $\mathcal{A}$ be the arrangement in $\mathbb{R}^n$ with equations
$$x_1 = x_2,\ x_2 = x_3,\ \ldots,\ x_{n-1} = x_n,\ x_n = x_1.$$
Compute the characteristic polynomial $\chi_{\mathcal{A}}(t)$, and compute the number $r(\mathcal{A})$ of regions of $\mathcal{A}$.

(4) [2+] Let $\mathcal{A}$ be an arrangement in $\mathbb{R}^n$ with m hyperplanes. Find the maximum possible number $f(n,m)$ of regions of $\mathcal{A}$.

(5) [2] Let $\mathcal{A}$ be an arrangement in the n-dimensional vector space V whose normals span a subspace W, and let $\mathcal{B}$ be another arrangement in V whose normals span a subspace Y. Suppose that $W \cap Y = \{0\}$. Show that
$$\chi_{\mathcal{A}\cup\mathcal{B}}(t) = t^{-n}\chi_{\mathcal{A}}(t)\chi_{\mathcal{B}}(t).$$

(6) [2] Let $\mathcal{A}$ be an arrangment in a vector space V. Suppose that $\chi_{\mathcal{A}}(t)$ is divisible by t^k but not t^{k+1}. Show that $\mathrm{rank}(\mathcal{A}) = n - k$.

(7) Let $\mathcal{A}$ be an essential arrangement in $\mathbb{R}^n$. Let Γ be the union of the bounded faces of $\mathcal{A}$.
   (a) [3] Show that Γ is contractible.
   (b) [2] Show that Γ need not be homeomorphic to a closed ball.
   (c) [2+] Show that Γ need not be starshaped. (A subset S of $\mathbb{R}^n$ is *starshaped* if there is a point $x \in S$ such that for all $y \in S$, the line segment from x to y lies in S.)
   (d) [3] Show that Γ is pure, i.e., all maximal faces of Γ have the same dimension. (This was an open problem solved by Xun Dong at the PCMI Summer Session in Geometric Combinatorics, July 11–31, 2004.)
   (e) [5] Suppose that $\mathcal{A}$ is in general position. Is Γ homeomorphic to an n-dimensional closed ball?

LECTURE 2
Properties of the Intersection Poset and Graphical Arrangements

2.1. Properties of the Intersection Poset

Let $\mathcal{A}$ be an arrangement in the vector space V. A *subarrangement* of $\mathcal{A}$ is a subset $\mathcal{B} \subseteq \mathcal{A}$. Thus $\mathcal{B}$ is also an arrangement in V. If $x \in L(\mathcal{A})$, define the subarrangement $\mathcal{A}_x \subseteq \mathcal{A}$ by

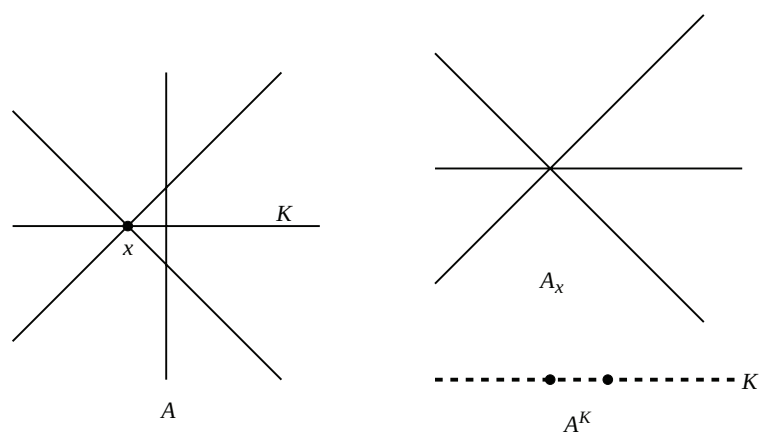
$$\mathcal{A}_x = \{H \in \mathcal{A} \,:\, x \subseteq H\}. \tag{6}$$

Also define an arrangement $\mathcal{A}^x$ in the affine subspace $x \in L(\mathcal{A})$ by

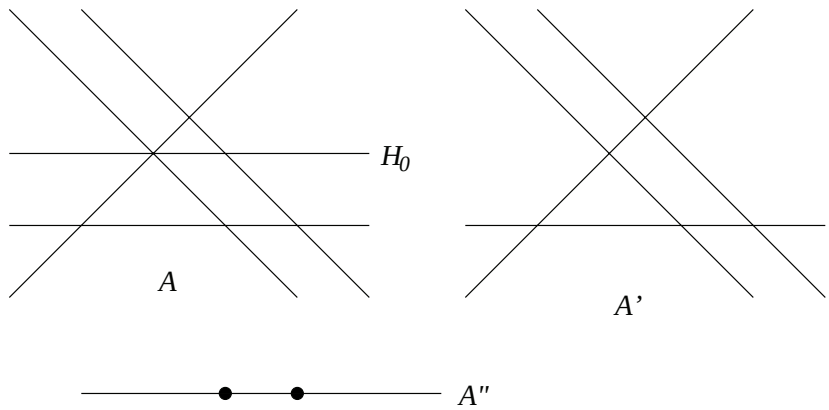
$$\mathcal{A}^x = \{x \cap H \neq \emptyset \,:\, H \in \mathcal{A} - \mathcal{A}_x\}.$$

Note that if $x \in L(\mathcal{A})$, then

$$\begin{aligned} L(\mathcal{A}_x) \cong \Lambda_x &:= \{y \in L(\mathcal{A}) \,:\, y \leq x\} \\ L(\mathcal{A}^x) \cong V_x &:= \{y \in L(\mathcal{A}) \,:\, y \geq x\} \end{aligned} \tag{7}$$



Choose $H_0 \in \mathcal{A}$. Let $\mathcal{A}' = \mathcal{A} - \{H_0\}$ and $\mathcal{A}'' = \mathcal{A}^{H_0}$. We call $(\mathcal{A}, \mathcal{A}', \mathcal{A}'')$ a *triple* of arrangements with *distinguished hyperplane* H_0.

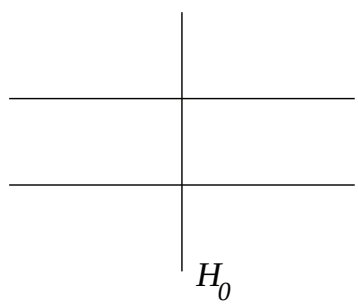


The main goal of this section is to give a formula in terms of $\chi_{\mathcal{A}}(t)$ for $r(\mathcal{A})$ and $b(\mathcal{A})$ when $K = \mathbb{R}$ (Theorem 2.5). We first establish recurrences for these two quantities.

Lemma 2.1. *Let $(\mathcal{A}, \mathcal{A}', \mathcal{A}'')$ be a triple of real arrangements with distinguished hyperplane H_0. Then*

$$\begin{aligned} r(\mathcal{A}) &= r(\mathcal{A}') + r(\mathcal{A}'') \\ b(\mathcal{A}) &= \begin{cases} b(\mathcal{A}') + b(\mathcal{A}''), & \text{if } \operatorname{rank}(\mathcal{A}) = \operatorname{rank}(\mathcal{A}') \\ 0, & \text{if } \operatorname{rank}(\mathcal{A}) = \operatorname{rank}(\mathcal{A}') + 1. \end{cases} \end{aligned}$$

NOTE. If $\operatorname{rank}(\mathcal{A}) = \operatorname{rank}(\mathcal{A}')$, then also $\operatorname{rank}(\mathcal{A}) = 1 + \operatorname{rank}(\mathcal{A}'')$. The figure below illustrates the situation when $\operatorname{rank}(\mathcal{A}) = \operatorname{rank}(\mathcal{A}') + 1$.



Proof. Note that $r(\mathcal{A})$ equals $r(\mathcal{A}')$ plus the number of regions of $\mathcal{A}'$ cut into two regions by H_0. Let R' be such a region of $\mathcal{A}'$. Then $R' \cap H_0 \in \mathcal{R}(\mathcal{A}'')$. Conversely, if $R'' \in \mathcal{R}(\mathcal{A}'')$ then points near R'' on either side of H_0 belong to the same region $R' \in \mathcal{R}(\mathcal{A}')$, since any $H \in \mathcal{R}(\mathcal{A}')$ separating them would intersect R''. Thus R' is cut in two by H_0. We have established a bijection between regions of $\mathcal{A}'$ cut into two by H_0 and regions of $\mathcal{A}''$, establishing the first recurrence.

The second recurrence is proved analogously; the details are omitted. □

We now come to the fundamental recursive property of the characteristic polynomial.

Lemma 2.2. *(Deletion-Restriction) Let $(\mathcal{A}, \mathcal{A}', \mathcal{A}'')$ be a triple of real arrangements. Then*

$$\chi_{\mathcal{A}}(t) = \chi_{\mathcal{A}'}(t) - \chi_{\mathcal{A}''}(t).$$

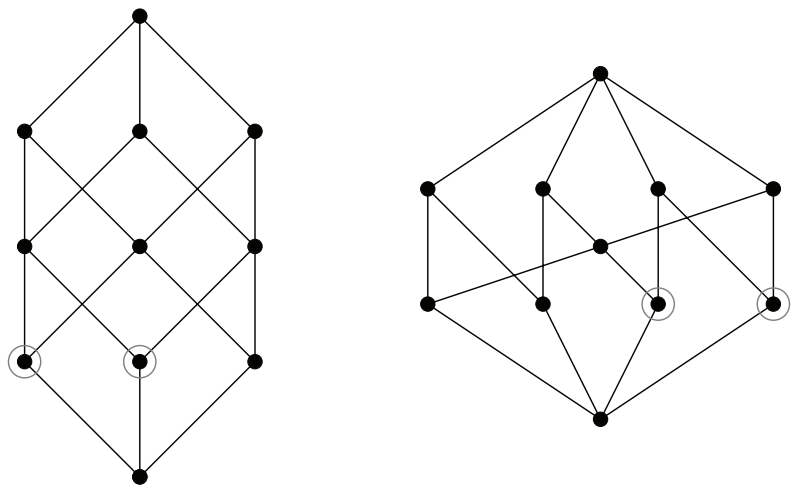


Figure 1. Two non-lattices

For the proof of this lemma, we will need some tools. (A more elementary proof could be given, but the tools will be useful later.)

Let P be a poset. An *upper bound* of $x, y \in P$ is an element $z \in P$ satisfying $z \geq x$ and $z \geq y$. A *least upper bound* or *join* of x and y, denoted $x \vee y$, is an upper bound z such that $z \leq z'$ for all upper bounds z'. Clearly if $x \vee y$ exists, then it is unique. Similarly define a *lower bound* of x and y, and a *greatest lower bound* or *meet*, denoted $x \wedge y$. A *lattice* is a poset L for which any two elements have a meet and join. A *meet-semilattice* is a poset P for which any two elements have a meet. Dually, a *join-semilattice* is a poset P for which any two elements have a join. Figure 1 shows two non-lattices, with a pair of elements circled which don't have a join.

Lemma 2.3. *A finite meet-semilattice L with a unique maximal element $\hat{1}$ is a lattice. Dually, a finite join-semilattice L with a unique minimal element $\hat{0}$ is a lattice.*

Proof. Let L be a finite meet-semilattice. If $x, y \in L$ then the set of upper bounds of x, y is nonempty since $\hat{1}$ is an upper bound. Hence

$$x \vee y = \bigwedge_{\substack{z \geq x \\ z \geq y}} z.$$

The statement for join-semilattices is by "duality," i.e., interchanging $\leq$ with $\geq$, and $\wedge$ with $\vee$. $\square$

The reader should check that Lemma 2.3 need not hold for *infinite* semilattices.

Proposition 2.3. *Let $\mathcal{A}$ be an arrangement. Then $L(\mathcal{A})$ is a meet-semilattice. In particular, every interval $[x, y]$ of $L(\mathcal{A})$ is a lattice. Moreover, $L(\mathcal{A})$ is a lattice if and only if $\mathcal{A}$ is central.*

Proof. If $\bigcap_{H \in \mathcal{A}} H = \emptyset$, then adjoin $\emptyset$ to $L(\mathcal{A})$ as the unique maximal element, obtaining the augmented intersection poset $L'(\mathcal{A})$. In $L'(\mathcal{A})$ it is clear that $x \vee y = x \cap y$. Hence $L'(\mathcal{A})$ is a join-semilattice. Since it has a $\hat{0}$, it is a lattice by Lemma 2.3.

Since $L(\mathcal{A}) = L'(\mathcal{A})$ or $L(\mathcal{A}) = L'(\mathcal{A}) - \{\hat{1}\}$, it follows that $L(\mathcal{A})$ is always a meet-semilattice, and is a lattice if $\mathcal{A}$ is central. If $\mathcal{A}$ isn't central, then $\bigvee_{x \in L(\mathcal{A})} x$ does not exist, so $L(\mathcal{A})$ is not a lattice. $\square$

We now come to a basic formula for the Möbius function of a lattice.

Theorem 2.2. *(the Cross-Cut Theorem) Let L be a finite lattice. Let X be a subset of L such that $\hat{0} \notin X$, and such that if $y \in L$, $y \neq \hat{0}$, then some $x \in X$ satisfies $x \leq y$. Let N_k be the number of k-element subsets of X with join $\hat{1}$. Then*

$$\mu_L(\hat{0}, \hat{1}) = N_0 - N_1 + N_2 - \cdots .$$

We will prove Theorem 2.2 by an algebraic method. Such a sophisticated proof is unnecessary, but the machinery we develop will be used later (Theorem 4.13). Let L be a finite lattice and K a field. The *Möbius algebra* of L, denoted $A(L)$, is the semigroup algebra of L over K with respect to the operation $\vee$. (Sometimes the operation is taken to be $\wedge$ instead of $\vee$, but for our purposes, $\vee$ is more convenient.) In other words, $A(L) = KL$ (the vector space with basis L) as a vector space. If $x, y \in L$ then we define $xy = x \vee y$. Multiplication is extended to all of $A(L)$ by bilinearity (or distributivity). Algebraists will recognize that $A(L)$ is a finite-dimensional commutative algebra with a basis of idempotents, and hence is isomorphic to $K^{\#L}$ (as an algebra). We will show this by exhibiting an explicit isomorphism $A(L) \xrightarrow{\cong} K^{\#L}$. For $x \in L$, define

$$\sigma_x = \sum_{y \geq x} \mu(x, y) y \in A(L), \tag{8}$$

where μ denotes the Möbius function of L. Thus by the Möbius inversion formula,

$$x = \sum_{y \geq x} \sigma_y, \text{ for all } x \in L. \tag{9}$$

Equation (9) shows that the σ_x's span $A(L)$. Since $\#\{\sigma_x : x \in L\} = \#L = \dim A(L)$, it follows that the σ_x's form a basis for $A(L)$.

Theorem 2.3. *Let $x, y \in L$. Then $\sigma_x \sigma_y = \delta_{xy} \sigma_x$, where δ_{xy} is the Kronecker delta. In other words, the σ_x's are orthogonal idempotents. Hence*

$$A(L) = \bigoplus_{x \in L} K \cdot \sigma_x \quad \textit{(algebra direct sum).}$$

Proof. Define a K-algebra $A'(L)$ with basis $\{\sigma'_x : x \in L\}$ and multiplication $\sigma'_x \sigma'_y = \delta_{xy} \sigma'_x$. For $x \in L$ set $x' = \sum_{s \geq x} \sigma'_s$. Then

$$\begin{aligned} x'y' &= \left(\sum_{s \geq x} \sigma'_s \right) \left(\sum_{t \geq y} \sigma'_t \right) \\ &= \sum_{\substack{s \geq x \\ s \geq y}} \sigma'_s \\ &= \sum_{s \geq x \vee y} \sigma'_s \\ &= (x \vee y)'. \end{aligned}$$

Hence the linear transformation $\varphi : A(L) \to A'(L)$ defined by $\varphi(x) = x'$ is an algebra isomorphism. Since $\varphi(\sigma_x) = \sigma'_x$, it follows that $\sigma_x \sigma_y = \delta_{xy} \sigma_x$. $\square$

NOTE. The algebra $A(L)$ has a multiplicative identity, viz., $1 = \hat{0} = \sum_{x \in L} \sigma_x$.

Proof of Theorem 2.2. Let $\mathrm{char}(K) = 0$, e.g., $K = \mathbb{Q}$. For any $x \in L$, we have in $A(L)$ that

$$\hat{0} - x = \sum_{y \geq \hat{0}} \sigma_y - \sum_{y \geq x} \sigma_y = \sum_{y \not\geq x} \sigma_y.$$

Hence by the orthogonality of the σ_y's we have

$$\prod_{x \in X} (\hat{0} - x) = \sum_y \sigma_y,$$

where y ranges over all elements of L satisfying $y \not\geq x$ for all $x \in X$. By hypothesis, the only such element is $\hat{0}$. Hence

$$\prod_{x \in X} (\hat{0} - x) = \sigma_{\hat{0}}.$$

If we now expand both sides as linear combinations of elements of L and equate coefficients of $\hat{1}$, the result follows. □

NOTE. In a finite lattice L, an *atom* is an element covering $\hat{0}$. Let T be the set of atoms of L. Then a set $X \subseteq L - \{\hat{0}\}$ satisfies the hypotheses of Theorem 2.2 if and only if $T \subseteq X$. Thus the simplest choice of X is just $X = T$.

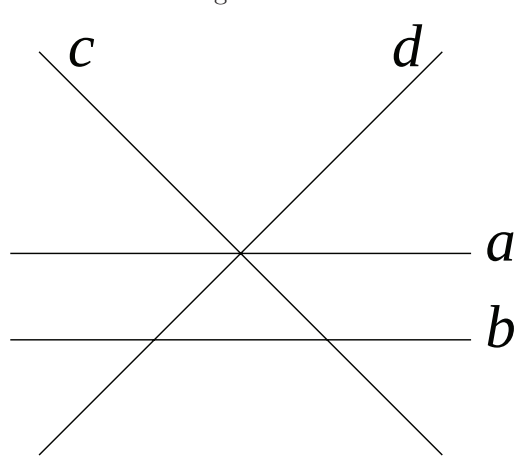
Example 2.5. Let $L = B_n$, the boolean algebra of all subsets of $[n]$. Let $X = T = \{\{i\} : i \in [n]\}$. Then $N_0 = N_1 = \cdots = N_{n-1} = 0$, $N_n = 1$. Hence $\mu(\hat{0}, \hat{1}) = (-1)^n$, agreeing with Proposition 1.2.

We will use the Crosscut Theorem to obtain a formula for the characteristic polynomial of an arrangement $\mathcal{A}$. Extending slightly the definition of a central arrangement, call any subset $\mathcal{B}$ of $\mathcal{A}$ *central* if $\bigcap_{H \in \mathcal{B}} H \neq \emptyset$. The following result is due to Hassler Whitney for linear arrangements. Its easy extension to arbitrary arrangements appears in [**24**, Lemma 2.3.8].

Theorem 2.4. *(Whitney's theorem) Let $\mathcal{A}$ be an arrangement in an n-dimensional vector space. Then*

$$\chi_{\mathcal{A}}(t) = \sum_{\substack{\mathcal{B} \subseteq \mathcal{A} \\ \mathcal{B} \text{ central}}} (-1)^{\#\mathcal{B}} t^{n - \mathrm{rank}(\mathcal{B})}. \tag{10}$$

Example 2.6. Let $\mathcal{A}$ be the arrangement in $\mathbb{R}^2$ shown below.



The following table shows all central subsets $\mathcal{B}$ of $\mathcal{A}$ and the values of $\#\mathcal{B}$ and $\mathrm{rank}(\mathcal{B})$.

$\mathcal{B}$	$\#\mathcal{B}$	$\mathrm{rank}(\mathcal{B})$
$\emptyset$	0	0
a	1	1
b	1	1
c	1	1
d	1	1
ac	2	2
ad	2	2
bc	2	2
bd	2	2
cd	2	2
acd	3	2

It follows that $\chi_{\mathcal{A}}(t) = t^2 - 4t + (5-1) = t^2 - 4t + 4$.

Proof of Theorem 2.4. Let $z \in L(\mathcal{A})$. Let

$$\Lambda_z = \{x \in L(\mathcal{A}) \,:\, x \leq z\},$$

the *principal order ideal* generated by z. Recall the definition

$$\mathcal{A}_z = \{H \in \mathcal{A} \,:\, H \leq z \text{ (i.e., } z \subseteq H)\}.$$

By the Crosscut Theorem (Theorem 2.2), we have

$$\mu(z) = \sum_k (-1)^k N_k(z),$$

where $N_k(z)$ is the number of k-subsets of $\mathcal{A}_z$ with join z. In other words,

$$\mu(z) = \sum_{\substack{\mathcal{B} \subseteq \mathcal{A}_z \\ z = \bigcap_{H \in \mathcal{B}} H}} (-1)^{\#\mathcal{B}}.$$

Note that $z = \bigcap_{H \in \mathcal{B}} H$ implies that $\mathrm{rank}(\mathcal{B}) = n - \dim z$. Now multiply both sides by $t^{\dim(z)}$ and sum over z to obtain equation (10). $\square$

We have now assembled all the machinery necessary to prove the Deletion-Restriction Lemma (Lemma 2.2) for $\chi_{\mathcal{A}}(t)$.

Proof of Lemma 2.2. Let $H_0 \in \mathcal{A}$ be the hyperplane defining the triple $(\mathcal{A}, \mathcal{A}', \mathcal{A}'')$. Split the sum on the right-hand side of (10) into two sums, depending on whether $H_0 \notin \mathcal{B}$ or $H_0 \in \mathcal{B}$. In the former case we get

$$\sum_{\substack{H_0 \notin \mathcal{B} \subseteq \mathcal{A} \\ \mathcal{B} \text{ central}}} (-1)^{\#\mathcal{B}} t^{n - \mathrm{rank}(\mathcal{B})} = \chi_{\mathcal{A}'}(t).$$

In the latter case, set $\mathcal{B}_1 = (\mathcal{B} - \{H_0\})^{H_0}$, a central arrangement in $H_0 \cong K^{n-1}$ and a subarrangement of $\mathcal{A}^{H_0} = \mathcal{A}''$. Since $\#\mathcal{B}_1 = \#\mathcal{B} - 1$ and $\mathrm{rank}(\mathcal{B}_1) = \mathrm{rank}(\mathcal{B}) - 1$, we get

$$\begin{aligned} \sum_{\substack{H_0 \in \mathcal{B} \subseteq \mathcal{A} \\ \mathcal{B} \text{ central}}} (-1)^{\#\mathcal{B}} t^{n - \mathrm{rank}(\mathcal{B})} &= \sum_{\mathcal{B}_1 \in \mathcal{A}''} (-1)^{\#\mathcal{B}_1 + 1} t^{(n-1) - \mathrm{rank}(\mathcal{B}_1)} \\ &= -\chi_{\mathcal{A}''}(t), \end{aligned}$$

and the proof follows. $\square$

2.2. The Number of Regions

The next result is perhaps the first major theorem in the subject of hyperplane arrangements, due to Thomas Zaslavsky in 1975.

Theorem 2.5. *Let $\mathcal{A}$ be an arrangement in an n-dimensional real vector space. Then*

$$r(\mathcal{A}) = (-1)^n \chi_{\mathcal{A}}(-1) \tag{11}$$

$$b(\mathcal{A}) = (-1)^{\mathrm{rank}(\mathcal{A})} \chi_{\mathcal{A}}(1). \tag{12}$$

First proof. Equation (11) holds for $\mathcal{A} = \emptyset$, since $r(\emptyset) = 1$ and $\chi_\emptyset(t) = t^n$. By Lemmas 2.1 and 2.2, both $r(\mathcal{A})$ and $(-1)^n \chi_{\mathcal{A}}(-1)$ satisfy the same recurrence, so the proof follows.

Now consider equation (12). Again it holds for $\mathcal{A} = \emptyset$ since $b(\emptyset) = 1$. (Recall that $b(\mathcal{A})$ is the number of *relatively* bounded regions. When $\mathcal{A} = \emptyset$, the entire ambient space $\mathbb{R}^n$ is relatively bounded.) Now

$$\chi_{\mathcal{A}}(1) = \chi_{\mathcal{A}'}(1) - \chi_{\mathcal{A}''}(1).$$

Let $d(\mathcal{A}) = (-1)^{\mathrm{rank}(\mathcal{A})} \chi_{\mathcal{A}}(1)$. If $\mathrm{rank}(\mathcal{A}) = \mathrm{rank}(\mathcal{A}') = \mathrm{rank}(\mathcal{A}'') + 1$, then $d(\mathcal{A}) = d(\mathcal{A}') + d(\mathcal{A}'')$. If $\mathrm{rank}(\mathcal{A}) = \mathrm{rank}(\mathcal{A}') + 1$ then $b(\mathcal{A}) = 0$ [why?] and $L(\mathcal{A}') \cong L(\mathcal{A}'')$ [why?]. Thus from Lemma 2.2 we have $d(\mathcal{A}) = 0$. Hence in all cases $b(\mathcal{A})$ and $d(\mathcal{A})$ satisfy the same recurrence, so $b(\mathcal{A}) = d(\mathcal{A})$. □

Second proof. Our second proof of Theorem 2.5 is based on Möbius inversion and some instructive topological considerations. For this proof we assume basic knowledge of the Euler characteristic $\psi(\Delta)$ of a topological space Δ. (Standard notation is $\chi(\Delta)$, but this would cause too much confusion with the characteristic polynomial.) In particular, if Δ is suitably decomposed into cells with f_i i-dimensional cells, then

$$\psi(\Delta) = f_0 - f_1 + f_2 - \cdots. \tag{13}$$

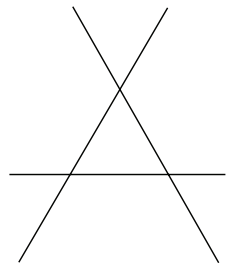
We take (13) as the definition of $\psi(\Delta)$. For "nice" spaces and decompositions, it is independent of the decomposition. In particular, $\psi(\mathbb{R}^n) = (-1)^n$. Write $\bar{R}$ for the closure of a region $R \in \mathcal{R}(\mathcal{A})$.

Definition 2.4. A (closed) *face* of a real arrangement $\mathcal{A}$ is a set $\emptyset \neq F = \bar{R} \cap x$, where $R \in \mathcal{R}(\mathcal{A})$ and $x \in L(\mathcal{A})$.

If we regard $\bar{R}$ as a convex polyhedron (possibly unbounded), then a face of $\mathcal{A}$ is just a face of some $\bar{R}$ in the usual sense of the face of a polyhedron, i.e., the intersection of $\bar{R}$ with a supporting hyperplane. In particular, each $\bar{R}$ is a face of $\mathcal{A}$. The *dimension* of a face F is defined by

$$\dim(F) = \dim(\mathrm{aff}(F)),$$

where $\mathrm{aff}(F)$ denotes the affine span of F. A *k-face* is a k-dimensional face of $\mathcal{A}$. For instance, the arrangement below has three 0-faces (vertices), nine 1-faces, and seven 2-faces (equivalently, seven regions). Hence $\psi(\mathbb{R}^2) = 3 - 9 + 7 = 1$.



Write $\mathcal{F}(\mathcal{A})$ for the set of faces of $\mathcal{A}$, and let relint denote relative interior. Then

$$\mathbb{R}^n = \bigsqcup_{F \in \mathcal{F}(\mathcal{A})} \operatorname{relint}(F),$$

where $\bigsqcup$ denotes disjoint union. If $f_k(\mathcal{A})$ denotes the number of k-faces of $\mathcal{A}$, it follows that

$$(-1)^n = \psi(\mathbb{R}^n) = f_0(\mathcal{A}) - f_1(\mathcal{A}) + f_2(\mathcal{A}) - \cdots .$$

Every k-face is a region of exactly one $\mathcal{A}^y$ for $y \in L(\mathcal{A})$. Hence

$$f_k(\mathcal{A}) = \sum_{\substack{y \in L(\mathcal{A}) \\ \dim(y)=k}} r(\mathcal{A}^y).$$

Multiply by $(-1)^k$ and sum over k to get

$$(-1)^n = \psi(\mathbb{R}^n) = \sum_{y \in L(\mathcal{A})} (-1)^{\dim(y)} r(\mathcal{A}^y).$$

Replacing $\mathbb{R}^n$ by $x \in L(\mathcal{A})$ gives

$$(-1)^{\dim(x)} = \psi(x) = \sum_{\substack{y \in L(\mathcal{A}) \\ y \geq x}} (-1)^{\dim(y)} r(\mathcal{A}^y).$$

Möbius inversion yields

$$(-1)^{\dim(x)} r(\mathcal{A}^x) = \sum_{\substack{y \in L(\mathcal{A}) \\ y \geq x}} (-1)^{\dim(y)} \mu(x,y).$$

Putting $x = \mathbb{R}^n$ gives

$$(-1)^n r(\mathcal{A}) = \sum_{y \in L(\mathcal{A})} (-1)^{\dim(y)} \mu(y) = \chi_{\mathcal{A}}(-1),$$

thereby proving (11).

The relatively bounded case (equation (12)) is similar, but with one technical complication. We may assume that $\mathcal{A}$ is essential, since $b(\mathcal{A}) = b(\operatorname{ess}(\mathcal{A}))$ and

$$\chi_{\mathcal{A}}(t) = t^{\dim(\mathcal{A}) - \dim(\operatorname{ess}(\mathcal{A}))} \chi_{\operatorname{ess}(\mathcal{A})}(t).$$

In this case, the relatively bounded regions are actually bounded. Let

$$\begin{aligned} \mathcal{F}_b(\mathcal{A}) &= \{F \in \mathcal{F}(\mathcal{A}) : F \text{ is relatively bounded}\} \\ \Gamma &= \bigcup_{F \in \mathcal{F}_b(\mathcal{A})} F. \end{aligned}$$

The difficulty lies in computing $\psi(\Gamma)$. Zaslavsky conjectured in 1975 that Γ is star-shaped, i.e., there exists $x \in \Gamma$ such that for every $y \in \Gamma$, the line segment

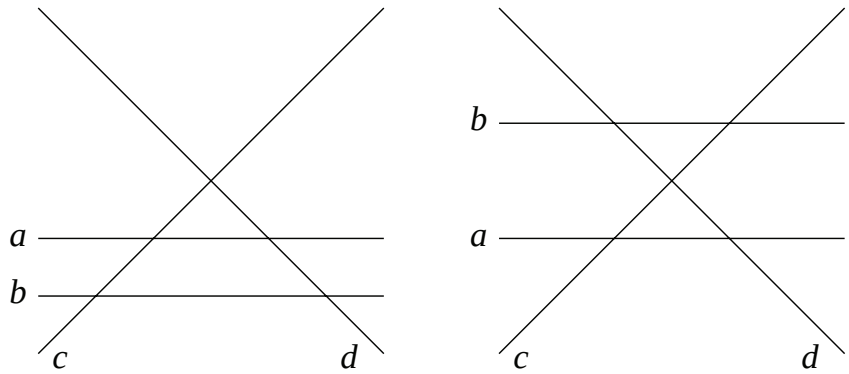


Figure 2. Two arrangements with the same intersection poset

joining x and y lies in Γ. This would imply that Γ is contractible, and hence (since Γ is compact when $\mathcal{A}$ is essential) $\psi(\Gamma) = 1$. A counterexample to Zaslavsky's conjecture appears as an exercise in [**7**, Exer. 4.29], but nevertheless Björner and Ziegler showed that Γ is indeed contractible. (See [**7**, Thm. 4.5.7(b)] and Lecture 1, Exercise 7.) The argument just given for $r(\mathcal{A})$ now carries over *mutatis mutandis* to $b(\mathcal{A})$. There is also a direct argument that $\psi(\Gamma) = 1$, circumventing the need to show that Γ is contractible. We will omit proving here that $\psi(\Gamma) = 1$. $\square$

Corollary 2.1. *Let $\mathcal{A}$ be a real arrangement. Then $r(\mathcal{A})$ and $b(\mathcal{A})$ depend only on $L(\mathcal{A})$.*

Figure 2 shows two arrangements in $\mathbb{R}^2$ with different "face structure" but the same $L(\mathcal{A})$. The first arrangement has for instance one triangular and one quadrilateral face, while the second has two triangular faces. Both arrangements, however, have ten regions and two bounded regions.

We now give two basic examples of arrangements and the computation of their characteristic polynomials.

Proposition 2.4. *(general position) Let $\mathcal{A}$ be an n-dimensional arrangement of m hyperplanes in general position. Then*

$$\chi_{\mathcal{A}}(t) = t^n - mt^{n-1} + \binom{m}{2}t^{n-2} - \cdots + (-1)^n\binom{m}{n}.$$

In particular, if $\mathcal{A}$ is a real arrangement, then

$$\begin{aligned} r(\mathcal{A}) &= 1 + m + \binom{m}{2} + \cdots + \binom{m}{n} \\ b(\mathcal{A}) &= (-1)^n\left(1 - m + \binom{m}{2} - \cdots + (-1)^n\binom{m}{n}\right) \\ &= \binom{m-1}{n}. \end{aligned}$$

Proof. Every $\mathcal{B} \subseteq \mathcal{A}$ with $\#\mathcal{B} \leq n$ defines an element $x_{\mathcal{B}} = \bigcap_{H\in\mathcal{B}} H$ of $L(\mathcal{A})$. Hence $L(\mathcal{A})$ is a *truncated boolean algebra*:

$$L(\mathcal{A}) \cong \{S \subseteq [m] \,:\, \#S \leq n\},$$

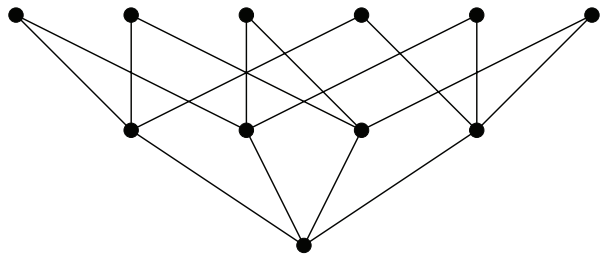


Figure 3. The truncated boolean algebra of rank 2 with four atoms

ordered by inclusion. Figure 3 shows the case $n = 2$ and $m = 4$, i.e., four lines in general position in $\mathbb{R}^2$. If $x \in L(\mathcal{A})$ and $\mathrm{rk}(x) = k$, then $[\hat{0}, x] \cong B_k$, a boolean algebra of rank k. By equation (4) there follows $\mu(x) = (1)^k$. Hence

$$\begin{aligned} \chi_{\mathcal{A}}(t) &= \sum_{\substack{S \subseteq [m] \\ \#S \leq n}} (-1)^{\#S} t^{n-\#S} \\ &= t^n - mt^{n-1} + \cdots + (-1)^n \binom{m}{n}. \quad \square \end{aligned}$$

Note. Arrangements whose hyperplanes are in general position were formerly called *free arrangements*. Now, however, free arrangements have another meaning discussed in the note following Example 4.11.

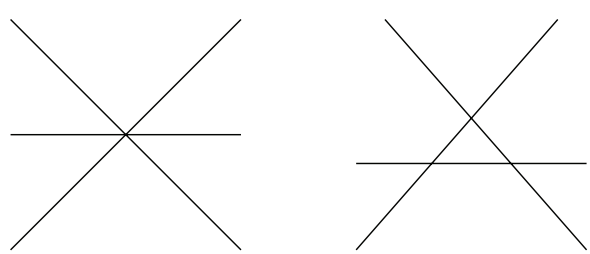
Our second example concerns generic translations of the hyperplanes of a linear arrangement. Let $L_1, \ldots, L_m$ be linear forms, not necessarily distinct, in the variables $v = (v_1, \ldots, v_n)$ over the field K. Let $\mathcal{A}$ be defined by

$$L_1(v) = a_1, \ldots, L_m(v) = a_m,$$

where $a_1, \ldots, a_m$ are *generic* elements of K. This means if $H_i = \ker(L_i(v) - a_i)$, then

$$H_{i_1} \cap \cdots \cap H_{i_k} \neq \emptyset \;\Leftrightarrow\; L_{i_1}, \ldots, L_{i_k} \text{ are linearly independent.}$$

For instance, if $K = \mathbb{R}$ and $L_1, \ldots, L_m$ are defined over $\mathbb{Q}$, then $a_1, \ldots, a_m$ are generic whenever they are linearly independent over $\mathbb{Q}$.



nongeneric generic

It follows that if $x = H_{i_1} \cap \cdots \cap H_{i_k} \in L(\mathcal{A})$, then $[\hat{0}, x] \cong B_k$. Hence

$$\chi_{\mathcal{A}}(t) = \sum_{\mathcal{B}} (-1)^{\#\mathcal{B}} t^{n-\#\mathcal{B}},$$

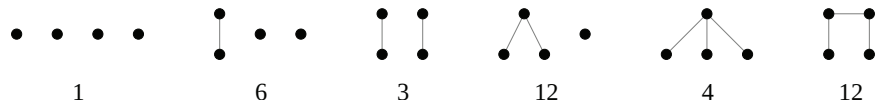


Figure 4. The forests on four vertices

where $\mathcal{B}$ ranges over all linearly independent subsets of $\mathcal{A}$. (We say that a set of hyperplanes are *linearly independent* if their normals are linearly independent.) Thus $\chi_{\mathcal{A}}(t)$, or more precisely $(-t)^n\chi_{\mathcal{A}}(-1/t)$, is the generating function for linearly independent subsets of $L_1, \ldots, L_m$ according to their number of elements. For instance, if $\mathcal{A}$ is given by Figure 2 (either arrangement) then the linearly independent subsets of hyperplanes are $\emptyset, a, b, c, d, ac, ad, bc, bd, cd$, so $\chi_{\mathcal{A}}(t) = t^2 - 4t + 5$.

Consider the more interesting example $x_i - x_j = a_{ij}$, $1 \le i < j \le n$, where the a_{ij} are generic. We could call this arrangement the *generic braid arrangement* $\mathcal{G}_n$. Identify the hyperplane $x_i - x_j = a_{ij}$ with the edge ij on the vertex set $[n]$. Thus a subset $\mathcal{B} \subseteq \mathcal{G}_n$ corresponds to a simple graph $G_{\mathcal{B}}$ on $[n]$. ("Simple" means that there is at most one edge between any two vertices, and no edge from a vertex to itself.) It is easy to see that $\mathcal{B}$ is linearly independent if and only if the graph $G_{\mathcal{B}}$ has no cycles, i.e., is a *forest*. Hence we obtain the interesting formula

$$\chi_{\mathcal{G}_n}(t) = \sum_F (-1)^{e(F)} t^{n-e(F)}, \tag{14}$$

where F ranges over all forests on $[n]$ and $e(F)$ denotes the number of edges of F. For instance, the isomorphism types of forests (with the number of distinct labelings written below the forest) on four vertices are given by Figure 4. Hence

$$\chi_{\mathcal{G}_4}(t) = t^4 - 6t^3 + 15t^2 - 16t.$$

Equation (11) can be rewritten as

$$r(\mathcal{A}) = \sum_{x \in L(\mathcal{A})} (-1)^{\mathrm{rk}(x)} \mu(x).$$

(Theorem 3.10 will show that $(-1)^{\mathrm{rk}(x)}\mu(x) > 0$, so we could also write $|\mu(x)|$ for this quantity.) It is easy to extend this result to count faces of $\mathcal{A}$ of all dimensions, not just the top dimension n. Let $f_k(\mathcal{A})$ denote the number of k-faces of the real arrangement $\mathcal{A}$.

Theorem 2.6. *We have*

$$f_k(\mathcal{A}) = \sum_{\substack{x \le y \text{ in } L(\mathcal{A}) \\ \dim(x)=k}} (-1)^{\dim(x)-\dim(y)} \mu(x,y) \tag{15}$$

$$= \sum_{\substack{x \le y \text{ in } L(\mathcal{A}) \\ \dim(x)=k}} |\mu(x,y)|. \tag{16}$$

Proof. As mentioned above, every face F is a region of a unique $\mathcal{A}^x$ for $x \in L(\mathcal{A})$, viz., $x = \mathrm{aff}(F)$. In particular, $\dim(F) = \dim(x)$. Hence if $\dim(F) = k$, then $r(\mathcal{A}^x)$ is the number of k-faces of $\mathcal{A}$ contained in x. By Theorem 2.5 and equation (7) we get

$$r(\mathcal{A}^x) = \sum_{y \ge x} (-1)^{\dim(y)-\dim(x)} \mu(x,y),$$

where we are dealing with the poset $L(\mathcal{A})$. Summing over all $x \in L(\mathcal{A})$ of dimension k yields (15), and (16) then follows from Theorem (3.10) below. $\square$

2.3. Graphical Arrangements

There are close connections between certain invariants of a graph G and an associated arrangement $\mathcal{A}_G$. Let G be a simple graph on the vertex set $[n]$. Let $E(G)$ denote the set of edges of G, regarded as two-element subsets of $[n]$. Write ij for the edge $\{i, j\}$.

Definition 2.5. The *graphical arrangement* $\mathcal{A}_G$ in K^n is the arrangement

$$x_i - x_j = 0, \ ij \in E(G).$$

Thus a graphical arrangement is simply a subarrangement of the braid arrangement $\mathcal{B}_n$. If $G = K_n$, the *complete graph* on $[n]$ (with all possible edges ij), then $\mathcal{A}_{K_n} = \mathcal{B}_n$.

Definition 2.6. A *coloring* of a graph G on $[n]$ is a map $\kappa : [n] \to \mathbb{P}$. The coloring κ is *proper* if $\kappa(i) \neq \kappa(j)$ whenever $ij \in E(G)$. If $q \in \mathbb{P}$ then let $\chi_G(q)$ denote the number of proper colorings $\kappa : [n] \to [q]$ of G, i.e., the number of proper colorings of G whose colors come from $1, 2, \ldots, q$. The function χ_G is called the *chromatic polynomial* of G.

For instance, suppose that G is the complete graph K_n. A proper coloring $\kappa : [n] \to [q]$ is obtained by choosing a vertex, say 1, and coloring it in q ways. Then choose another vertex, say 2, and color it in $q - 1$ ways, etc., obtaining

$$\chi_{K_n}(q) = q(q-1)\cdots(q-n+1). \tag{17}$$

A similar argument applies to the graph G of Figure 5. There are q ways to color vertex 1, then $q - 1$ to color vertex 2, then $q - 1$ to color vertex 3, etc., obtaining

$$\begin{aligned} \chi_G(q) &= q(q-1)(q-1)(q-2)(q-1)(q-1)(q-2)(q-2)(q-3) \\ &= q(q-1)^4(q-2)^3(q-3). \end{aligned}$$

Unlike the case of the complete graph, in order to obtain this nice product formula one factor at a time only certain orderings of the vertices are suitable. It is not always possible to evaluate the chromatic polynomials "one vertex at a time." For instance, let H be the 4-cycle of Figure 5. If a proper coloring $\kappa : [4] \to [q]$ satisfies $\kappa(1) = \kappa(3)$, then there are q choices for $\kappa(1)$, then $q - 1$ choices each for $\kappa(2)$ and

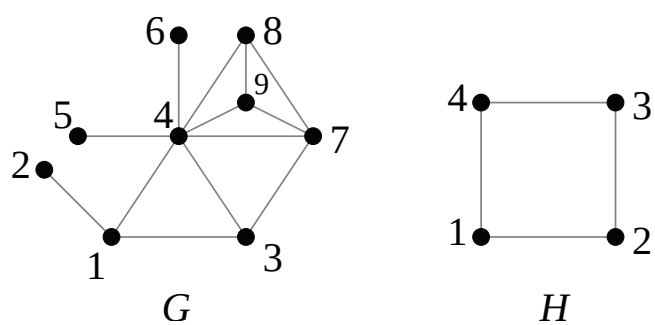


Figure 5. Two graphs

$\kappa(4)$. On the other hand, if $\kappa(1) \neq \kappa(3)$, then there are q choices for $\kappa(1)$, then $q-1$ choices for $\kappa(3)$, and then $q-2$ choices each for $\kappa(2)$ and $\kappa(4)$. Hence

$$\begin{aligned}\chi_H(q) &= q(q-1)^2 + q(q-1)(q-2)^2 \\ &= q(q-1)(q^2-3q+3).\end{aligned}$$

For further information on graphs whose chromatic polynomial *can* be evaluated one vertex at a time, see Corollary 4.10 and the note following it.

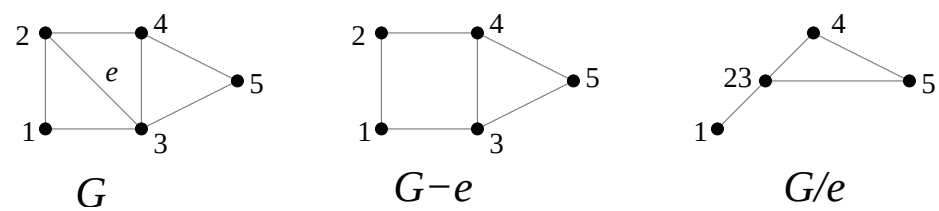
It is easy to see directly that $\chi_G(q)$ is a polynomial function of q. Let $e_i(G)$ denote the number of *surjective* proper colorings $\kappa : [n] \to [i]$ of G. We can choose an arbitrary proper coloring $\kappa : [n] \to [q]$ by first choosing the size $i = \#\kappa([n])$ of its image in $\binom{q}{i}$ ways, and then choose κ in e_i ways. Hence

$$\chi_G(q) = \sum_{i=0}^{n} e_i \binom{q}{i}. \tag{18}$$

Since $\binom{q}{i} = q(q-1)\cdots(q-i+1)/i!$, a polynomial in q (of degree i), we see that $\chi_G(q)$ is a polynomial. We therefore write $\chi_G(t)$, where t is an indeterminate. Moreover, any surjection (= bijection) $\kappa : [n] \to [n]$ is proper. Hence $e_n = n!$. It follows from equation (18) that $\chi_G(t)$ is monic of degree n. Using more sophisticated methods we will later derive further properties of the coefficients of $\chi_G(t)$.

Theorem 2.7. *For any graph G, we have $\chi_{\mathcal{A}_G}(t) = \chi_G(t)$.*

First proof. The first proof is based on deletion-restriction (which in the context of graphs is called *deletion-contraction*). Let $e = ij \in E(G)$. Let $G - e$ (also denoted $G \backslash e$) denote the graph G with edge e deleted, and let G/e denote G with the edge e contracted to a point and all multiple edges replaced by a single edge (i.e., whenever there is more than one edge between two vertices, replace these edges by a single edge). (In some contexts we want to keep track of multiple edges, but they are irrelevant in regard to proper colorings.)



Let $H_0 \in \mathcal{A} = \mathcal{A}_G$ be the hyperplane $x_i = x_j$. It is clear that $\mathcal{A} - \{H_0\} = \mathcal{A}_{G-e}$. We claim that

$$\mathcal{A}^{H_0} = \mathcal{A}_{G/e}, \tag{19}$$

so by Deletion-Restriction (Lemma 2.2) we have

$$\chi_{\mathcal{A}_G}(t) = \chi_{\mathcal{A}_{G-e}}(t) = \chi_{\mathcal{A}_{G/e}}(t).$$

To prove (19), define an affine isomorphism $\varphi : H_0 \xrightarrow{\cong} \mathbb{R}^{n-1}$ by

$$(x_1, x_2, \ldots, x_n) \mapsto (x_1, \ldots, x_i, \ldots, \hat{x_j}, \ldots, x_n),$$

where $\hat{x_j}$ denotes that the jth coordinate is omitted. (Hence the coordinates in $\mathbb{R}^{n-1}$ are $1, 2, \ldots, \hat{j}, \ldots, n$.) Write H_{ab} for the hyperplane $x_a = x_b$ of $\mathcal{A}$. If neither

of a, b are equal to i or j, then $\varphi(H_{ab} \cap H_0)$ is the hyperplane $x_a = x_b$ in $\mathbb{R}^{n-1}$. If $a \neq i, j$ then $\varphi(H_{ia} \cap H_0) = \varphi(H_{aj} \cap H_0)$, the hyperplane $x_a = x_i$ in $\mathbb{R}^{n-1}$. Hence φ defines an isomorphism between $\mathcal{A}^{H_0}$ and the arrangement $\mathcal{A}_{G/e}$ in $\mathbb{R}^{n-1}$, proving (19).

Let $n\bullet$ denote the graph with n vertices and no edges, and let $\emptyset$ denote the empty arrangement in $\mathbb{R}^n$. The theorem will be proved by induction (using Lemma 2.2) if we show:

(a) Initialization: $\chi_{n\bullet}(t) = \chi_\emptyset(t)$
(b) Deletion-contraction:

$$\chi_G(t) = \chi_{G-e}(t) - \chi_{G/e}(t) \tag{20}$$

To prove (a), note that both sides are equal to t^n. To prove (b), observe that $\chi_{G-e}(q)$ is the number of colorings of $\kappa : [n] \to [q]$ that are proper except possibly $\kappa(i) = \kappa(j)$, while $\chi_{G/e}(q)$ is the number of colorings $\kappa : [n] \to [q]$ of G that are proper except that $\kappa(i) = \kappa(j)$. □

Our second proof of Theorem 2.7 is based on Möbius inversion. We first obtain a combinatorial description of the intersection lattice $L(\mathcal{A}_G)$. Let H_{ij} denote the hyperplane $x_i = x_j$ as above, and let $F \subseteq E(G)$. Consider the element $X = \bigcap_{ij \in F} H_{ij}$ of $L(\mathcal{A}_G)$. Thus

$$(x_1, \ldots, x_n) \in X \Leftrightarrow x_i = x_j \text{ whenever } ij \in F.$$

Let $C_1, \ldots, C_k$ be the connected components of the spanning subgraph G_F of G with edge set F. (A subgraph of G is *spanning* if it contains all the vertices of G. Thus if the edges of F do not span all of G, we need to include all remaining vertices as isolated vertices of G_F.) If i, j are vertices of some C_m, then there is a path from i to j whose edges all belong to F. Hence $x_i = x_j$ for all $(x_1, \ldots, x_n) \in X$. On the other hand, if i and j belong to different C_m's, then there is no such path. Let

$$\bar{F} = \{e = ij \in E(G) \,:\, i, j \in V(C_m) \text{ for some } m\},$$

where $V(C_m)$ denotes the vertex set of C_m. Figure 6 illustrates a graph G with a set F of edges indicated by thickening. The set $\bar{F}$ is shown below G, with the additional edges $\bar{F} - F$ not in F drawn as dashed lines.

A *partition* π of a finite set S is a collection $\{B_1, \ldots, B_k\}$ of subsets of S, called *blocks*, that are nonempty, pairwise disjoint, and whose union is S. The set of all

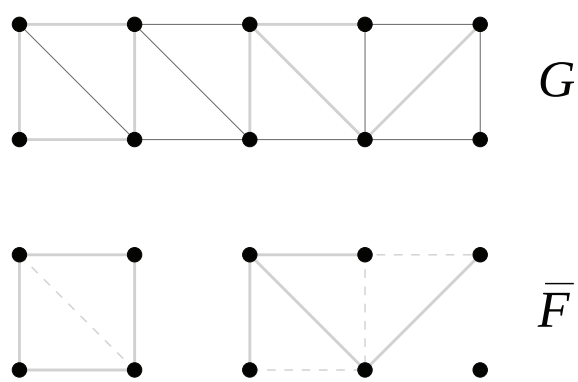


Figure 6. A graph G with edge subset F and closure $\bar{F}$

partitions of S is denoted Π_S, and when $S = [n]$ we write simply Π_n for $\Pi_{[n]}$. It follows from the above discussion that the elements X_π of $L(\mathcal{A}_G)$ correspond to the *connected partitions* of $V(G)$, i.e., the partitions $\pi = \{B_1, \ldots, B_k\}$ of $V(G) = [n]$ such that the restriction of G to each block B_i is connected. Namely,

$$X_\pi = \{(x_1, \ldots, x_n) \in K^n \,:\, i, j \in B_m \text{ for some } m \Rightarrow x_i = x_j\}.$$

We have $X_\pi \leq X_\sigma$ in $L(\mathcal{A})$ if and only if every block of π is contained in a block of σ. In other words, π is a *refinement* of σ. This refinement order is the "standard" ordering on Π_n, so $L(\mathcal{A}_G)$ is isomorphic to an induced subposet L_G of Π_n, called the *bond lattice* or *lattice of contractions* of G. ("Induced" means that if $\pi \leq \sigma$ in Π_n and $\pi, \sigma \in L(\mathcal{A}_G)$, then $\pi \leq \sigma$ in $L(\mathcal{A}_G)$.) In particular, $\Pi_n \cong L(\mathcal{A}_{K_n})$. Note that in general L_G is not a sublattice of Π_n, but only a sub-join-semilattice of Π_n [why?]. The bottom element $\hat{0}$ of L_G is the partition of $[n]$ into n one-element blocks, while the top element $\hat{1}$ is the partition into one block. The case $G = K_n$ shows that the intersection lattice $L(\mathcal{B}_n)$ of the braid arrangement $\mathcal{B}_n$ is isomorphic to the full partition lattice Π_n. Figure 7 shows a graph G and its bond lattice L_G (singleton blocks are omitted from the labels of the elements of L_G).

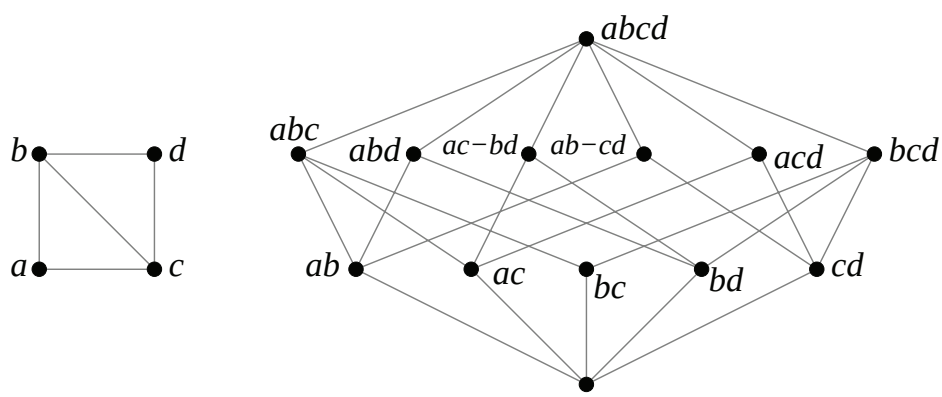


Figure 7. A graph G and its bond lattice L_G

Second proof of Theorem 2.7. Let $\pi \in L_G$. For $q \in \mathbb{P}$ define $\chi_\pi(q)$ to be the number of colorings $\kappa : [n] \to [q]$ of G satisfying:

- If i, j are in the same block of π, then $\kappa(i) = \kappa(j)$.
- If i, j are in different blocks of π and $ij \in E(G)$, then $\kappa(i) \neq \kappa(j)$.

Given *any* $\kappa : [n] \to [q]$, there is a unique $\sigma \in L_G$ such that κ is enumerated by $\chi_\sigma(q)$. Moreover, κ will be constant on the blocks of some $\pi \in L_G$ if and only if $\sigma \geq \pi$ in L_G. Hence

$$q^{|\pi|} = \sum_{\sigma \geq \pi} \chi_\sigma(q) \quad \forall \pi \in L_G,$$

where $|\pi|$ denotes the number of blocks of π. By Möbius inversion,

$$\chi_\pi(q) = \sum_{\sigma \geq \pi} q^{|\sigma|} \mu(\pi, \sigma),$$

where μ denotes the Möbius function of L_G. Let $\pi = \hat{0}$. We get

$$\chi_G(q) = \chi_{\hat{0}}(q) = \sum_{\sigma \in L_G} \mu(\sigma) q^{|\sigma|}. \tag{21}$$

It is easily seen that $|\sigma| = \dim X_\sigma$, so comparing equation (21) with Definition 1.3 shows that $\chi_G(t) = \chi_{\mathcal{A}_G}(t)$. □

Corollary 2.2. *The characteristic polynomial of the braid arrangement $\mathcal{B}_n$ is given by*

$$\chi_{\mathcal{B}_n}(t) = t(t-1)\cdots(t-n+1).$$

Proof. Since $\mathcal{B}_n = \mathcal{A}_{K_n}$ (the graphical arrangement of the complete graph K_n), we have from Theorem 2.7 that $\chi_{\mathcal{B}_n}(t) = \chi_{K_n}(t)$. The proof follows from equation (17). □

There is a further invariant of a graph G that is closely connected with the graphical arrangement $\mathcal{A}_G$.

Definition 2.7. An *orientation* $\mathfrak{o}$ of a graph G is an assignment of a direction $i \to j$ or $j \to i$ to each edge ij of G. A *directed cycle* of $\mathfrak{o}$ is a sequence of vertices $i_0, i_1, \ldots, i_k$ of G such that $i_0 \to i_1 \to i_2 \to \cdots \to i_k \to i_0$ in $\mathfrak{o}$. An orientation $\mathfrak{o}$ is *acyclic* if it contains no directed cycles.

A graph G with no loops (edges from a vertex to itself) thus has $2^{\#E(G)}$ orientations. Let $R \in \mathcal{R}(\mathcal{A}_G)$, and let $(x_1, \ldots, x_n) \in R$. In choosing R, we have specified for all $ij \in E(G)$ whether $x_i < x_j$ or $x_i > x_j$. Indicate by an arrow $i \to j$ that $x_i < x_j$, and by $j \to i$ that $x_i > x_j$. In this way the region R defines an orientation $\mathfrak{o}_R$ of G. Clearly if $R \neq R'$, then $\mathfrak{o}_R \neq \mathfrak{o}_{R'}$. Which orientations can arise in this way?

Proposition 2.5. *Let $\mathfrak{o}$ be an orientation of G. Then $\mathfrak{o} = \mathfrak{o}_R$ for some $R \in \mathcal{R}(\mathcal{A}_G)$ if and only if $\mathfrak{o}$ is acyclic.*

Proof. If $\mathfrak{o}_R$ had a cycle $i_1 \to i_2 \to \cdots \to i_k \to i_1$, then a point $(x_1, \ldots, x_n) \in R$ would satisfy $x_{i_1} < x_{i_2} < \cdots < x_{i_k} < x_{i_1}$, which is absurd. Hence $\mathfrak{o}_R$ is acyclic.

Conversely, let $\mathfrak{o}$ be an acyclic orientation of G. First note that $\mathfrak{o}$ must have a *sink*, i.e., a vertex with no arrows pointing out. To see this, walk along the edges of $\mathfrak{o}$ by starting at any vertex and following arrows. Since $\mathfrak{o}$ is acyclic, we can never return to a vertex so the process will end in a sink. Let j_n be a sink vertex of $\mathfrak{o}$. When we remove j_n from $\mathfrak{o}$ the remaining orientation is still acyclic, so it contains a sink j_{n-1}. Continuing in this manner, we obtain an ordering $j_1, j_2, \ldots, j_n$ of $[n]$ such that j_i is a sink of the restriction of $\mathfrak{o}$ to $j_1, \ldots, j_i$. Hence if $x_1, \ldots, x_n \in \mathbb{R}$ satisfy $x_{j_1} < x_{j_2} < \cdots < x_{j_n}$ then the region $R \in \mathcal{R}(\mathcal{A})$ containing $(x_1, \ldots, x_n)$ satisfies $\mathfrak{o} = \mathfrak{o}_R$. □

NOTE. The transitive, reflexive closure $\bar{\mathfrak{o}}$ of an acyclic orientation $\mathfrak{o}$ is a partial order. The construction of the ordering $j_1, j_2, \ldots, j_n$ above is equivalent to constructing a linear extension of $\mathfrak{o}$.

Let $\mathrm{AO}(G)$ denote the set of acyclic orientations of G. We have constructed a bijection between $\mathrm{AO}(G)$ and $\mathcal{R}(\mathcal{A}_G)$. Hence from Theorem 2.5 we conclude:

Corollary 2.3. *For any graph G with n vertices, we have*

$$\#\mathrm{AO}(G) = (-1)^n \chi_G(-1).$$

Corollary 2.3 was first proved by Stanley in 1973 by a "direct" argument based on deletion-contraction (see Exercise 7). The proof we have just given based on arrangements is due to Greene and Zaslavsky in 1983.

NOTE. Given a graph G on n vertices, let $\mathcal{A}_G^{\#}$ be the arrangement defined by

$$x_i - x_j = a_{ij}, \quad ij \in E(G),$$

where the a_{ij}'s are generic. Just as we obtained equation (14) (the case $G = K_n$) we have

$$\chi_{\mathcal{A}_G^{\#}}(t) = \sum_F (-1)^{e(F)} t^{n-e(F)},$$

where F ranges over all spanning forests of G.

Exercises

(1) [3–] Show that for any arrangement $\mathcal{A}$, we have $\chi_{c\mathcal{A}}(t) = (t-1)\chi_{\mathcal{A}}(t)$, where $c\mathcal{A}$ denotes the cone over $\mathcal{A}$. (Use Whitney's theorem.)

(2) [2–] Let G be a graph on the vertex set $[n]$. Show that the bond lattice L_G is a sub-join-semilattice of the partition lattice Π_n but is not in general a sublattice of Π_n.

(3) [2–] Let G be a forest (graph with no cycles) on the vertex set $[n]$. Show that $L_G \cong B_{E(G)}$, the boolean algebra of all subsets of $E(G)$.

(4) [2] Let G be a graph with n vertices and $\mathcal{A}_G$ the corresponding graphical arrangement. Suppose that G has a k-element clique, i.e., k vertices such that any two are adjacent. Show that $k!|r(\mathcal{A})$.

(5) [2+] Let G be a graph on the vertex set $[n] = \{1, 2, \ldots, n\}$, and let $\mathcal{A}_G$ be the corresponding graphical arrangement (over any field K, but you may assume $K = \mathbb{R}$ if you wish). Let $\mathcal{C}_n$ be the coordinate hyperplane arrangement, consisting of the hyperplanes $x_i = 0$, $1 \le i \le n$. Express $\chi_{\mathcal{A}_G \cup \mathcal{C}_n}(t)$ in terms of $\chi_{\mathcal{A}_G}(t)$.

(6) [4] Let G be a planar graph, i.e., G can be drawn in the plane without crossing edges. Show that $\chi_{\mathcal{A}_G}(4) \neq 0$.

(7) [2+] Let G be a graph with n vertices. Show directly from the the deletion-contraction recurrence (20) that

$$(-1)^n \chi_G(-1) = \#\mathrm{AO}(G).$$

(8) [2+] Let $\chi_G(t) = t^n - c_{n-1}t^{n-1} + \cdots + (-1)^{n-1}c_1 t$ be the chromatic polynomial of the graph G. Let i be a vertex of G. Show that c_1 is equal to the number of acyclic orientations of G whose unique source is i. (A *source* is a vertex with no arrows pointing in. In particular, an isolated vertex is a source.)

(9) [5] Let $\mathcal{A}$ be an arrangement with characteristic polynomial $\chi_{\mathcal{A}}(t) = t^n - c_{n-1}t^{n-1} + c_{n-2}t^{n-2} - \cdots + (-1)^n c_0$. Show that the sequence $c_0, c_1, \ldots, c_n = 1$ is *unimodal*, i.e., for some j we have

$$c_0 \le c_1 \le \cdots \le c_j \ge c_{j+1} \ge \cdots \ge c_n.$$

(10) [2+] Let $f(n)$ be the total number of faces of the braid arrangement $\mathcal{B}_n$. Find a simple formula for the generating function

$$\sum_{n\ge 0} f(n)\frac{x^n}{n!} = 1 + x + 3\frac{x^2}{2!} + 13\frac{x^3}{3!} + 75\frac{x^4}{4!} + 541\frac{x^5}{5!} + 4683\frac{x^6}{6!} + \cdots.$$

More generally, let $f_k(n)$ denote the number of k-dimensional faces of $\mathcal{B}_n$. For instance, $f_1(n) = 1$ (for $n \ge 1$) and $f_n(n) = n!$. Find a simple formula for the generating function

$$\sum_{n\ge 0}\sum_{k\ge 0} f_k(n) y^k \frac{x^n}{n!} = 1 + yx + (y + 2y^2)\frac{x^2}{2!} + (y + 6y^2 + 6y^3)\frac{x^3}{3!} + \cdots.$$

LECTURE 3
Matroids and Geometric Lattices

3.1. Matroids

A matroid is an abstraction of a set of vectors in a vector space (for us, the normals to the hyperplanes in an arrangement). Many basic facts about arrangements (especially linear arrangements) and their intersection posets are best understood from the more general viewpoint of matroid theory. There are many equivalent ways to define matroids. We will define them in terms of independent sets, which are an abstraction of linearly independent sets. For any set S we write

$$2^S = \{T : T \subseteq S\}.$$

Definition 3.8. A (finite) *matroid* is a pair $M = (S, \mathcal{I})$, where S is a finite set and $\mathcal{I}$ is a collection of subsets of S, satisfying the following axioms:

(1) $\mathcal{I}$ is a nonempty (abstract) *simplicial complex*, i.e., $\mathcal{I} \neq \emptyset$, and if $J \in \mathcal{I}$ and $I \subset J$, then $I \in \mathcal{I}$.
(2) For all $T \subseteq S$, the maximal elements of $\mathcal{I} \cap 2^T$ have the same cardinality. In the language of simplicial complexes, every induced subcomplex of $\mathcal{I}$ is *pure*.

The elements of $\mathcal{I}$ are called *independent sets*. All matroids considered here will be assumed to be finite. By standard abuse of notation, if $M = (S, \mathcal{I})$ then we write $x \in M$ to mean $x \in S$. The archetypal example of a matroid is a finite subset S of a vector space, where independence means linear independence. A closely related matroid consists of a finite subset S of an affine space, where independence now means affine independence.

It should be clear what is meant for two matroids $M = (S, \mathcal{I})$ and $M' = (S', \mathcal{I}')$ to be *isomorphic*, viz., there exists a bijection $f : S \to S'$ such that $\{x_1, \ldots, x_j\} \in \mathcal{I}$ if and only if $\{f(x_1), \ldots, f(x_j)\} \in \mathcal{I}'$. Let M be a matroid and S a set of points in $\mathbb{R}^n$, regarded as a matroid with independence meaning affine independence. If M and S are isomorphic matroids, then S is called an *affine diagram* of M. (Not all matroids have affine diagrams.)

Example 3.7. (a) Regard the configuration in Figure 1 as a set of five points in the two-dimensional affine space $\mathbb{R}^2$. These five points thus define the affine diagram of a matroid M. The lines indicate that the points 1,2,3 and 3,4,5 lie on straight

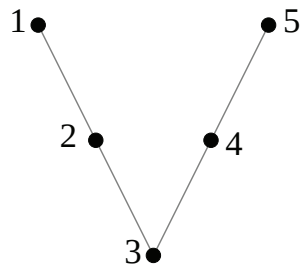


Figure 1. A five-point matroid in the affine space $\mathbb{R}^2$

lines. Hence the sets $\{1,2,3\}$ and $\{3,4,5\}$ are affinely dependent in $\mathbb{R}^2$ and therefore dependent (i.e., not independent) in M. The independent sets of M consist of all subsets of [5] with at most two elements, together with all three-element subsets of [5] except 123 and 345 (where 123 is short for $\{1,2,3\}$, etc.).

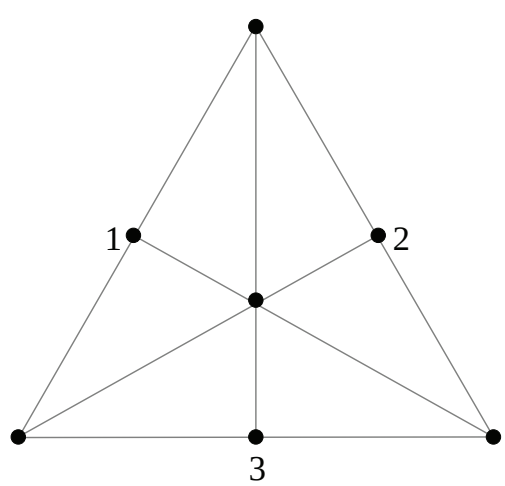
(b) Write $\mathcal{I} = \langle S_1, \ldots, S_k \rangle$ for the simplicial complex $\mathcal{I}$ *generated* by $S_1, \ldots, S_k$, i.e.,

$$\begin{aligned} \langle S_1, \ldots, S_k \rangle &= \{T : T \subseteq S_i \text{ for some } i\} \\ &= 2^{S_1} \cup \cdots \cup 2^{S_k}. \end{aligned}$$

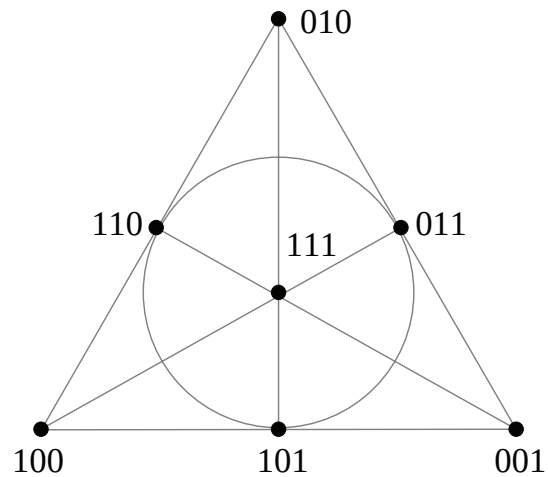
Then $\mathcal{I} = \langle 13, 14, 23, 24 \rangle$ is the set of independent sets of a matroid M on [4]. This matroid is realized by a *multiset* of vectors in a vector space or affine space, e.g., by the points 1,1,2,2 in the affine space $\mathbb{R}$. The affine diagam of this matroid is given by

(c) Let $\mathcal{I} = \langle 12, 23, 34, 45, 15 \rangle$. Then $\mathcal{I}$ is *not* the set of independent sets of a matroid. For instance, the maximal elements of $\mathcal{I} \cap 2^{\{1,2,4\}}$ are 12 and 4, which do not have the same cardinality.

(d) The affine diagram below shows a seven point matroid.



If we further require the points labelled 1,2,3 to lie on a line (i.e., remove 123 from $\mathcal{I}$), we still have a matroid M, but not one that can be realized by real vectors. In fact, M is isomorphic to the set of nonzero vectors in the vector space $\mathbb{F}_2^3$, where $\mathbb{F}_2$ denotes the two-element field.



Let us now define a number of important terms associated to a matroid M. A *basis* of M is a maximal independent set. A *circuit* C is a minimal dependent set, i.e., C is not independent but becomes independent when we remove any point from it. For example, the circuits of the matroid of Figure 1 are 123, 345, and 1245.

If $M = (S, \mathcal{I})$ is a matroid and $T \subseteq S$ then define the *rank* $\mathrm{rk}(T)$ of T by

$$\mathrm{rk}(T) = \max\{\#I \,:\, I \in \mathcal{I} \text{ and } I \subseteq T\}.$$

In particular, $\mathrm{rk}(\emptyset) = 0$. We define the rank of the matroid M itself by $\mathrm{rk}(M) = \mathrm{rk}(S)$. A *$k$-flat* is a maximal subset of rank k. For instance, if M is an affine matroid, i.e., if S is a subset of an affine space and independence in M is given by affine independence, then the flats of M are just the intersections of S with affine subspaces. Note that if F and F' are flats of a matroid M, then so is $F \cap F'$ (see Exercise 2). Since the intersection of flats is a flat, we can define the *closure* $\overline{T}$ of a subset $T \subseteq S$ to be the smallest flat containing T, i.e.,

$$\overline{T} = \bigcap_{\text{flats } F \supseteq T} F.$$

This closure operator has a number of nice properties, such as $\overline{\overline{T}} = \overline{T}$ and $T' \subseteq T \Rightarrow \overline{T}' \subseteq \overline{T}$.

3.2. The Lattice of Flats and Geometric Lattices

For a matroid M define $L(M)$ to be the poset of flats of M, ordered by inclusion. Since the intersection of flats is a flat, $L(M)$ is a meet-semilattice; and since $L(M)$ has a top element S, it follows from Lemma 2.3 that $L(M)$ is a lattice, which we call the *lattice of flats* of M. Note that $L(M)$ has a unique minimal element $\hat{0}$, viz., $\overline{\emptyset}$ or equivalently, the intersection of all flats. It is easy to see that $L(M)$ is graded by rank, i.e., every maximal chain of $L(M)$ has length $m = \mathrm{rk}(M)$. Thus if $x \lessdot y$ in

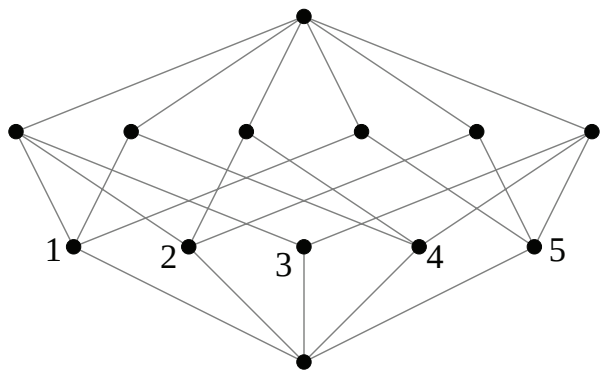


Figure 2. The lattice of flats of the matroid of Figure 1

$L(M)$ then $\mathrm{rk}(y) = 1 + \mathrm{rk}(x)$. We now define the *characteristic polynomial* $\chi_M(t)$, in analogy to the definition (3) of $\chi_{\mathcal{A}}(t)$, by

$$\chi_M(t) = \sum_{x \in L(M)} \mu(\hat{0}, x) t^{m - \mathrm{rk}(x)}, \tag{22}$$

where μ denotes the Möbius function of $L(M)$ and $m = \mathrm{rk}(M)$. Figure 2 shows the lattice of flats of the matroid M of Figure 1. From this figure we see easily that

$$\chi_M(t) = t^3 - 5t^2 + 8t - 4.$$

Let M be a matroid and $x \in M$. If the set $\{x\}$ is dependent (i.e., if $\mathrm{rk}(\{x\}) = 0$) then we call x a *loop*. Thus $\bar{\emptyset}$ is just the set of loops of M. Suppose that $x, y \in M$, neither x nor y are loops, and $\mathrm{rk}(\{x, y\}) = 1$. We then call x and y *parallel* points. A matroid is *simple* if it has no loops or pairs of parallel points. It is clear that the following three conditions are equivalent:

- M is simple.
- $\bar{\emptyset} = \emptyset$ and $\bar{x} = x$ for all $x \in M$.
- $\mathrm{rk}(\{x, y\}) = 2$ for all points $x \neq y$ of M (assuming M has at least two points).

For any matroid M and $x, y \in M$, define $x \sim y$ if $\bar{x} = \bar{y}$. It is easy to see that $\sim$ is an equivalence relation. Let

$$\widehat{M} = \{\bar{x} \,:\, x \in M,\ x \notin \bar{\emptyset}\}, \tag{23}$$

with an obvious definition of independence, i.e.,

$$\{\bar{x}_1, \ldots, \bar{x}_k\} \in \mathcal{I}(\widehat{M}) \Leftrightarrow \{x_1, \ldots, x_k\} \in \mathcal{I}(M).$$

Then $\widehat{M}$ is simple, and $L(M) \cong L(\widehat{M})$. Thus insofar as intersection lattices $L(M)$ are concerned, we may assume that M is simple. (Readers familiar with point set topology will recognize the similarity between the conditions for a matroid to be simple and for a topological space to be T_0.)

Example 3.8. Let S be any finite set and V a vector space. If $f : S \to V$, then define a matroid M_f on S by the condition that given $I \subseteq S$,

$$I \in \mathcal{I}(M) \Leftrightarrow \{f(x) \,:\, x \in I\} \text{ is linearly independent.}$$

Then a loop is any element x satisfying $f(x) = 0$, and $x \sim y$ if and only if $f(x)$ is a nonzero scalar multiple of $f(y)$.

NOTE. If $M = (S, \mathcal{I})$ is simple, then $L(M)$ determines M. For we can identify S with the set of atoms of $L(M)$, and we have

$$\{x_1, \ldots, x_k\} \in \mathcal{I} \;\Leftrightarrow\; \mathrm{rk}(x_1 \vee \cdots \vee x_k) = k \text{ in } L(M).$$

See the proof of Theorem 3.8 for further details.

We now come to the primary connection between hyperplane arrangements and matroid theory. If H is a hyperplane, write $\mathfrak{n}_H$ for some (nonzero) normal vector to H.

Proposition 3.6. *Let $\mathcal{A}$ be a* central *arrangement in the vector space V. Define a matroid $M = M_{\mathcal{A}}$ on $\mathcal{A}$ by letting $\mathcal{B} \in \mathcal{I}(M)$ if $\mathcal{B}$ is linearly independent (i.e., $\{\mathfrak{n}_H : H \in \mathcal{B}\}$ is linearly independent). Then M is simple and $L(M) \cong L(\mathcal{A})$.*

Proof. M has no loops, since every $H \in \mathcal{A}$ has a nonzero normal. Two distinct nonparallel hyperplanes have linearly independent normals, so the points of M are closed. Hence M is simple.

Let $\mathcal{B}, \mathcal{B}' \subseteq \mathcal{A}$, and set

$$X = \bigcap_{H \in \mathcal{B}} H = X_{\mathcal{B}}, \quad X' = \bigcap_{H \in \mathcal{B}'} H = X_{\mathcal{B}'}.$$

Then $X = X'$ if and only if

$$\mathrm{span}\{\mathfrak{n}_H : H \in \mathcal{B}\} = \mathrm{span}\{\mathfrak{n}_H : H \in \mathcal{B}'\}.$$

Now the closure relation in M is given by

$$\overline{\mathcal{B}} = \{H' \in \mathcal{A} : \mathfrak{n}_{H'} \in \mathrm{span}\{\mathfrak{n}_H : H \in \mathcal{B}\}\}.$$

Hence $X = X'$ if and only if $\overline{B} = \overline{B}'$, so $L(M) \cong L(\mathcal{A})$. □

It follows that for a central arrangement $\mathcal{A}$, $L(\mathcal{A})$ depends only on the *matroidal* structure of $\mathcal{A}$, i.e., which subsets of hyperplanes are linearly independent. Thus the matroid $M_{\mathcal{A}}$ encapsulates the essential information about $\mathcal{A}$ needed to define $L(\mathcal{A})$.

Our next goal is to characterize those lattices L which have the form $L(M)$ for some matroid M.

Proposition 3.7. *Let L be a finite graded lattice. The following two conditions are equivalent.*

(1) *For all $x, y \in L$, we have* $\mathrm{rk}(x) + \mathrm{rk}(y) \geq \mathrm{rk}(x \wedge y) + \mathrm{rk}(x \vee y)$.
(2) *If x and y both cover $x \wedge y$, then $x \vee y$ covers both x and y.*

Proof. Assume (1). Let $x, y \gtrdot x \wedge y$, so $\mathrm{rk}(x) = \mathrm{rk}(y) = \mathrm{rk}(x \wedge y) + 1$ and $\mathrm{rk}(x \vee y) > \mathrm{rk}(x) = \mathrm{rk}(y)$. By (1),

$$\begin{aligned} \mathrm{rk}(x) + \mathrm{rk}(y) &\geq (\mathrm{rk}(x) - 1) + \mathrm{rk}(x \vee y) \\ \Rightarrow \mathrm{rk}(y) &\geq \mathrm{rk}(x \vee y) - 1 \\ \Rightarrow x \vee y &\gtrdot x. \end{aligned}$$

Similarly $x \vee y \gtrdot y$, proving (2).

For (2)⇒(1), see [**31**, Prop. 3.3.2]. □

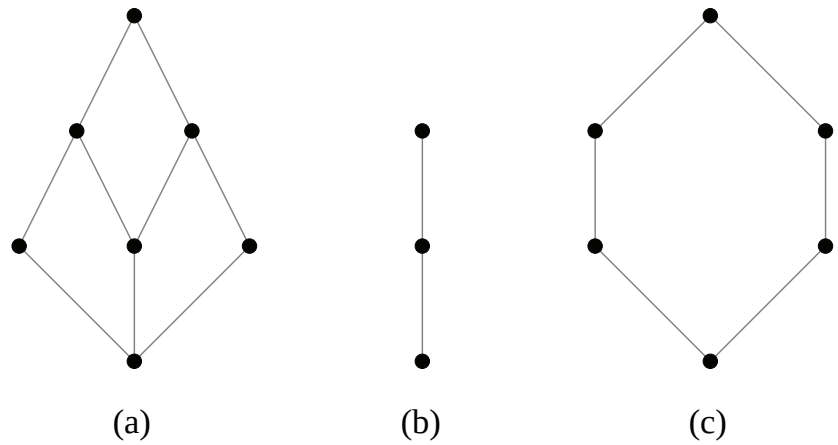


Figure 3. Three nongeometric lattices

Definition 3.9. A finite lattice L satisfying condition (1) or (2) above is called *(upper) semimodular*. A finite lattice L is *atomic* if every $x \in L$ is a join of atoms (where we regard $\hat{0}$ as an empty join of atoms). Equivalently, if $x \in L$ is join-irreducible (i.e., covers a unique element), then x is an atom. Finally, a finite lattice is *geometric* if it is both semimodular and atomic.

To illustrate these definitions, Figure 3(a) shows an atomic lattice that is not semimodular, (b) shows a semimodular lattice that is not atomic, and (c) shows a graded lattice that is neither semimodular nor atomic.

We are now ready to characterize the lattice of flats of a matroid.

Theorem 3.8. *Let L be a finite lattice. The following two conditions are equivalent.*

(1) *L is a geometric lattice.*
(2) *$L \cong L(M)$ for some (simple) matroid M.*

Proof. Assume that L is geometric, and let A be the set of atoms of L. If $T \subseteq A$ then write $\bigvee T = \bigvee_{x \in T} x$, the join of all elements of T. Let

$$\mathcal{I} = \{I \subseteq A \,:\, \mathrm{rk}(\vee I) = \#I\}.$$

Note that by semimodularity, we have for any $S \subseteq A$ and $x \in A$ that $\mathrm{rk}((\bigvee S) \vee x) \leq \mathrm{rk}(\bigvee S) + 1$. (Hence in particular, $\mathrm{rk}(\bigvee S) \leq \#S$.) It follows that $\mathcal{I}$ is a simplicial complex. Let $S \subseteq A$, and let T, T' be maximal elements of $2^S \cap \mathcal{I}$. We need to show that $\#T = \#T'$.

Assume $\#T < \#T'$, say. If $y \in S$ then $y \leq \bigvee T'$, else $T'' = T' \cup y$ satisfies $\mathrm{rk}(\bigvee T'') = \#T''$, contradicting the maximality of T'. Since $\#T < \#T'$ and $T \subseteq S$, it follows that $\bigvee T < \bigvee T'$ [why?]. Since L is atomic, there exists $y \in S$ such that $y \in S$ but $y \not\leq \bigvee T$. But then $\mathrm{rk}(\bigvee(T \cup y)) = 1 + \#T$, contradicting the maximality of T. Hence $M = (A, \mathcal{I})$ is a matroid, and $L \cong L(M)$.

Conversely, given a matroid M, which we may assume is simple, we need to show that $L(M)$ is a geometric lattice. Clearly $L(M)$ is atomic, since every flat is the join of its elements. Let $S, T \subseteq M$. We will show that

$$\mathrm{rk}(S) + \mathrm{rk}(T) \geq \mathrm{rk}(S \cap T) + \mathrm{rk}(S \cup T). \tag{24}$$

Note that if S and T are *flats* (i.e., $S, T \in L(M)$) then $S \cap T = S \wedge T$ and $\mathrm{rk}(S \cup T) = \mathrm{rk}(S \vee T)$. Hence taking S and T to be flats in (24) shows that $L(M)$ is semimodular and thus geometric. Suppose (24) is false, so

$$\mathrm{rk}(S \cup T) > \mathrm{rk}(S) + \mathrm{rk}(T) - \mathrm{rk}(S \cap T).$$

Let B be a basis for $S \cup T$ extending a basis for $S \cup T$. Then either $\#(B \cap S) > \mathrm{rk}(S)$ or $\#(B \cap T) > \mathrm{rk}(T)$, a contradiction completing the proof. □

Note that by Proposition 3.6 and Theorem 3.8, any results we prove about geometric lattices hold a fortiori for the intersection lattice $L_{\mathcal{A}}$ of a central arrangement $\mathcal{A}$.

NOTE. If L is geometric and $x \leq y$ in L, then it is easy to show using semimodularity that the interval $[x, y]$ is also a geometric lattice. (See Exercise 3.) In general, however, an interval of an atomic lattice need not be atomic.

For noncentral arrangements $L(\mathcal{A})$ is not a lattice, but there is still a connection with geometric lattices. For a stronger statement, see Exercise 4.

Proposition 3.8. *Let $\mathcal{A}$ be an arrangement. Then every interval $[x, y]$ of $L(\mathcal{A})$ is a geometric lattice.*

Proof. By Exercise 3, it suffices to take $x = \hat{0}$. Now $[\hat{0}, y] \cong L(\mathcal{A}_y)$, where $\mathcal{A}_y$ is given by (6). Since $\mathcal{A}_y$ is a central arrangement, the proof follows from Proposition 3.6. □

The proof of our next result about geometric lattices will use a fundamental formula concerning Möbius functions known as *Weisner's theorem*. For a proof, see [**31**, Cor. 3.9.3] (where it is stated in dual form).

Theorem 3.9. *Let L be a finite lattice with at least two elements and with Möbius function μ. Let $\hat{0} \neq a \in L$. Then*

$$\sum_{x\,:\,x \vee a = \hat{1}} \mu(x) = 0. \tag{25}$$

Note that Theorem 3.9 gives a "shortening" of the recurrence (2) defining μ. Normally we take a to be an atom, since that produces fewer terms in (25) than choosing any $b > a$. As an example, let $L = B_n$, the boolean algebra of all subsets of $[n]$, and let $a = \{n\}$. There are two elements $x \in B_n$ such that $x \vee a = \hat{1} = [n]$, viz., $x_1 = [n-1]$ and $x_2 = [n]$. Hence $\mu(x_1) + \mu(x_2) = 0$. Since $[\hat{0}, x_1] = B_{n-1}$ and $[\hat{0}, x_2] = B_n$, we easily obtain $\mu_{B_n}(\hat{1}) = (-1)^n$, agreeing with (4).

If $x \leq y$ in a graded lattice L, write $\mathrm{rk}(x, y) = \mathrm{rk}(y) - \mathrm{rk}(x)$, the length of every saturated chain from x to y. The next result may be stated as "the Möbius function of a geometric lattice strictly alternates in sign."

Theorem 3.10. *Let L be a finite geometric lattice with Möbius function μ, and let $x \leq y$ in L. Then*

$$(-1)^{\mathrm{rk}(x,y)} \mu(x, y) > 0.$$

Proof. Since every interval of a geometric lattice is a geometric lattice (Exercise 3), it suffices to prove the theorem for $[x, y] = [\hat{0}, \hat{1}]$. The proof is by induction on the rank of L. It is clear if $\mathrm{rk}(L) = 1$, in which case $\mu(\hat{0}, \hat{1}) = -1$. Assume the result for geometric lattices of rank $< n$, and let $\mathrm{rk}(L) = n$. Let a be an atom of L in Theorem 3.9. For any $y \in L$ we have by semimodularity that

$$\mathrm{rk}(y \wedge a) + \mathrm{rk}(y \vee a) \leq \mathrm{rk}(y) + \mathrm{rk}(a) = \mathrm{rk}(y) + 1.$$

Hence $x \vee a = \hat{1}$ if and only if $x = \hat{1}$ or x is a coatom (i.e., $x \lessdot \hat{1}$) satisfying $a \not\leq x$. From Theorem 3.9 there follows

$$\mu(\hat{0}, \hat{1}) = -\sum_{a \not\leq x \lessdot \hat{1}} \mu(\hat{0}, x).$$

The sum on the right is nonempty since L is atomic, and by induction every x indexing the sum satisfies $(-1)^{n-1}\mu(\hat{0}, x) > 0$. Hence $(-1)^n \mu(\hat{0}, \hat{1}) > 0$. □

Combining Proposition 3.8 and Theorem 3.10 yields the following result.

Corollary 3.4. *Let $\mathcal{A}$ be any arrangement and $x \leq y$ in $L(\mathcal{A})$. Then*

$$(-1)^{\mathrm{rk}(x,y)}\mu(x, y) > 0,$$

where μ denotes the Möbius function of $L(\mathcal{A})$.

Similarly, combining Theorem 3.10 with the definition (22) of $\chi_M(t)$ gives the next corollary.

Corollary 3.5. *Let M be a matroid of rank n. Then the characteristic polynomial $\chi_M(t)$ strictly alternates in sign, i.e., if*

$$\chi_M(t) = a_n t^n + a_{n-1}t^{n-1} + \cdots + a_0,$$

then $(-1)^{n-i}a_i > 0$ for $0 \leq i \leq n$.

Let $\mathcal{A}$ be an n-dimensional arrangement of rank r. If $M_{\mathcal{A}}$ is the matroid corresponding to $\mathcal{A}$, as defined in Proposition 3.6, then

$$\chi_{\mathcal{A}}(t) = t^{n-r}\chi_M(t). \tag{26}$$

It follows from Corollary 3.5 and equation (26) that we can write

$$\chi_{\mathcal{A}}(t) = b_n t^n + b_{n-1}t^{n-1} + \cdots + b_{n-r}t^{n-r},$$

where $(-1)^{n-i}b_i > 0$ for $n - r \leq i \leq n$.

Exercises

(1) (a) [1+] Let $\chi_G(t)$ be the characteristic polynomial of the graphical arrangement $\mathcal{A}_G$. Suppose that $\chi_G(i) = 0$, where $i \in \mathbb{Z}$, $i > 1$. Show that $\chi_G(i-1) = 0$.
  (b) [2] Is the same conclusion true for *any* central arrangement $\mathcal{A}$?

(2) [2] Show that if F and F' are flats of a matroid M, then so is $F \cap F'$.

(3) [2] Prove the assertion in the Note following the proof of Theorem 3.8 that an interval $[x, y]$ of a geometric lattice L is also a geometric lattice.

(4) [2–] Let $\mathcal{A}$ be an arrangement (not necessarily central), and let $c\mathcal{A}$ denote the cone over $\mathcal{A}$. Show that there exists an atom a of $L(c\mathcal{A})$ such that $L(\mathcal{A}) \cong L(c\mathcal{A}) - V_a$, where $V_a = \{x \in L : x \geq a\}$.

(5) [2–] Let L be a geometric lattice of rank n, and define the *truncation* $T(L)$ to be the subposet of L consisting of all elements of rank $\neq n-1$. Show that $T(L)$ is a geometric lattice.

(6) Let W_i be the number of elements of rank i in a geometric lattice (or just in the intersection poset of a central hyperplane arrangement, if you prefer) of rank n.
  (a) [3] Show that for $k \leq n/2$,

$$W_1 + W_2 + \cdots + W_k \leq W_{n-k} + W_{n-k+1} + \cdots + W_{n-1}.$$

  (b) [2–] Deduce from (a) and Exercise 5 that $W_1 \leq W_k$ for all $1 \leq k \leq n-1$.

(c) [5] Show that $W_i \leq W_{n-i}$ for $i < n/2$ and that the sequence $W_0, W_1, \ldots, W_n$ is unimodal. (Compare Lecture 2, Exercise 9.)

(7) [3–] Let $x \leq y$ in a geometric lattice L. Show that $\mu(x, y) = \pm 1$ if and only if the interval $[x, y]$ is isomorphic to a boolean algebra. (Use Weisner's theorem.) **Note.** This problem becomes much easier using Theorem 4.12 (the Broken Circuit Theorem); see Exercise 4.13.

LECTURE 4
Broken Circuits, Modular Elements, and Supersolvability

This lecture is concerned primarily with matroids and geometric lattices. Since the intersection lattice of a central arrangement is a geometric lattice, all our results can be applied to arrangements.

4.1. Broken Circuits

For any geometric lattice L and $x \leq y$ in L, we have seen (Theorem 3.10) that $(-1)^{\mathrm{rk}(x,y)}\mu(x,y)$ is a positive integer. It is thus natural to ask whether this integer has a direct combinatorial interpretation. To this end, let M be a matroid on the set $S = \{u_1, \ldots, u_m\}$. Linearly order the elements of S, say $u_1 < u_2 < \cdots < u_m$. Recall that a circuit of M is a minimal dependent subset of S.

Definition 4.10. A *broken circuit* of M (with respect to the linear ordering $\mathcal{O}$ of S) is a set $C - \{u\}$, where C is a circuit and u is the largest element of C (in the ordering $\mathcal{O}$). The *broken circuit complex* $\mathrm{BC}_{\mathcal{O}}(M)$ (or just $\mathrm{BC}(M)$ if no confusion will arise) is defined by

$$\mathrm{BC}(M) = \{T \subseteq S \,:\, T \text{ contains no broken circuit}\}.$$

Figure 1 shows two linear orderings $\mathcal{O}$ and $\mathcal{O}'$ of the points of the affine matroid M of Figure 1 (where the ordering of the points is $1 < 2 < 3 < 4 < 5$). With respect to the first ordering $\mathcal{O}$ the circuits are 123, 345, 1245, and the broken circuits are 12, 34, 124. With respect to the second ordering $\mathcal{O}'$ the circuits are 123, 145, 2345, and the broken circuits are 12, 14, 234.

It is clear that the broken circuit complex $\mathrm{BC}(M)$ is an abstract simplicial complex, i.e., if $T \in \mathrm{BC}(M)$ and $U \subseteq T$, then $U \in \mathrm{BC}(M)$. In Figure 1 we

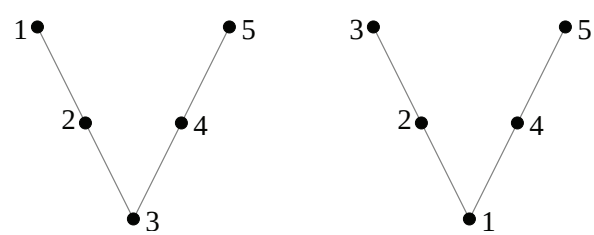
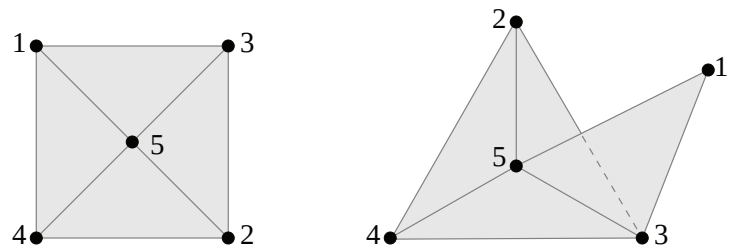


Figure 1. Two linear orderings of the matroid M of Figure 1

have $\mathrm{BC}_{\mathcal{O}}(M) = \langle 135, 145, 235, 245\rangle$, while $\mathrm{BC}_{\mathcal{O}'}(M) = \langle 135, 235, 245, 345\rangle$. These simplicial complexes have geometric realizations as follows:



Note that the two simplicial complexes $\mathrm{BC}_{\mathcal{O}}(M)$ and $\mathrm{BC}_{\mathcal{O}'}(M)$ are not isomorphic (as abstract simplicial complexes); in fact, their geometric realizations are not even homeomorphic. On the other hand, if $f_i(\Delta)$ denotes the number of i-dimensional faces (or faces of cardinality $i-1$) of the abstract simplicial complex Δ, then for Δ given by either $\mathrm{BC}_{\mathcal{O}}(M)$ or $\mathrm{BC}_{\mathcal{O}'}(M)$ we have

$$f_{-1}(\Delta) = 1,\ f_0(\Delta) = 5,\ f_1(\Delta) = 8,\ f_2(\Delta) = 4.$$

Note, moreover, that

$$\chi_M(t) = t^3 - 5t^2 + 8t - 4.$$

In order to generalize this observation to arbitrary matroids, we need to introduce a fair amount of machinery, much of it of interest for its own sake. First we give a fundamental formula, known as *Philip Hall's theorem*, for the Möbius function value $\mu(\hat{0}, \hat{1})$.

Lemma 4.4. *Let P be a finite poset with $\hat{0}$ and $\hat{1}$, and with Möbius function μ. Let c_i denote the number of chains $\hat{0} = y_0 < y_1 < \cdots < y_i = \hat{1}$ in P. Then*

$$\mu(\hat{0}, \hat{1}) = -c_1 + c_2 - c_3 + \cdots .$$

Proof. We work in the incidence algebra $\mathcal{I}(P)$. We have

$$\begin{aligned} \mu(\hat{0}, \hat{1}) &= \zeta^{-1}(\hat{0}, \hat{1}) \\ &= (\delta + (\zeta - \delta))^{-1}(\hat{0}, \hat{1}) \\ &= \delta(\hat{0}, \hat{1}) - (\zeta - \delta)(\hat{0}, \hat{1}) + (\zeta - \delta)^2(\hat{0}, \hat{1}) - \cdots . \end{aligned}$$

This expansion is easily justified since $(\zeta - \delta)^k(\hat{0}, \hat{1}) = 0$ if the longest chain of P has length less than k. By definition of the product in $\mathcal{I}(P)$ we have $(\zeta - \delta)^i(\hat{0}, \hat{1}) = c_i$, and the proof follows. □

Note. Let P be a finite poset with $\hat{0}$ and $\hat{1}$, and let $P' = P - \{\hat{0}, \hat{1}\}$. Define $\Delta(P')$ to be the set of chains of P', so $\Delta(P')$ is an abstract simplicial complex. The *reduced Euler characteristic* of a simplicial complex Δ is defined by

$$\tilde{\chi}(P) = -f_{-1} + f_0 - f_1 + \cdots ,$$

where f_i is the number of i-dimensional faces $F \in \Delta$ (or $\#F = i + 1$). Comparing with Lemma 4.4 shows that

$$\mu(\hat{0}, \hat{1}) = \tilde{\chi}(\Delta(P')).$$

Readers familiar with topology will know that $\tilde{\chi}(\Delta)$ has important topological significance related to the homology of Δ. It is thus natural to ask whether results

concerning Möbius functions can be generalized or refined topologically. Such results are part of the subject of "topological combinatorics," about which we will say a little more later.

Now let P be a finite graded poset with $\hat{0}$ and $\hat{1}$. Let

$$\mathcal{E}(P) = \{(x, y) : x \lessdot y \text{ in } P\},$$

the set of (directed) edges of the Hasse diagram of P.

Definition 4.11. An *E-labeling* of P is a map $\lambda : \mathcal{E}(P) \to \mathbb{P}$ such that if $x < y$ in P then there exists a unique saturated chain

$$C : x = x_0 \lessdot x_1 \lessdot x_1 \lessdot \cdots \lessdot x_k = y$$

satisfying

$$\lambda(x_0, x_1) \leq \lambda(x_1, x_2) \leq \cdots \leq \lambda(x_{k-1}, x_k).$$

We call C the *increasing chain* from x to y.

Figure 2 shows three examples of posets P with a labeling of their edges, i.e. a map $\lambda : \mathcal{E}(P) \to \mathbb{P}$. Figure 2(a) is the boolean algebra B_3 with the labeling $\lambda(S, S \cup \{i\}) = i$. (The one-element subsets $\{i\}$ are also labelled with a small i.) For any boolean algebra B_n, this labeling is the archetypal example of an E-labeling. The unique increasing chain from S to T is obtained by adjoining to S the elements of $T - S$ one at a time in increasing order. Figures 2(b) and (c) show two different E-labelings of the same poset P. These labelings have a number of different properties, e.g., the first has a chain whose edge labels are not all different, while every maximal chain label of Figure 2(c) is a permutation of $\{1, 2\}$.

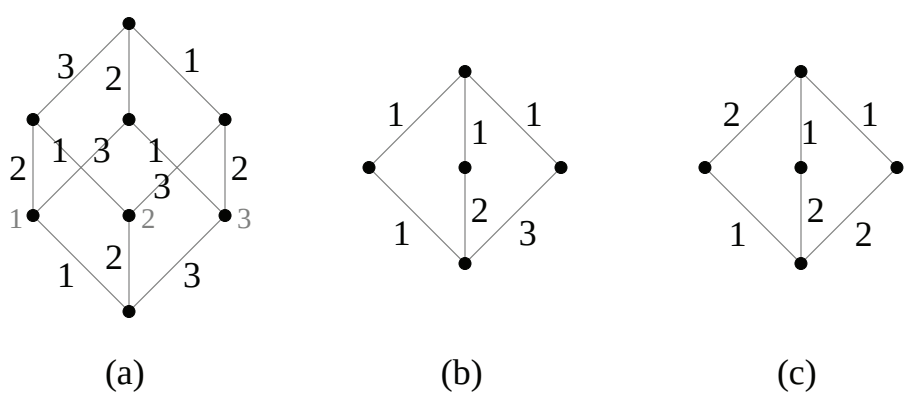


Figure 2. Three examples of edge-labelings

Theorem 4.11. *Let λ be an E-labeling of P, and let $x \leq y$ in P. Let μ denote the Möbius function of P. Then $(-1)^{\mathrm{rk}(x,y)}\mu(x, y)$ is equal to the number of strictly decreasing saturated chains from x to y, i.e.,*

$$(-1)^{\mathrm{rk}(x,y)}\mu(x, y) =$$

$$\#\{x = x_0 \lessdot x_1 \lessdot \cdots \lessdot x_k = y : \lambda(x_0, x_1) > \lambda(x_1, x_2) > \cdots > \lambda(x_{k-1}, x_k)\}.$$

Proof. Since λ restricted to $[x,y]$ (i.e., to $\mathcal{E}([x,y])$) is an E-labeling, we can assume $[x,y]=[\hat{0},\hat{1}]=P$. Let $S=\{a_1,a_2,\ldots,a_{j-1}\}\subseteq[n-1]$, with $a_1<a_2<\cdots<a_{j-1}$. Define $\alpha_P(S)$ to be the number of chains $\hat{0}<y_1<\cdots<y_{j-1}<\hat{1}$ in P such that $\mathrm{rk}(y_i)=a_i$ for $1\le i\le j-1$. The function α_P is called the *flag f-vector* of P.

Claim. $\alpha_P(S)$ is the number of maximal chains $\hat{0}=x_0\lessdot x_1\lessdot\cdots\lessdot x_n=\hat{1}$ such that

$$\lambda(x_{i-1},x_i)>\lambda(x_i,x_{i+1})\Rightarrow i\in S,\ 1\le i\le n. \tag{27}$$

To prove the claim, let $\hat{0}=y_0<y_1<\cdots<y_{j-1}<y_j=\hat{1}$ with $\mathrm{rk}(y_i)=a_i$ for $1\le i\le j-1$. By the definition of E-labeling, there exists a unique refinement

$$\hat{0}=y_0=x_0\lessdot x_1\lessdot\cdots\lessdot x_{a_1}=y_1\lessdot x_{a_1+1}\lessdot\cdots\lessdot x_{a_2}=y_2\lessdot\cdots\lessdot x_n=y_j=\hat{1}$$

satisfying

$$\begin{gathered}\lambda(x_0,x_1)\le\lambda(x_1,x_2)\le\cdots\le\lambda(x_{a_1-1},x_{a_1})\\ \lambda(x_{a_1},x_{a_1+1})\le\lambda(x_{a_1+1},x_{a_1+2})\le\cdots\le\lambda(x_{a_2-1},x_{a_2})\\ \cdots\end{gathered}$$

Thus if $\lambda(x_{i-1},x_i)>\lambda(x_i,x_{i+1})$, then $i\in S$, so (27) is satisfied. Conversely, given a maximal chain $\hat{0}=x_0\lessdot x_1\lessdot\cdots\lessdot x_n=\hat{1}$ satisfying the above conditions on λ, let $y_i=x_{a_i}$. Therefore we have a bijection between the chains counted by $\alpha_P(S)$ and the maximal chains satisfying (27), so the claim follows.

Now for $S\subseteq[n-1]$ define

$$\beta_P(S)=\sum_{T\subseteq S}(-1)^{\#(S-T)}\alpha_P(T). \tag{28}$$

The function β_P is called the *flag h-vector* of P. A simple Inclusion-Exclusion argument gives

$$\alpha_P(S)=\sum_{T\subseteq S}\beta_P(T), \tag{29}$$

for all $S\subseteq[n-1]$. It follows from the claim and equation (29) that $\beta_P(T)$ is equal to the number of maximal chains $\hat{0}=x_0\lessdot x_1\lessdot\cdots\lessdot x_n=\hat{1}$ such that $\lambda(x_i)>\lambda(x_{i+1})$ if and only if $i\in T$. In particular, $\beta_P([n-1])$ is equal to the number of strictly decreasing maximal chains $\hat{0}=x_0\lessdot x_1\lessdot\cdots\lessdot x_n=\hat{1}$ of P, i.e.,

$$\lambda(x_0,x_1)>\lambda(x_1,x_2)>\cdots>\lambda(x_{n-1},x_n).$$

Now by (28) we have

$$\begin{aligned}\beta_P([n-1]) &= \sum_{T\subseteq[n-1]}(-1)^{n-1-\#T}\alpha_P(T)\\ &= \sum_{k\ge1}\sum_{\hat{0}=y_0<y_1<\cdots<y_k=\hat{1}}(-1)^{n-k}\\ &= (-1)^n\sum_{k\ge1}(-1)^k c_k,\end{aligned}$$

where c_i is the number of chains $\hat{0}=y_0<y_1<\cdots<y_i=\hat{1}$ in P. The proof now follows from Philip Hall's theorem (Lemma 4.4). $\square$

We come to the main result of this section, a combinatorial interpretation of the coefficients of the characteristic polynomial $\chi_M(t)$ for any matroid M.

Theorem 4.12. *Let M be a matroid of rank n with a linear ordering $x_1 < x_2 < \cdots < x_m$ of its points (so the broken circuit complex $BC(M)$ is defined), and let $0 \le i \le n$. Then*

$$(-1)^i [t^{n-i}]\chi_M(t) = f_{i-1}(\mathrm{BC}(M)).$$

Proof. We may assume M is simple since the "simplification" $\widehat{M}$ has the same lattice of flats and same broken circuit complex as M (Exercise 1). The atoms x_i of $L(M)$ can then be identified with the points of M. Define a labeling $\tilde{\lambda} : \mathcal{E}(L(M)) \to \mathbb{P}$ as follows. Let $x \lessdot y$ in $L(M)$. Then set

$$\tilde{\lambda}(x, y) = \max\{i \,:\, x \vee x_i = y\}. \tag{30}$$

Note that $\tilde{\lambda}(x, y)$ is defined since $L(M)$ is atomic.

As an example, Figure 3 shows the lattice of flats of the matroid M of Figure 1 with the edge labeling (30).

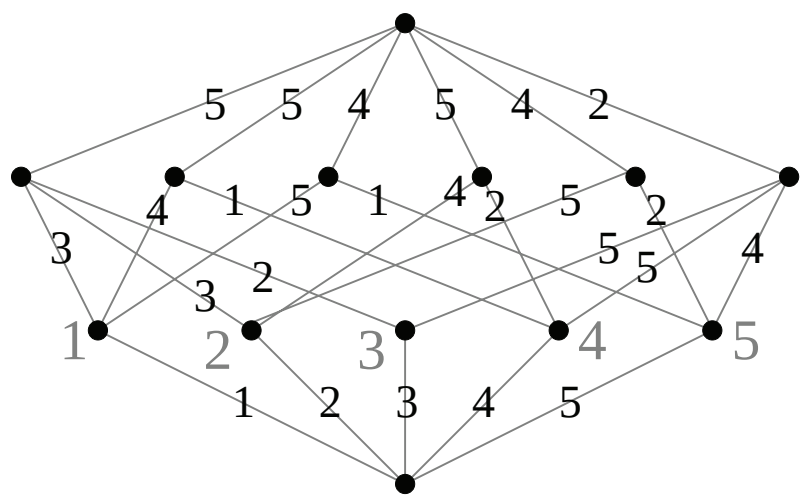


Figure 3. The edge labeling $\tilde{\lambda}$ of a geometric lattice $L(M)$

Claim 1. Define $\lambda : \mathcal{E}(L(M)) \to \mathbb{P}$ by

$$\lambda(x, y) = m + 1 - \tilde{\lambda}(x, y).$$

Then λ is an E-labeling.

To prove this claim, we need to show that for all $x < y$ in $L(M)$ there is a unique saturated chain $x = y_0 \lessdot y_1 \lessdot \cdots \lessdot y_k = y$ satisfying

$$\tilde{\lambda}(y_0, y_1) \ge \tilde{\lambda}(y_1, y_2) \ge \cdots \ge \tilde{\lambda}(y_{k-1}, y_k).$$

The proof is by induction on k. There is nothing to prove for $k = 1$. Let $k > 1$ and assume the assertion for $k - 1$. Let

$$j = \max\{i \,:\, x_i \le y,\ x_i \not\le x\}.$$

For *any* saturated chain $x = z_0 \lessdot z_1 \lessdot \cdots \lessdot z_k = y$, there is some i for which $x_j \not\le z_i$ and $x_j \le z_{i+1}$. Hence $\tilde{\lambda}(z_i, z_{i+1}) = j$. Thus if $\tilde{\lambda}(z_0, z_1) \ge \cdots \ge \tilde{\lambda}(z_{k-1}, z_k)$, then $\tilde{\lambda}(z_0, z_1) = j$. Moreover, there is a unique y_1 satisfying $x = x_0 \lessdot y_1 \le y$ and $\tilde{\lambda}(x_0, y_1) = j$, viz., $y_1 = x_0 \vee x_j$. (Note that $y_1 \gtrdot x_0$ by semimodularity.)

By the induction hypothesis there exists a unique saturated chain $y_1 \lessdot y_2 \lessdot \cdots \lessdot y_k = y$ satisfying $\tilde{\lambda}(y_1, y_2) \geq \cdots \geq \tilde{\lambda}(y_{k-1}, y_k)$. Since $\tilde{\lambda}(y_0, y_1) = j > \tilde{\lambda}(y_1, y_2)$, the proof of Claim 1 follows by induction.

Claim 2. The broken circuit complex $\mathrm{BC}(M)$ consists of all chain labels $\lambda(C)$, where C is a saturated increasing chain (with respect to $\tilde{\lambda}$) from $\hat{0}$ to some $x \in L(M)$. Moreover, all such $\lambda(C)$ are distinct.

To prove the distinctness of the labels $\lambda(C)$, suppose that C is given by $\hat{0} = y_0 \lessdot y_1 \lessdot \cdots \lessdot y_k$, with $\tilde{\lambda}(C) = (a_1, a_2, \ldots, a_k)$. Then $y_i = y_{i-1} \vee x_{a_i}$, so C is the only chain with its label.

Now let C and $\tilde{\lambda}(C)$ be as in the previous paragraph. We claim that the set $\{x_{a_1}, \ldots, x_{a_k}\}$ contains no broken circuit. (We don't even require that C is increasing for this part of the proof.) Write $z_i = x_{a_i}$, and suppose to the contrary that $B = \{z_{i_1}, \ldots, z_{i_j}\}$ is a broken circuit, with $1 \leq i_1 < \cdots < i_j \leq k$. Let $B \cup \{x_r\}$ be a circuit with $r > a_{i_t}$ for $1 \leq t \leq j$. Now for any circuit $\{u_1, \ldots, u_h\}$ and any $1 \leq i \leq h$ we have

$$u_1 \vee u_2 \vee \cdots \vee u_h = u_1 \vee \cdots \vee u_{i-1} \vee u_{i+1} \vee \cdots \vee u_h.$$

Thus

$$z_{i_1} \vee z_{i_2} \vee \cdots \vee z_{i_{j-1}} \vee x_r = \bigvee_{z \in B} z = z_{i_1} \vee z_{i_2} \vee \cdots \vee z_{i_j}.$$

Then $y_{i_j - 1} \vee x_r = y_{i_j}$, contradicting the maximality of the label a_{i_j}. Hence $\{x_{a_1}, \ldots, x_{a_k}\} \in \mathrm{BC}(M)$.

Conversely, suppose that $T := \{x_{a_1}, \ldots, x_{a_k}\}$ contains no broken circuit, with $a_1 < \cdots < a_k$. Let $y_i = x_{a_1} \vee \cdots \vee x_{a_i}$, and let C be the chain $\hat{0} := y_0 \lessdot y_1 \lessdot \cdots \lessdot y_k$. (Note that C is saturated by semimodularity.) We claim that $\tilde{\lambda}(C) = (a_1, \ldots, a_k)$. If not, then $y_{i-1} \vee x_j = y_i$ for some $j > a_i$. Thus

$$\mathrm{rk}(T) = \mathrm{rk}(T \cup \{x_j\}) = i.$$

Since T is independent, $T \cup \{x_j\}$ contains a circuit Q satisfying $x_j \in Q$, so T contains a broken circuit. This contradiction completes the proof of Claim 2.

To complete the proof of the theorem, note that we have shown that $f_{i-1}(\mathrm{BC}(M))$ is the number of chains $C : \hat{0} = y_0 \lessdot y_1 \lessdot \cdots \lessdot y_i$ such that $\tilde{\lambda}(C)$ is strictly increasing, or equivalently, $\lambda(C)$ is strictly decreasing. Since λ is an E-labeling, the proof follows from Theorem 4.11. $\square$

Corollary 4.6. *The broken circuit complex* $\mathrm{BC}(M)$ *is* pure*, i.e., every maximal face has the same dimension.*

The proof is left as an exercise (Exercise 21).

Note (for readers with some knowledge of topology). (a) Let M be a matroid on the linearly ordered set $u_1 < u_2 < \cdots < u_m$. Note that $F \in \mathrm{BC}(M)$ if and only if $F \cup \{u_m\} \in \mathrm{BC}(M)$. Define the *reduced broken circuit complex* $\mathrm{BC}_r(M)$ by

$$\mathrm{BC}_r(M) = \{F \in \mathrm{BC}(M) \,:\, u_m \notin F\}.$$

Thus

$$\mathrm{BC}(M) = \mathrm{BC}_r(M) * u_m,$$

the *join* of $\mathrm{BC}_r(M)$ and the vertex u_m. Equivalently, $\mathrm{BC}(M)$ is a *cone* over $\mathrm{BC}_r(M)$ with apex u_m. As a consequence, $\mathrm{BC}(M)$ is contractible and therefore has the homotopy type of a point. A more interesting problem is to determine the topological nature of $\mathrm{BC}_r(M)$. It can be shown that $\mathrm{BC}_r(M)$ has the homotopy type of a wedge

of $\beta(M)$ spheres of dimension $\mathrm{rank}(M)-2$, where $(-1)^{\mathrm{rank}(M)-1}\beta(M) = \chi'_M(1)$ (the derivative of $\chi_M(t)$ at $t=1$). See Exercise 22 for more information on $\beta(M)$.

As an example of the applicability of our results on matroids and geometric lattices to arrangements, we have the following purely combinatorial description of the number of regions of a real central arrangement.

Corollary 4.7. *Let $\mathcal{A}$ be a central arrangement in $\mathbb{R}^n$, and let M be the matroid defined by the normals to $H \in \mathcal{A}$, i.e., the independent sets of M are the linearly independent normals. Then with respect to any linear ordering of the points of M, $r(\mathcal{A})$ is the total number of subsets of M that don't contain a broken circuit.*

Proof. Immediate from Theorems 2.5 and 4.12. □

4.2. Modular Elements

We next discuss a situation in which the characteristic polynomial $\chi_M(t)$ factors in a nice way.

Definition 4.12. An element x of a geometric lattice L is *modular* if for all $y \in L$ we have

$$\mathrm{rk}(x) + \mathrm{rk}(y) = \mathrm{rk}(x \wedge y) + \mathrm{rk}(x \vee y). \tag{31}$$

Example 4.9. Let L be a geometric lattice.

(a) $\hat{0}$ and $\hat{1}$ are clearly modular (in any finite lattice).

(b) We claim that atoms a are modular.

Proof. Suppose that $a \leq y$. Then $a \wedge y = a$ and $a \vee y = y$, so equation (31) holds. (We don't need that a is an atom for this case.) Now suppose $a \not\leq y$. By semimodularity, $\mathrm{rk}(a \vee y) = 1 + \mathrm{rk}(y)$, while $\mathrm{rk}(a) = 1$ and $\mathrm{rk}(a \wedge y) = \mathrm{rk}(\hat{0}) = 0$, so again (31) holds. □

(c) Suppose that $\mathrm{rk}(L) = 3$. All elements of rank 0, 1, or 3 are modular by (a) and (b). Suppose that $\mathrm{rk}(x) = 2$. Then x is modular if and only if for all elements $y \neq x$ and $\mathrm{rk}(y) = 2$, we have that $\mathrm{rk}(x \wedge y) = 1$.

(d) Let $L = B_n$. If $x \in B_n$ then $\mathrm{rk}(x) = \#x$. Moreover, for any $x, y \in B_n$ we have $x \wedge y = x \cap y$ and $x \vee y = x \cup y$. Since for any finite sets x and y we have

$$\#x + \#y = \#(x \cap y) + \#(x \cup y),$$

it follows that *every* element of B_n is modular. In other words, B_n is a *modular lattice*.

(e) Let q be a prime power and $\mathbb{F}_q$ the finite field with q elements. Define $B_n(q)$ to be the lattice of subspaces, ordered by inclusion, of the vector space $\mathbb{F}_q^n$. Note that $B_n(q)$ is also isomorphic to the intersection lattice of the arrangement of *all* linear hyperplanes in the vector space $\mathbb{F}_n(q)$. Figure 4 shows the Hasse diagrams of $B_2(3)$ and $B_3(2)$.

Note that for $x, y \in B_n(q)$ we have $x \wedge y = x \cap y$ and $x \vee y = x + y$ (subspace sum). Clearly $B_n(q)$ is atomic: every vector space is the join (sum) of its one-dimensional subspaces. Moreover, $B_n(q)$ is graded of rank n, with rank function given by $\mathrm{rk}(x) = \dim(x)$. Since for any subspaces x and y we have

$$\dim(x) + \dim(y) = \dim(x \cap y) + \dim(x + y),$$

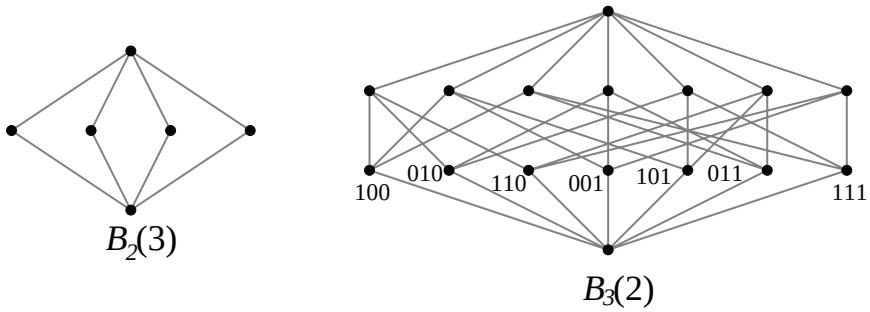


Figure 4. The lattices $B_2(3)$ and $B_3(2)$

it follows that L is a modular geometric lattice. Thus every $x \in L$ is modular.

NOTE. A *projective plane* R consists of a set (also denoted R) of points, and a collection of subsets of R, called lines, such that: (a) every two points lie on a unique line, (b) every two lines intersect in exactly one point, and (c) (non-degeneracy) there exist four points, no three of which are on a line. The *incidence lattice* $L(R)$ of R is the set of all points and lines of R, ordered by $p < L$ if $p \in L$, with $\hat{0}$ and $\hat{1}$ adjoined. It is an immediate consequence of the axioms that when R is finite, $L(R)$ is a modular geometric lattice of rank 3. It is an open (and probably intractable) problem to classify all finite projective planes. Now let P and Q be posets and define their *direct product* (or *cartesian product*) to be the set

$$P \times Q = \{(x, y) \,:\, x \in P,\ y \in Q\},$$

ordered componentwise, i.e., $(x, y) \leq (x', y')$ if $x \leq x'$ and $y \leq y'$. It is easy to see that if P and Q are geometric (respectively, atomic, semimodular, modular) lattices, then so is $P \times Q$ (Exercise 7). It is a consequence of the "fundamental theorem of projective geometry" that every finite modular geometric lattice is a direct product of boolean algebras B_n, subspace lattices $B_n(q)$ for $n \geq 3$, lattices of rank 2 with at least five elements (which may be regarded as $B_2(q)$ for any $q \geq 2$) and incidence lattices of finite projective planes.

(f) The following result characterizes the modular elements of Π_n, which is the lattice of partitions of $[n]$ or the intersection lattice of the braid arrangement $\mathcal{B}_n$.

Proposition 4.9. *A partition $\pi \in \Pi_n$ is a modular element of Π_n if and only if π has at most one nonsingleton block. Hence the number of modular elements of Π_n is $2^n - n$.*

Proof. If all blocks of π are singletons, then $\pi = \hat{0}$, which is modular by (a). Assume that π has the block A with $r > 1$ elements, and all other blocks are singletons. Hence the number $|\pi|$ of blocks of π is given by $n - r + 1$. For any $\sigma \in \Pi_n$, we have $\mathrm{rk}(\sigma) = n - |\sigma|$. Let $k = |\sigma|$ and

$$j = \#\{B \in \sigma \,:\, A \cap B \neq \emptyset\}.$$

Then $|\pi \wedge \sigma| = j + (n - r)$ and $|\pi \vee \sigma| = k - j + 1$. Hence $\mathrm{rk}(\pi) = r - 1$, $\mathrm{rk}(\sigma) = n - k$, $\mathrm{rk}(\pi \wedge \sigma) = r - j$, and $\mathrm{rk}(\pi \vee \sigma) = n - k + j - 1$, so π is modular.

Conversely, let $\pi = \{B_1, B_2, \ldots, B_k\}$ with $\#B_1 > 1$ and $\#B_2 > 1$. Let $a \in B_1$ and $b \in B_2$, and set

$$\sigma = \{(B_1 \cup b) - a, (B_2 \cup a) - b, B_3, \ldots, B_k\}.$$

Then

$$|\pi| = |\sigma| = k$$

$$\begin{aligned} \pi \wedge \sigma &= \{a, b, B_1 - a, B_2 - b, \ldots, B_3, \ldots, B_k\} &\Rightarrow\ & |\pi \wedge \sigma| = k + 2 \\ \pi \vee \sigma &= \{B_1 \cup B_2, B_3, \ldots, B_l\} &\Rightarrow\ & |\pi \vee \sigma| = k - 1. \end{aligned}$$

Hence $\mathrm{rk}(\pi) + \mathrm{rk}(\sigma) \neq \mathrm{rk}(\pi \wedge \sigma) + \mathrm{rk}(\pi \vee \sigma)$, so π is not modular. □

In a finite lattice L, a *complement* of $x \in L$ is an element $y \in L$ such that $x \wedge y = \hat{0}$ and $x \vee y = \hat{1}$. For instance, in the boolean algebra B_n every element has a unique complement. (See Exercise 3 for the converse.) The following proposition collects some useful properties of modular elements. The proof is left as an exercise (Exercises 4–5).

Proposition 4.10. *Let L be a geometric lattice of rank n.*

(a) *Let $x \in L$. The following four conditions are equivalent.*
  (i) *x is a modular element of L.*
  (ii) *If $x \wedge y = \hat{0}$, then $\mathrm{rk}(x) + \mathrm{rk}(y) = \mathrm{rk}(x \vee y)$.*
  (iii) *If x and y are complements, then $\mathrm{rk}(x) + \mathrm{rk}(y) = n$.*
  (iv) *All complements of x are incomparable.*

(b) *(transitivity of modularity) If x is a modular element of L and y is modular in the interval $[\hat{0}, x]$, then y is a modular element of L.*

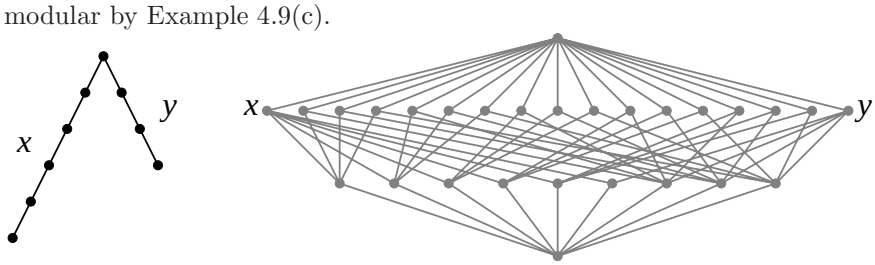
(c) *If x and y are modular elements of L, then $x \wedge y$ is also modular.*

The next result, known as the *modular element factorization theorem* [**28**], is our primary reason for defining modular elements — such an element induces a factorization of the characteristic polynomial.

Theorem 4.13. *Let z be a modular element of the geometric lattice L of rank n. Write $\chi_z(t) = \chi_{[\hat{0},z]}(t)$. Then*

$$\chi_L(t) = \chi_z(t) \left[\sum_{y\,:\,y \wedge z = \hat{0}} \mu_L(y) t^{n - \mathrm{rk}(y) - \mathrm{rk}(z)} \right]. \tag{32}$$

Example 4.10. Before proceeding to the proof of Theorem 4.13, let us consider an example. The illustration below is the affine diagram of a matroid M of rank 3, together with its lattice of flats. The two lines (flats of rank 2) labelled x and y are modular by Example 4.9(c).



Hence by equation (32) $\chi_M(t)$ is divisible by $\chi_x(t)$. Moreover, any atom a of the interval $[\hat{0}, x]$ is modular, so $\chi_x(t)$ is divisible by $\chi_a(t) = t - 1$. From this it is immediate (e.g., because the characteristic polynomial $\chi_G(t)$ of any geometric lattice G of rank n begins $x^n - ax^{n-1} + \cdots$, where a is the number of atoms of G) that $\chi_x(t) = (t-1)(t-5)$ and $\chi_M(t) = (t-1)(t-3)(t-5)$. On the other hand, since y is modular, $\chi_M(t)$ is divisible by $\chi_y(t)$, and we get as before $\chi_y(t) = (t-1)(t-3)$ and $\chi_M(t) = (t-1)(t-3)(t-5)$. Geometric lattices whose characteristic polynomial factors into linear factors in a similar way due to a maximal chain of modular elements are discussed further beginning with Definition 4.13.

Our proof of Theorem 4.13 will depend on the following lemma of Greene [**19**]. We give a somewhat simpler proof than Greene.

Lemma 4.5. *Let L be a finite lattice with Möbius function μ, and let $z \in L$. The following identity is valid in the Möbius algebra $A(L)$ of L:*

$$\sigma_{\hat{0}} := \sum_{x \in L} \mu(x)x = \left(\sum_{v \le z} \mu(v)v \right) \left(\sum_{y \wedge z = \hat{0}} \mu(y)y \right). \tag{33}$$

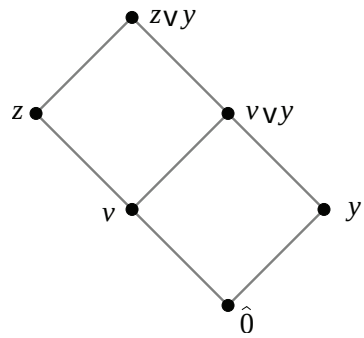
Proof. Let σ_s for $s \in L$ be given by (8). The right-hand side of equation (33) is then given by

$$\begin{aligned}
\sum_{\substack{v \le z \\ y \wedge z = \hat{0}}} \mu(v)\mu(y)(v \vee y) &= \sum_{\substack{v \le z \\ y \wedge z = \hat{0}}} \mu(v)\mu(y) \sum_{s \ge v \vee y} \sigma_s \\
&= \sum_s \sigma_s \sum_{\substack{v \le s, v \le z \\ y \le s, y \wedge z = \hat{0}}} \mu(v)\mu(y) \\
&= \sum_s \sigma_s \left(\underbrace{\sum_{v \le s \wedge z} \mu(v)}_{\delta_{\hat{0}, s \wedge z}} \right) \left(\sum_{\substack{y \le s \\ y \wedge z = \hat{0}}} \mu(y) \right) \\
&= \sum_{s \wedge z = \hat{0}} \sigma_s \left(\underbrace{\sum_{\substack{y \le s \\ y \wedge z = \hat{0} \text{ (redundant)}}} \mu(y)}_{\delta_{\hat{0}, s}} \right) \\
&= \sigma_{\hat{0}}.
\end{aligned}$$

□

Proof of Theorem 4.13. We are assuming that z is a modular element of the geometric lattice L.

Claim 1. Let $v \le z$ and $y \wedge z = \hat{0}$ (so $v \wedge y = \hat{0}$). Then $z \wedge (v \vee y) = v$ (as illustrated below).



Proof of Claim 1. Clearly $z \wedge (v \vee y) \geq v$, so it suffices to show that $\mathrm{rk}(z \wedge (v \vee y)) \leq \mathrm{rk}(v)$. Since z is modular we have

$$\begin{aligned}
\mathrm{rk}(z \wedge (v \vee y)) &= \mathrm{rk}(z) + \mathrm{rk}(v \vee y) - \mathrm{rk}(z \vee y) \\
&= \mathrm{rk}(z) + \mathrm{rk}(v \vee y) - (\mathrm{rk}(z) + \mathrm{rk}(y) - \underbrace{\mathrm{rk}(z \wedge y)}_{0}) \\
&= \mathrm{rk}(v \vee y) - \mathrm{rk}(y) \\
&\leq (\mathrm{rk}(v) + \mathrm{rk}(y) - \underbrace{\mathrm{rk}(v \wedge y)}_{0}) - \mathrm{rk}(y) \text{ by semimodularity} \\
&= \mathrm{rk}(v),
\end{aligned}$$

proving Claim 1.

Claim 2. With v and y as above, we have $\mathrm{rk}(v \vee y) = \mathrm{rk}(v) + \mathrm{rk}(y)$.

Proof of Claim 2. By the modularity of z we have

$$\mathrm{rk}(z \wedge (v \vee y)) + \mathrm{rk}(z \vee (v \vee y)) = \mathrm{rk}(z) + \mathrm{rk}(v \vee y).$$

By Claim 1 we have $\mathrm{rk}(z \wedge (v \vee y)) = \mathrm{rk}(v)$. Moreover, again by the modularity of z we have

$$\mathrm{rk}(z \vee (v \vee y)) = \mathrm{rk}(z \vee y) = \mathrm{rk}(z) + \mathrm{rk}(y) - \mathrm{rk}(z \wedge y) = \mathrm{rk}(z) + \mathrm{rk}(y).$$

It follows that $\mathrm{rk}(v) + \mathrm{rk}(y) = \mathrm{rk}(v \vee y)$, as claimed.

Now substitute $\mu(v)v \to \mu(v)t^{\mathrm{rk}(z)-\mathrm{rk}(v)}$ and $\mu(y)y \to \mu(y)t^{n-\mathrm{rk}(y)-\mathrm{rk}(z)}$ in the right-hand side of equation (33). Then by Claim 2 we have

$$vy \to t^{n-\mathrm{rk}(v)-\mathrm{rk}(y)} = t^{n-\mathrm{rk}(v \vee y)}.$$

Now $v \vee y$ is just vy in the Möbius algebra $A(L)$. Hence if we further substitute $\mu(x)x \to \mu(x)t^{n-\mathrm{rk}(x)}$ in the left-hand side of (33), then the product will be preserved. We thus obtain

$$\underbrace{\sum_{x \in L} \mu(x) t^{n-\mathrm{rk}(x)}}_{\chi_L(t)} = \left(\underbrace{\sum_{v \leq z} \mu(v) t^{\mathrm{rk}(z)-\mathrm{rk}(v)}}_{\chi_z(t)} \right) \left(\sum_{y \wedge z = \hat{0}} \mu(y) t^{n-\mathrm{rk}(y)-\mathrm{rk}(z)} \right),$$

as desired. □

Corollary 4.8. *Let L be a geometric lattice of rank n and a an atom of L. Then*

$$\chi_L(t) = (t-1) \sum_{y \wedge a = \hat{0}} \mu(y) t^{n-1-\mathrm{rk}(y)}.$$

Proof. The atom a is modular (Example 4.9(b)), and $\chi_a(t) = t - 1$. □

Corollary 4.8 provides a nice context for understanding the operation of coning defined in Chapter 1, in particular, Exercise 2.1. Recall that if $\mathcal{A}$ is an affine arrangement in K^n given by the equations

$$L_1(x) = a_1, \ \ldots, \ L_m(x) = a_m,$$

then the cone $x\mathcal{A}$ is the arrangement in $K^n \times K$ (where y denotes the last coordinate) with equations

$$L_1(x) = a_1 y, \ \ldots, L_m(x) = a_m y, \ y = 0.$$

Let H_0 denote the hyperplane $y = 0$. It is easy to see by elementary linear algebra that

$$L(\mathcal{A}) \cong L(c\mathcal{A}) - \{x \in L(\mathcal{A}) \,:\, x \geq H_0\} = L(\mathcal{A}) - L(\mathcal{A}^{H_0}).$$

Now H_0 is a modular element of $L(\mathcal{A})$ (since it's an atom), so Corollary 4.8 yields

$$\begin{aligned} \chi_{c\mathcal{A}}(t) = (t-1) \sum_{y \not\geq H_0} \mu(y) t^{(n+1)-1-\mathrm{rk}(y)} \\ = \quad (t-1)\chi_{\mathcal{A}}(t). \end{aligned}$$

There is a left inverse to the operation of coning. Let $\mathcal{A}$ be a nonempty linear arrangement in K^{n+1}. Let $H_0 \in \mathcal{A}$. Choose coordinates $(x_0, x_1, \ldots, x_n)$ in K^{n+1} so that $H_0 = \ker(x_0)$. Let $\mathcal{A}$ be defined by the equations

$$x_0 = 0, \ L_1(x_0, \ldots, x_n) = 0, \ \ldots, \ L_m(x_0, \ldots, x_n) = 0.$$

Define the *deconing* $c^{-1}\mathcal{A}$ (with respect to H_0) in K^n by the equations

$$L_1(1, x_1, \ldots, x_n) = 0, \ \ldots \ L_m(1, x_1, \ldots, x_n) = 0.$$

Clearly $c(c^{-1}\mathcal{A}) = \mathcal{A}$ and $L(c^{-1}\mathcal{A}) \cong L(\mathcal{A}) - \{x \in L(\mathcal{A}) \,:\, x \geq H_0\}$.

4.3. Supersolvable Lattices

For some geometric lattices L, there are "enough" modular elements to give a factorization of $\chi_L(t)$ into linear factors.

Definition 4.13. A geometric lattice L is *supersolvable* if there exists a *modular maximal chain*, i.e., a maximal chain $\hat{0} = x_0 \lessdot x_1 \lessdot \cdots \lessdot x_n = \hat{1}$ such that each x_i is modular. A central arrangement $\mathcal{A}$ is *supersolvable* if its intersection lattice $L_{\mathcal{A}}$ is supersolvable.

NOTE. Let $\hat{0} = x_0 \lessdot x_1 \lessdot \cdots \lessdot x_n = \hat{1}$ be a modular maximal chain of the geometric lattice L. Clearly then each x_{i-1} is a modular element of the interval $[\hat{0}, x_i]$. The converse follows from Proposition 4.10(b): if $\hat{0} = x_0 \lessdot x_1 \lessdot \cdots \lessdot x_n = \hat{1}$ is a maximal chain for which each x_{i-1} is modular in $[\hat{0}, x_i]$, then each x_i is modular in L.

NOTE. The term "supersolvable" comes from group theory. A finite group Γ is *supersolvable* if and only if its subgroup lattice contains a maximal chain all of whose elements are normal subgroups of Γ. Normal subgroups are "nice" analogues of modular elements; see [**29**, Example 2.5] for further details.

Corollary 4.9. *Let L be a supersolvable geometric lattice of rank n, with modular maximal chain $\hat{0} = x_0 \lessdot x_1 \lessdot \cdots \lessdot x_n = \hat{1}$. Let T denote the set of atoms of L, and set*

$$e_i = \#\{a \in T \,:\, a \leq x_i,\ a \not\leq x_{i-1}\}. \tag{34}$$

Then $\chi_L(t) = (t - e_1)(t - e_2) \cdots (t - e_n)$.

Proof. Since x_{n-1} is modular, we have

$$y \wedge x_{n-1} = \hat{0} \Leftrightarrow y \in T \text{ and } y \not\leq x_{n-1}, \text{ or } y = \hat{0}.$$

By Theorem 4.13 we therefore have

$$\chi_L(t) = \chi_{x_{n-1}}(t) \left[\sum_{\substack{a \in T \\ a \not\leq x_{n-1}}} \mu(a) t^{n - \mathrm{rk}(a) - \mathrm{rk}(x_{n-1})} + \mu(\hat{0}) t^{n - \mathrm{rk}(\hat{0}) - \mathrm{rk}(x_{n-1})} \right].$$

Since $\mu(a) = -1$, $\mu(\hat{0}) = 1$, $\mathrm{rk}(a) = 1$, $\mathrm{rk}(\hat{0}) = 0$, and $\mathrm{rk}(x_{n-1}) = n - 1$, the expression in brackets is just $t - e_n$. Now continue this with L replaced by $[\hat{0}, x_{n-1}]$ (or use induction on n). □

NOTE. The positive integers $e_1, \ldots, e_n$ of Corollary 4.9 are called the *exponents* of L.

Example 4.11. (a) Let $L = B_n$, the boolean algebra of rank n. By Example 4.9(d) every element of B_n is modular. Hence B_n is supersolvable. Clearly each $e_i = 1$, so $\chi_{B_n}(t) = (t - 1)^n$.

(b) Let $L = B_n(q)$, the lattice of subspaces of $\mathbb{F}_n^q$. By Example 4.9(e) every element of $B_n(q)$ is modular, so $B_n(q)$ is supersolvable. If $\left[{k \atop j}\right]$ denotes the number of j-dimensional subspaces of a k-dimensional vector space over $\mathbb{F}_q$, then

$$\begin{aligned} e_i &= \left[{i \atop 1}\right] - \left[{i-1 \atop 1}\right] \\ &= \frac{q^i - 1}{q - 1} - \frac{q^{i-1} - 1}{q - 1} \\ &= q^{i-1}. \end{aligned}$$

Hence

$$\chi_{B_n(q)}(t) = (t - 1)(t - q)(t - q^2) \cdots (t - q^{n-1}).$$

In particular, setting $t = 0$ gives

$$\mu_{B_n(q)}(\hat{1}) = (-1)^n q^{\binom{n}{2}}.$$

NOTE. The expression $\left[{k \atop j}\right]$ is called a *q-binomial coefficient*. It is a polynomial in q with many interesting properties. For the most basic properties, see e.g. [**31**, pp. 27–30].

(c) Let $L = \Pi_n$, the lattice of partitions of the set $[n]$ (a geometric lattice of rank $n - 1$). By Proposition 4.9, a maximal chain of Π_n is modular if and only if it has the form $\hat{0} = \pi_0 \lessdot \pi_1 \lessdot \cdots \lessdot \pi_{n-1} = \hat{1}$, where π_i for $i > 0$ has exactly one nonsingleton block B_i (necessarily with $i + 1$ elements), with $B_1 \subset B_2 \cdots \subset B_{n-1} = [n]$. In particular, Π_n is supersolvable and has exactly $n!/2$ modular chains for $n > 1$. The atoms covered by π_i are the

partitions with one nonsingleton block $\{j,k\} \subseteq B_i$. Hence π_i lies above exactly $\binom{i+1}{2}$ atoms, so

$$e_i = \binom{i+1}{2} - \binom{i}{2} = i.$$

It follows that $\chi_{\Pi_n}(t) = (t-1)(t-2)\cdots(t-n+1)$ and $\mu_{\Pi_n}(\hat{1}) = (-1)^{n-1}(n-1)!$. Compare Corollary 2.2. The polynomials $\chi_{\mathcal{B}_n}(t)$ and $\chi_{\Pi_n}(t)$ differ by a factor of t because $\mathcal{B}_n(t)$ is an arrangement in K^n of rank $n-1$. In general, if $\mathcal{A}$ is an arrangement and $\mathrm{ess}(\mathcal{A})$ its essentialization, then

$$t^{\mathrm{rk}(\mathrm{ess}(\mathcal{A}))}\chi_{\mathcal{A}}(t) = t^{\mathrm{rk}(\mathcal{A})}\chi_{\mathrm{ess}(\mathcal{A})}(t). \tag{35}$$

(See Lecture 1, Exercise 2.)

NOTE. It is natural to ask whether there is a more general class of geometric lattices L than the supersolvable ones for which $\chi_L(t)$ factors into linear factors (over $\mathbb{Z}$). There is a profound such generalization due to Terao [**35**] when L is an intersection poset of a linear arrangement $\mathcal{A}$ in K^n. Write $K[x] = K[x_1,\ldots,x_n]$ and define

$$\mathcal{T}(\mathcal{A}) = \{(p_1,\ldots,p_n) \in K[x]^n \,:\, p_i(H) \subseteq H \text{ for all } H \in \mathcal{A}\}.$$

Here we are regarding $(p_1,\ldots,p_n) : K^n \to K^n$, viz., if $(a_1,\ldots,a_n) \in K^n$, then

$$(p_1,\ldots,p_n)(a_1,\ldots,a_n) = (p_1(a_1,\ldots,a_n),\ldots,p_n(a_1,\ldots,a_n)).$$

The $K[x]$-module structure $K[x] \times \mathcal{T}(\mathcal{A}) \to \mathcal{T}(\mathcal{A})$ is given explicitly by

$$q \cdot (p_1,\ldots,p_n) = (qp_1,\ldots,qp_n).$$

Note, for instance, that we always have $(x_1,\ldots,x_n) \in \mathcal{T}(\mathcal{A})$. Since $\mathcal{A}$ is a linear arrangement, $\mathcal{T}(\mathcal{A})$ is indeed a $K[x]$-module. (We have given the most intuitive definition of the module $\mathcal{T}(\mathcal{A})$, though it isn't the most useful definition for proofs.) It is easy to see that $\mathcal{T}(\mathcal{A})$ has rank n as a $K[x]$-module, i.e., $\mathcal{T}(\mathcal{A})$ contains n, but not $n+1$, elements that are linearly independent over $K[x]$. We say that $\mathcal{A}$ is a *free* arrangement if $\mathcal{T}(\mathcal{A})$ is a free $K[x]$-module, i.e., there exist $Q_1,\ldots,Q_n \in \mathcal{T}(\mathcal{A})$ such that every element $Q \in \mathcal{T}(\mathcal{A})$ can be uniquely written in the form $Q = q_1Q_1 + \cdots + q_nQ_n$, where $q_i \in K[x]$. It is easy to see that if $\mathcal{T}(\mathcal{A})$ is free, then the basis $\{Q_1,\ldots,Q_n\}$ can be chosen to be *homogeneous*, i.e., all coordinates of each Q_i are homogeneous polynomials of the same degree d_i. We then write $d_i = \deg Q_i$. It can be shown that supersolvable arrangements are free, but there are also nonsupersolvable free arrangements. The property of freeness seems quite subtle; indeed, it is unknown whether freeness is a matroidal property, i.e., depends only on the intersection lattice $L_{\mathcal{A}}$ (regarding the ground field K as fixed). The remarkable "factorization theorem" of Terao is the following.

Theorem 4.14. *Suppose that $\mathcal{T}(\mathcal{A})$ is free with homogeneous basis $Q_1,\ldots,Q_n$. If $\deg Q_i = d_i$ then*

$$\chi_{\mathcal{A}}(t) = (t-d_1)(t-d_2)\cdots(t-d_n).$$

We will not prove Theorem 4.14 here. A good reference for this subject is [**24**, Ch. 4].

Returning to supersolvability, we can try to characterize the supersolvable property for various classes of geometric lattices. Let us consider the case of the bond

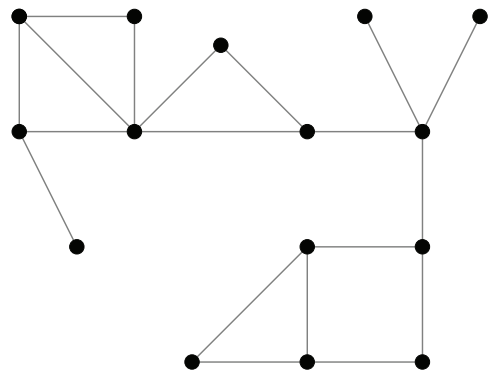


Figure 5. A graph with eight blocks

lattice L_G of the graph G. A graph H with at least one edge is *doubly connected* if it is connected and remains connected upon the removal of any vertex (and all incident edges). A maximal doubly connected subgraph of a graph G is called a *block* of G. For instance, if G is a forest then its blocks are its edges. Two different blocks of G intersect in at most one vertex. Figure 5 shows a graph with eight blocks, five of which consist of a single edge. The following proposition is straightforward to prove (Exercise 16).

Proposition 4.11. *Let G be a graph with blocks $G_1, \ldots, G_k$. Then*

$$L_G \cong L_{G_1} \times \cdots \times L_{G_k}.$$

It is also easy to see that if L_1 and L_2 are geometric lattices, then L_1 and L_2 are supersolvable if and only if $L_1 \times L_2$ is supersolvable (Exercise 18). Hence in characterizing *supersolvable graphs* G (i.e., graphs whose bond lattice L_G is supersolvable) we may assume that G is doubly connected. Note that for any connected (and hence a fortiori doubly connected) graph G, any coatom π of L_G has exactly two blocks.

Proposition 4.12. *Let G be a doubly connected graph, and let $\pi = \{A, B\}$ be a coatom of the bond lattice L_G, where $\#A \leq \#B$. Then π is a modular element of L_G if and only if $\#A = 1$, say $A = \{v\}$, and the neighborhood $N(v)$ (the set of vertices adjacent to v) forms a clique (i.e., any two distinct vertices of $N(v)$ are adjacent).*

Proof. The proof parallels that of Proposition 4.9, which is a special case. Suppose that $\#A > 1$. Since G is doubly connected, there exist $u, v \in A$ and $u', v' \in B$ such that $u \neq v$, $u' \neq v'$, $uu' \in E(G)$, and $vv' \in E(G)$. Set $\sigma = \{(A \cup u') - v, (B \cup v) - u'\}$. If G has n vertices then $\mathrm{rk}(\pi) = \mathrm{rk}(\sigma) = n-2$, $\mathrm{rk}(\pi \vee \sigma) = n-1$, and $\mathrm{rk}(\pi \wedge \sigma) = n-4$. Hence π is not modular.

Assume then that $A = \{v\}$. Suppose that $av, bv \in E(G)$ but $ab \notin E(G)$. We need to show that π is not modular. Let $\sigma = \{A - \{a, b\}, \{a, b, v\}\}$. Then

$$\sigma \vee \pi = \hat{1}, \quad \sigma \wedge \pi = \{A - \{a, b\}, a, b, v\}$$

$$\mathrm{rk}(\sigma) = \mathrm{rk}(\pi) = n - 2, \quad \mathrm{rk}(\sigma \vee \pi) = n - 1, \quad \mathrm{rk}(\sigma \wedge \pi) = n - 4.$$

Hence π is not modular.

Conversely, let $\pi = \{A, v\}$. Assume that if $av, bv \in E(G)$ then $ab \in E(G)$. It is then straightforward to show (Exercise 8) that π is modular, completing the proof. $\square$

As an immediate consequence of Propositions 4.10(b) and 4.12 we obtain a characterization of supersolvable graphs.

Corollary 4.10. *A graph G is supersolvable if and only if there exists an ordering $v_1, v_2, \ldots, v_n$ of its vertices such that if $i < k$, $j < k$, $v_i v_k \in E(G)$ and $v_j v_k \in E(G)$, then $v_i v_j \in E(G)$. Equivalently, in the restriction of G to the vertices $v_1, v_2, \ldots, v_i$, the neighborhood of v_i is a clique.*

Note. Supersolvable graphs G had appeared earlier in the literature under the names *chordal*, *rigid circuit*, or *triangulated* graphs. One of their many characterizations is that any circuit of length at least four contains a chord. Equivalently, no induced subgraph of G is a k-cycle for $k \geq 4$.

Exercises

(1) [2–] Let M be a matroid on a linearly ordered set. Show that $\mathrm{BC}(M) = \mathrm{BC}(\widehat{M})$, where $\widehat{M}$ is defined by equation (23).

(2) [2+] Let M be a matroid of rank at least one. Show that the coefficients of the polynomial $\chi_M(t)/(t-1)$ alternate in sign.

(3) (a) [2+] Let L be finite lattice for which every element has a unique complement. Show that L is isomorphic to a boolean algebra B_n.

(b) [3] A lattice L is *distributive* if

$$\begin{aligned} x \vee (y \wedge z) &= (x \vee y) \wedge (x \vee z) \\ x \wedge (y \vee z) &= (x \wedge y) \vee (x \wedge z) \end{aligned}$$

for all $x, y, z \in L$. Let L be an infinite lattice with $\hat{0}$ and $\hat{1}$. If every element of L has a unique complement, then is L a distributive lattice?

(4) [3–] Let x be an element of a geometric lattice L. Show that the following four conditions are equivalent.

(i) x is a modular element of L.

(ii) If $x \wedge y = \hat{0}$, then

$$\mathrm{rk}(x) + \mathrm{rk}(y) = \mathrm{rk}(x \vee y).$$

(iii) If x and y are complements, then $\mathrm{rk}(x) + \mathrm{rk}(y) = n$.

(iv) All complements of x are incomparable.

(5) [2+] Let x, y be modular elements of a geometric lattice L. Show that $x \wedge y$ is also modular.

(6) [2] Let L be a geometric lattice. Prove or disprove: if x is modular in L and y is modular in the interval $[x, \hat{1}]$, then y is modular in L.

(7) [2–] Let L and L' be finite lattices. Show that if both L and L' are geometric (respectively, atomic, semimodular, modular) lattices, then so is $L \times L'$.

(8) [2] Let G be a (loopless) connected graph and $v \in V(G)$. Let $A = V(G) - v$ and $\pi = \{A, v\} \in L_G$. Suppose that whenever $av, bv \in E(G)$ we have $ab \in E(G)$. Show that π is a modular element of L_G.

(9) [2+] Generalize the previous exercise as follows. Let G be a doubly-connected graph with lattice of contractions L_G. Let $\pi \in L_G$. Show that the following two conditions are equivalent.

(a) π is a modular element of L_G.
(b) π satisfies the following two properties:
  (i) At most one block B of π contains more than one vertex of G.
  (ii) Let H be the subgraph induced by the block B of (i). Let K be any connected component of the subgraph induced by $G - B$, and let H_1 be the graph induced by the set of vertices in H that are connected to some vertex in K. Then H_1 is a clique (complete subgraph) of G.

(10) [2+] Let L be a geometric lattice of rank n, and fix $x \in L$. Show that

$$\chi_L(t) = \sum_{\substack{y \in L \\ x \wedge y = \hat{0}}} \mu(y) \chi_{L_y}(t) t^{n - \mathrm{rk}(x \vee y)},$$

where L_y is the image of the interval $[\hat{0}, x]$ under the map $z \mapsto z \vee y$.

(11) [2+] Let $\mathcal{I}(M)$ be the set of independent sets of a matroid M. Find another matroid N and a labeling of its points for which $\mathcal{I}(M) = \mathrm{BC}_r(N)$, the reduced broken circuit complex of N.

(12) (a) [2+] If Δ and Γ are simplicial complexes on disjoint sets A and B, respectively, then define the *join* $\Delta * \Gamma$ to be the simplicial complex on the set $A \cup B$ with faces $F \cup G$, where $F \in \Delta$ and $G \in \Gamma$. (E.g., if Γ consists of a single point then $\Delta * \Gamma$ is the *cone* over Δ. If Γ consists of two disjoint points, then $\Delta * \Gamma$ is the *suspension* of Δ.) We say that Δ and Γ are *join-factors* of $\Delta * \Gamma$. Now let M be a matroid and $S \subset M$ a modular flat, i.e., S is a modular element of L_M. Order the points of M such that if $p \in S$ and $q \notin S$, then $p < q$. Show that $\mathrm{BC}(S)$ is a join-factor of $\mathrm{BC}(M)$. Deduce that $\chi_M(t)$ is divisible by $\chi_S(t)$.
(b) [2+] Conversely, let M be a matroid and $S \subset M$. Label the points of M so that if $p \in S$ and $q \notin S$, then $p < q$. Suppose that $\mathrm{BC}(S)$ is a join-factor of $\mathrm{BC}(M)$. Show that S is modular.

(13) [2] Do Exercise 3.7, this time using Theorem 4.12 (the Broken Circuit Theorem).

(14) [1] Show that all geometric lattices of rank two are supersolvable.

(15) [2] Give an example of two nonisomorphic supersolvable geometric lattices of rank 3 with the same characteristic polynomials.

(16) [2] Prove Proposition 4.11: if G is a graph with blocks $G_1, \ldots, G_k$, then $L_G \cong L_{G_1} \times \cdots \times L_{G_k}$.

(17) [2+] Give an example of a nonsupersolvable geometric lattice of rank three whose characteristic polynomial has only integer zeros.

(18) [2] Let L_1 and L_2 be geometric lattices. Show that L_1 and L_2 are supersolvable if and only if $L_1 \times L_2$ is supersolvable.

(19) [3–] Let L be a supersolvable geometric lattice. Show that every interval of L is also supersolvable.

(20) [2] (a) Find the number of maximal chains of the partition lattice Π_n.
(b) Find the number of modular maximal chains of Π_n.

(21) [2+] Show that the broken circuit complex of a matroid is pure (Corollary 4.6).

(22) Let M be a matroid with a linear ordering of its points. The *internal activity* of a basis B is the number of points $p \in B$ such that $p < q$ for all points $q \neq p$ not in the closure $\overline{B - p}$ of $B - p$. The *external activity* of B is the number of points $p' \in M - B$ such that $p' < q'$ for all $q' \neq p'$ contained in the unique circuit that

is a subset of $B \cup \{p'\}$. Define the *Crapo beta invariant* of M by

$$\beta(M) = (-1)^{\mathrm{rk}(M)-1}\chi'_M(1),$$

where $'$ denotes differentiation.

(a) [1+] Show that $1-\chi'_M(1) = \psi(\mathrm{BC}_r)$, the Euler characteristic of the reduced broken circuit complex of M.

(b) [3–] Show that $\beta(M)$ is equal to the number of bases of M with internal activity 0 and external activity 0.

(c) [2] Let $\mathcal{A}$ be a real central arrangement with associated matroid $M_{\mathcal{A}}$. Suppose that $\mathcal{A} = c\mathcal{A}'$ for some arrangement $\mathcal{A}'$, where $c\mathcal{A}'$ denotes the cone over $\mathcal{A}'$. Show that $\beta(M_{\mathcal{A}}) = b(\mathcal{A}')$.

(d) [2+] With $\mathcal{A}$ as in (c), let H' be a (proper) translate of some hyperplane $H \in \mathcal{A}$. Show that $\beta(M_{\mathcal{A}}) = b(\mathcal{A} \cup \{H'\})$.

LECTURE 5
Finite Fields

5.1. The Finite Field Method

In this lecture we will describe a method based on finite fields for computing the characteristic polynomial of an arrangement defined over $\mathbb{Q}$. We will then discuss several interesting examples. The main result (Theorem 5.15) is implicit in the work of Crapo and Rota [**13**, §17]. It was first developed into a systematic tool for computing characteristic polynomials by Athanasiadis [**1**][**2**], after a closely related but not as general technique was presented by Blass and Sagan [**9**].

Suppose that the arrangement $\mathcal{A}$ is defined over $\mathbb{Q}$. By multiplying each hyperplane equation by a suitable integer, we may assume $\mathcal{A}$ is defined over $\mathbb{Z}$. In that case we can take coefficients modulo a prime p and get an arrangement $\mathcal{A}_q$ defined over the finite field $\mathbb{F}_q$, where $q = p^r$. We say that $\mathcal{A}$ has *good reduction* mod p (or over $\mathbb{F}_q$) if $L(\mathcal{A}) \cong L(\mathcal{A}_q)$.

For instance, let $\mathcal{A}$ be the affine arrangement in $\mathbb{Q}^1 = \mathbb{Q}$ consisting of the points 0 and 10. Then $L(\mathcal{A})$ contains three elements, viz., $\mathbb{Q}$, $\{0\}$, and $\{10\}$. If $p \neq 2, 5$ then 0 and 10 remain distinct, so $\mathcal{A}$ has good reduction. On the other hand, if $p = 2$ or $p = 5$ then $0 = 10$ in $\mathbb{F}_p$, so $L(\mathcal{A}_p)$ contains just two elements. Hence $\mathcal{A}$ has bad reduction when $p = 2, 5$.

Proposition 5.13. *Let $\mathcal{A}$ be an arrangement defined over $\mathbb{Z}$. Then $\mathcal{A}$ has good reduction for all but finitely many primes p.*

Proof. Let $H_1, \ldots, H_j$ be affine hyperplanes, where H_i is given by the equation $v_i \cdot x = a_i$ ($v_i, a_i \in \mathbb{Z}^n$). By linear algebra, we have $H_1 \cap \cdots \cap H_j \neq \emptyset$ if and only if

$$\operatorname{rank} \begin{bmatrix} v_1 & a_1 \\ \vdots & \vdots \\ v_j & a_j \end{bmatrix} = \operatorname{rank} \begin{bmatrix} v_1 \\ \vdots \\ v_j \end{bmatrix}. \tag{36}$$

Moreover, if (36) holds then

$$\dim(H_1 \cap \cdots \cap H_j) = n - \operatorname{rank} \begin{bmatrix} v_1 \\ \vdots \\ v_j \end{bmatrix}.$$

Now for any $r \times s$ matrix A, we have $\operatorname{rank}(A) \geq t$ if and only if some $t \times t$ submatrix B satisfies $\det(B) \neq 0$. It follows that $L(\mathcal{A}) \ncong L(\mathcal{A}_p)$ if and only if at least one member S of a certain finite collection $\mathcal{S}$ of subsets of integer matrices B satisfies the following condition:

$$(\forall B \in S)\ \det(B) \neq 0 \text{ but } \det(B) \equiv 0 \pmod{p}.$$

This can only happen for finitely many p, viz., for certain B we must have $p | \det(B)$, so $L(\mathcal{A}) \cong L(\mathcal{A}_p)$ for p sufficiently large. □

The main result of this section is the following. Like many fundamental results in combinatorics, the proof is easy but the applicability very broad.

Theorem 5.15. *Let $\mathcal{A}$ be an arrangement in $\mathbb{Q}^n$, and suppose that $L(\mathcal{A}) \cong L(\mathcal{A}_q)$ for some prime power q. Then*

$$\begin{aligned} \chi_{\mathcal{A}}(q) &= \#\left(\mathbb{F}_q^n - \bigcup_{H \in \mathcal{A}_q} H\right) \\ &= q^n - \# \bigcup_{H \in \mathcal{A}_q} H. \end{aligned}$$

Proof. Let $x \in L(\mathcal{A}_q)$ so $\#x = q^{\dim(x)}$. Here $\dim(x)$ can be computed either over $\mathbb{Q}$ or F_q. Define two functions $f, g : L(\mathcal{A}_q) \to \mathbb{Z}$ by

$$\begin{aligned} f(x) &= \#x \\ g(x) &= \#\left(x - \bigcup_{y > x} y\right). \end{aligned}$$

In particular,

$$g(\hat{0}) = g(\mathbb{F}_q^n) = \#\left(\mathbb{F}_q^n - \bigcup_{H \in \mathcal{A}_q} H\right).$$

Clearly

$$f(x) = \sum_{y \geq x} g(y).$$

Let μ denote the Möbius function of $L(\mathcal{A}) \cong L(\mathcal{A}_q)$. By Möbius inversion (Theorem 1.1),

$$\begin{aligned} g(x) &= \sum_{y \geq x} \mu(x, y) f(y) \\ &= \sum_{y \geq x} \mu(x, y) q^{\dim(y)}. \end{aligned}$$

Put $x = \hat{0}$ to get

$$g(\hat{0}) = \sum_{y} \mu(y) q^{\dim(y)} = \chi_{\mathcal{A}}(q).$$

□

For the remainder of this lecture, we will be concerned with applications of Theorem 5.15 and further interesting examples of arrangements.

Example 5.12. Let G be a graph with vertices $1, 2, \ldots, n$, so

$$Q_{\mathcal{A}_G}(x) = \prod_{ij \in E(G)} (x_i - x_j).$$

Then by Theorem 5.15,

$$\begin{aligned} \chi_{\mathcal{A}_G}(q) &= q^n - \#\{(\alpha_1, \ldots, \alpha_n) \in \mathbb{F}_1^n \,:\, \alpha_i = \alpha_j \text{ for some } ij \in E(G)\} \\ &= \#\{(\beta_1, \ldots, \beta_n) \in \mathbb{F}_q^n \,:\, \beta_i \neq \beta_j \ \forall\, ij \in E(G)\} \\ &= \chi_G(q), \end{aligned}$$

in agreement with Theorem 2.7. Note that this equality holds for *all* prime powers q, not just for p^m with $p \gg 0$. This is because the matrix with rows $e_i - e_j$, where $ij \in E(G)$ and e_i is the ith unit coordinate vector in $\mathbb{Q}^n$, is *totally unimodular*, i.e., every minor (determinant of a square submatrix) is $0, \pm 1$. Hence the nonvanishing of a minor is independent of the ambient field.

A very interesting class of arrangements, including the braid arrangement, is associated with root systems, or more generally, finite reflection groups. We will simply mention some basic results here without proof. A *root system* is a finite set R of nonzero vectors in $\mathbb{R}^n$ satisfying certain properties that we will not give here. (References include [**6**][**10**][**20**].) The *Coxeter arrangement* $\mathcal{A}(R)$ consists of the hyperplanes $\alpha \cdot x = 0$, where $\alpha \in R$. There are four infinite (irreducible) classes of root systems (all in $\mathbb{R}^n$):

$$\begin{aligned} A_{n-1} &= \{e_i - e_j \,:\, 1 \leq i < j \leq n\} = \mathcal{B}_n \\ D_n &= \{e_i - e_j, e_i + e_j \,:\, 1 \leq i < j \leq n\} \\ B_n &= D_n \cup \{e_i \,:\, 1 \leq i \leq n\} \\ C_n &= D_n \cup \{2e_i \,:\, 1 \leq i \leq n\}. \end{aligned}$$

We should really regard A_{n-1} as being a subset of the space

$$\{(\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n \,:\, \sum \alpha_i = 0\} \cong \mathbb{R}^{n-1}.$$

We thus obtain the following Coxeter arrangements. In all cases $1 \leq i < j \leq n$ and $1 \leq k \leq n$.

$$\begin{aligned} \mathcal{A}(A_{n-1}) = \mathcal{B}_n &: \ x_i - x_j = 0 \\ \mathcal{A}(B_n) = \mathcal{A}(C_n) &: \ x_i - x_j = 0,\ x_i + x_j = 0,\ x_k = 0 \\ \mathcal{A}(D_n) &: \ x_i - x_j = 0,\ x_i + x_j = 0. \end{aligned}$$

See Figure 1 for the arrangements $\mathcal{A}(B_2)$ and $\mathcal{A}(D_2)$.

Let us compute the characteristic polynomial $\chi_{\mathcal{A}(B_n)}(q)$. For $p \gg 0$ (actually $p > 2$) and $q = p^m$ we have

$$\chi_{\mathcal{A}(B_n)}(q) = \#\{(\alpha_1, \ldots, \alpha_n) \in \mathbb{F}_q^n \,:\, \alpha_i \neq \pm\alpha_j \ (i \neq j),\ \alpha_i \neq 0 \ (1 \leq i \leq n)\}.$$

Choose $\alpha_1 \in \mathbb{F}_q^* = \mathbb{F}_q - \{0\}$ in $q - 1$ ways. Then choose $\alpha_2 \in \mathbb{F}_q^* - \{\alpha_1, -\alpha_1\}$ in $q - 3$ ways, then α_3 in $q - 5$ ways, etc., to obtain:

$$\chi_{\mathcal{A}(B_n)}(t) = (t-1)(t-3)\cdots(t-(2n-1)).$$

In particular,

$$r(\mathcal{A}(B_n)) = (-1)^n \chi_{\mathcal{A}(B_n)}(-1) = 2 \cdot 4 \cdot 6 \cdots (2n) = 2^n n!.$$

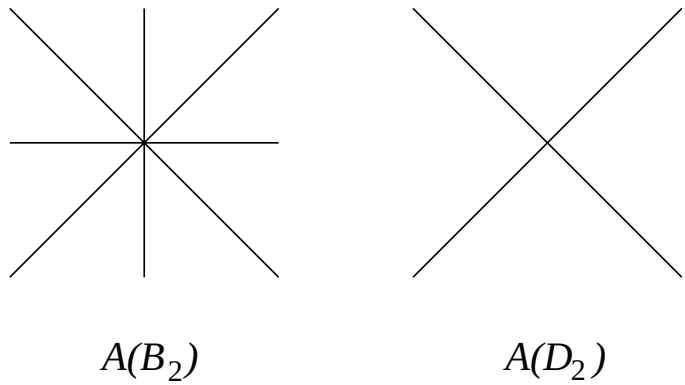


Figure 1. The arrangements $\mathcal{A}(B_2)$ and $\mathcal{A}(D_2)$

By a similar but slightly more complicated argument we get (Exercise 1)

$$\chi_{\mathcal{A}(D_n)}(t) = (t-1)(t-3)\cdots(t-(2n-3))\cdot(t-n+1). \tag{37}$$

NOTE. Coxeter arrangements are always free in the sense of Theorem 4.14 (a result of Terao [**34**]), but need not be supersolvable. In fact, $\mathcal{A}(A_n)$ and $\mathcal{A}(B_n)$ are supersolvable, but $\mathcal{A}(D_n)$ is not supersolvable for $n \geq 4$ [**4**, Thm. 5.1].

5.2. The Shi Arrangement

We next consider a modification (or deformation) of the braid arrangement called the *Shi arrangement* [**27**, §7] and denoted $\mathcal{S}_n$. It consists of the hyperplanes

$$x_i - x_j = 0, 1, \quad 1 \leq i < j \leq n.$$

Thus $\mathcal{S}_n$ has $n(n-1)$ hyperplanes and $\operatorname{rank}(\mathcal{S}_n) = n-1$. Figure 2 shows the Shi arrangement $\mathcal{S}_3$ in $\ker(x_1+x_2+x_3) \cong \mathbb{R}^2$ (i.e., the space $\{(x_1,x_2,x_3) \in \mathbb{R}^3 : x_1+x_2+x_3 = 0\}$).

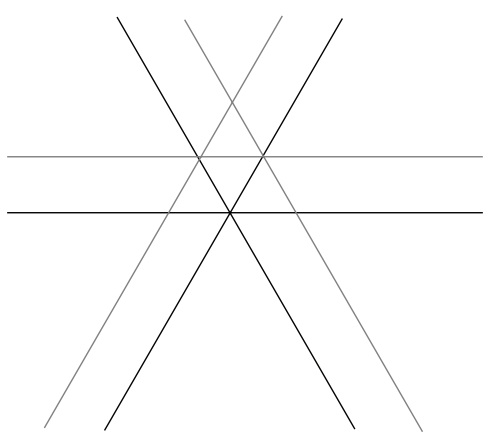


Figure 2. The Shi arrangement $\mathcal{S}_3$ in $\ker(x_1+x_2+x_3)$

Theorem 5.16. *The characteristic polynomial of $\mathcal{S}_n$ is given by*

$$\chi_{\mathcal{S}_n}(t) = t(t-n)^{n-1}.$$

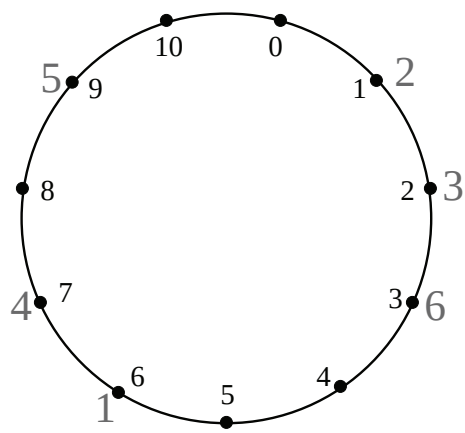
Proof. Let p be a large prime. By Theorem 5.15 we have

$$\chi_{\mathcal{S}_n}(p) = \#\{(\alpha_1, \ldots, \alpha_n) \in \mathbb{F}_p^n \,:\, i < j \Rightarrow \alpha_i \neq \alpha_j \text{ and } \alpha_i \neq \alpha_j + 1\}.$$

Choose a weak ordered partition $\pi = (B_1, \ldots, B_{p-n})$ of $[n]$ into $p-n$ blocks, i.e., $\bigcup B_i = [n]$ and $B_i \cap B_j = \emptyset$ if $i \neq j$, such that $1 \in B_1$. ("Weak" means that we allow $B_i = \emptyset$.) For $2 \leq i \leq n$ there are $p-n$ choices for j such that $i \in B_j$, so $(p-n)^{n-1}$ choices in all. We will illustrate the following argument with the example $p = 11$, $n = 6$, and

$$\pi = (\{1,4\}, \{5\}, \emptyset, \{2,3,6\}, \emptyset). \tag{38}$$

Arrange the elements of $\mathbb{F}_p$ clockwise on a circle. Place $1, 2, \ldots, n$ on some n of these points as follows. Place elements of B_1 consecutively (clockwise) in increasing order with 1 placed at some element $\alpha_1 \in \mathbb{F}_p$. Skip a space and place the elements of B_2 consecutively in increasing order. Skip another space and place the elements of B_3 consecutively in increasing order, etc. For our example (38), say $\alpha_1 = 6$.



Let α_i be the position (element of $\mathbb{F}_p$) at which i was placed. For our example we have

$$(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6) = (6, 1, 2, 7, 9, 3).$$

It is easily verified that we have defined a bijection from the $(p-n)^{n-1}$ weak ordered partitions $\pi = (B_1, \ldots, B_{p-n})$ of $[n]$ into $p-n$ blocks such that $1 \in B_1$, together with the choice of $\alpha_1 \in \mathbb{F}_p$, to the set $\mathbb{F}_p^n - \cup_{H \in (\mathcal{S}_n)_p} H$. There are $(p-n)^{n-1}$ choices for π and p choices for α_1, so it follows from Theorem 5.15 that $\chi_{\mathcal{S}_n}(t) = t(t-n)^{n-1}$. $\square$

We obtain the following corollary immediately from Theorem 2.5.

Corollary 5.11. *We have $r(\mathcal{S}_n) = (n+1)^{n-1}$ and $b(\mathcal{S}_n) = (n-1)^{n-1}$.*

Note. Since $r(\mathcal{S}_n)$ and $b(\mathcal{S}_n)$ have such simple formulas, it is natural to ask for a direct bijective proof of Corollary 5.11. A number of such proofs are known; a sketch that $r(\mathcal{S}_n) = (n+1)^{n-1}$ is given in Exercise 3.

Note. It can be shown that the cone $c\mathcal{S}_n$ is not supersolvable for $n \geq 3$ (Exercise 4) but is free in the sense of Theorem 4.14.

5.3. Exponential Sequences of Arrangements

The braid arrangement (in fact, any Coxeter arrangement) is highly symmetrical; indeed, the group of linear transformations that preserves the arrangement acts transitively on the regions. Thus all regions "look the same." The Shi arrangement lacks this symmetry, but it still possesses a kind of "combinatorial symmetry" that allows us to express the characteristic polynomials $\chi_{\mathcal{S}_n}(t)$, for all $n \geq 1$, in terms of the number $r(\mathcal{S}_n)$ of regions.

Definition 5.14. A sequence $\mathfrak{A} = (\mathcal{A}_1, \mathcal{A}_2, \dots)$ of arrangements is called an *exponential sequence of arrangements* (ESA) if it satisfies the following three conditions.

(1) $\mathcal{A}_n$ is in K^n for some field K (independent of n).
(2) Every $H \in \mathcal{A}_n$ is parallel to some hyperplane H' in the braid arrangement $\mathcal{B}_n$ (over K).
(3) Let S be a k-element subset of $[n]$, and define

$$\mathcal{A}_n^S = \{H \in \mathcal{A}_n \,:\, H \text{ is parallel to } x_i - x_j = 0 \text{ for some } i, j \in S\}.$$

Then $L(\mathcal{A}_n^S) \cong L(\mathcal{A}_k)$.

Examples of ESA's are given by $\mathcal{A}_n = \mathcal{B}_n$ or $\mathcal{A}_n = \mathcal{S}_n$. In fact, in these cases we have $\mathcal{A}_n^S \cong \mathcal{A}_k \times K^{n-k}$.

The combinatorial properties of ESA's are related to the *exponential formula* in the theory of exponential generating functions [**32**, §5.1], which we now review. Informally, we are dealing with "structures" that can be put on a vertex set V such that each structure is a disjoint union of its "connected components." We obtain a structure on V by partitioning V and placing a connected structure on each block (independently). Examples of such structures are graphs, forests, and posets, but not trees or groups. Let $h(n)$ be the total number of structures on an n-set V (with $h(0) = 1$), and let $f(n)$ be the number that are connected. The exponential formula states that

$$\sum_{n \geq 0} h(n) \frac{x^n}{n!} = \exp \sum_{n \geq 1} f(n) \frac{x^n}{n!}. \tag{39}$$

More precisely, let $f : \mathbb{P} \to R$, where R is a commutative ring. (For our purposes, $R = \mathbb{Z}$ will do.) Define a new function $h : \mathbb{N} \to R$ by $h(0) = 1$ and

$$h(n) = \sum_{\pi = \{B_1, \dots, B_k\} \in \Pi_n} f(\#B_1) f(\#B_2) \cdots f(\#B_k). \tag{40}$$

Then equation (39) holds. A straightforward proof can be given by considering the expansion

$$\begin{aligned} \exp \sum_{n \geq 1} f(n) \frac{x^n}{n!} &= \prod_{n \geq 1} \exp f(n) \frac{x^n}{n!} \\ &= \prod_{n \geq 1} \left(\sum_{k \geq 0} f(n)^k \frac{x^{kn}}{n!^k k!} \right). \end{aligned}$$

We omit the details (Exercise 5).

For any arrangement $\mathcal{A}$ in K^n, define $r(\mathcal{A}) = (-1)^n \chi_{\mathcal{A}}(-1)$. Of course if $K = \mathbb{R}$ this coincides with the definition of $r(\mathcal{A})$ as the number of regions of $\mathcal{A}$. We come to the main result concering ESA's.

Theorem 5.17. *Let $\mathfrak{A} = (\mathcal{A}_1, \mathcal{A}_2, \dots)$ be an ESA. Then*

$$\sum_{n\geq 0} \chi_{\mathcal{A}_n}(t)\frac{x^n}{n!} = \left(\sum_{n\geq 0}(-1)^n r(\mathcal{A}_n)\frac{x^n}{n!}\right)^{-t}.$$

Example 5.13. For $\mathfrak{A} = (\mathcal{B}_1, \mathcal{B}_2, \dots)$ Theorem 5.17 asserts that

$$\sum_{n\geq 0} t(t-1)\cdots(t-n+1)\frac{x^n}{n!} = \left(\sum_{n\geq 0}(-1)^n n!\frac{x^n}{n!}\right)^{-t},$$

as immediately follows from the binomial theorem. On the other hand, if $\mathfrak{A} = (\mathcal{S}_1, \mathcal{S}_2, \dots)$, then we obtain the much less obvious identity

$$\sum_{n\geq 0} t(t-n)^{n-1}\frac{x^n}{n!} = \left(\sum_{n\geq 0}(-1)^n (n+1)^{n-1}\frac{x^n}{n!}\right)^{-t}.$$

Proof of Theorem 5.17. By Whitney's theorem (Theorem 2.4) we have for any arrangement $\mathcal{A}$ in K^n that

$$\chi_{\mathcal{A}}(t) = \sum_{\substack{\mathcal{B}\subseteq\mathcal{A}\\ \mathcal{B}\text{ central}}} (-1)^{\#\mathcal{B}} t^{n-\operatorname{rank}(\mathcal{B})}.$$

Let $\mathfrak{A} = (\mathcal{A}_1, \mathcal{A}_2, \dots)$, and let $\mathcal{B} \subseteq \mathcal{A}_n$ for some n. Define $\pi(\mathcal{B}) \in \Pi_n$ to have blocks that are the vertex sets of the connected components of the graph G on $[n]$ with edges

$$E(G) = \{ij \,:\, \exists\, x_i - x_j = c \text{ in } \mathcal{B}\}. \tag{41}$$

Define

$$\tilde{\chi}_{\mathcal{A}_n}(t) = \sum_{\substack{\mathcal{B}\subseteq\mathcal{A}\\ \mathcal{B}\text{ central}\\ \pi(\mathcal{B})=[n]}} (-1)^{\#\mathcal{B}} t^{n-\operatorname{rk}(\mathcal{B})}.$$

Then

$$\begin{aligned}\chi_{\mathcal{A}_n}(t) &= \sum_{\pi=\{B_1,\dots,B_k\}\in\Pi_n} \sum_{\substack{\mathcal{B}\subseteq\mathcal{A}\\ \mathcal{B}\text{ central}\\ \pi(\mathcal{B})=\pi}} (-1)^{\#\mathcal{B}} t^{n-\operatorname{rk}(\mathcal{B})}\\ &= \sum_{\pi=\{B_1,\dots,B_k\}\in\Pi_n} \tilde{\chi}_{\mathcal{A}_{\#B_1}}(t)\tilde{\chi}_{\mathcal{A}_{\#B_2}}(t)\cdots\tilde{\chi}_{\mathcal{A}_{\#B_k}}(t).\end{aligned}$$

Thus by the exponential formula (39),

$$\sum_{n\geq 0}\chi_{\mathcal{A}_n}(t)\frac{x^n}{n!} = \exp\sum_{n\geq 1}\tilde{\chi}_{\mathcal{A}_n}(t)\frac{x^n}{n!}.$$

But $\pi(\mathcal{B}) = [n]$ if and only if $\mathrm{rk}(\mathcal{B}) = n-1$, so $\tilde{\chi}_{\mathcal{A}_n}(t) = c_n t$ for some $c_n \in \mathbb{Z}$. We therefore get

$$\begin{aligned}\sum_{n\geq 0} \chi_{\mathcal{A}_n}(t)\frac{x^n}{n!} &= \exp t \sum_{n\geq 1} c_n \frac{x^n}{n!} \\ &= \left(\sum_{n\geq 0} b_n \frac{x^n}{n!}\right)^t,\end{aligned}$$

where $\exp \sum_{n\geq 1} c_n \frac{x^n}{n!} = \sum_{n\geq 0} b_n \frac{x^n}{n!}$. Put $t = -1$ to get

$$\sum_{n\geq 0} (-1)^n r(\mathcal{A}_n)\frac{x^n}{n!} = \left(\sum_{n\geq 0} b_n \frac{x^n}{n!}\right)^{-1},$$

from which it follows that

$$\sum_{n\geq 0} \chi_{\mathcal{A}_n}(t)\frac{x^n}{n!} = \left(\sum_{n\geq 0} (-1)^n r(\mathcal{A}_n)\frac{x^n}{n!}\right)^{-t}.$$

□

For a generalization of Theorem 5.17, see Exercise 10.

5.4. The Catalan Arrangement

Define the *Catalan arrangement* $\mathcal{C}_n$ in K^n, where $\mathrm{char}(K) \neq 2$, by

$$Q_{\mathcal{C}_n}(x) = \prod_{1\leq i<j\leq n} (x_i - x_j)(x_i - x_j - 1)(x_i - x_j + 1).$$

Equivalently, the hyperplanes of $\mathcal{C}_n$ are given by

$$x_i - x_j = -1, 0, 1, \quad 1 \leq i < j \leq n.$$

Thus $\mathcal{C}_n$ has $3\binom{n}{2}$ hyperplanes, and $\mathrm{rank}(\mathcal{C}_n) = n-1$.

Assume now that $K = \mathbb{R}$. The symmetric group $\mathfrak{S}_n$ acts on $\mathbb{R}^n$ by permuting coordinates, i.e.,

$$w \cdot (x_1, \ldots, x_n) = (x_{w(1)}, \ldots, x_{w(n)}).$$

Here we are multiplying permutations left-to-right, e.g., $(1,2)(2,3) = (1,3,2)$ (in cycle form), so $vw \cdot \alpha = v \cdot (w \cdot \alpha)$. Both $\mathcal{B}_n$ and $\mathcal{C}_n$ are $\mathfrak{S}_n$-invariant, i.e., $\mathfrak{S}_n$ permutes the hyperplanes of these arrangements. Hence $\mathfrak{S}_n$ also permutes their regions, and each region $x_{w(1)} > x_{w(2)} > \cdots > x_{w(n)}$ of $\mathcal{B}_n$ is divided "in the same way" in $\mathcal{C}_n$. In particular, if $r_0(\mathcal{C}_n)$ denotes the number of regions of $\mathcal{C}_n$ contained in some fixed region of $\mathcal{B}_n$, then $r(\mathcal{C}_n) = n! r_0(\mathcal{C}_n)$. See Figure 3 for $\mathcal{C}_3$ in the ambient space $\ker(x_1 + x_2 + x_3)$, where the hyperplanes of $\mathcal{B}_3$ are drawn as solid lines and the remaining hyperplanes as dashed lines. Each region of $\mathcal{B}_3$ contains five regions of $\mathcal{C}_3$, so $r(\mathcal{C}_3) = 6 \cdot 5 = 30$.

We can compute $r(\mathcal{C}_n)$ (or equivalently $r_0(\mathcal{C}_n)$) by a direct combinatorial argument. Let R_0 denote the region $x_1 > x_2 > \cdots > x_n$ of $\mathcal{B}_n$. The regions of $\mathcal{C}_n$ contained in R_0 are determined by those $i < j$ such that $x_i - x_j < 1$. We need only specify the *maximal* intervals $[i,j]$ such that $x_i - x_j < 1$, i.e., if $a \leq i < j \leq b$ and $x_a - x_b < 1$, then $a = i$ and $b = j$. It is easy to see that any such specification of maximal intervals determines a region of $\mathcal{C}_n$ contained in R_0. Thus $r_0(\mathcal{C}_n)$ is equal

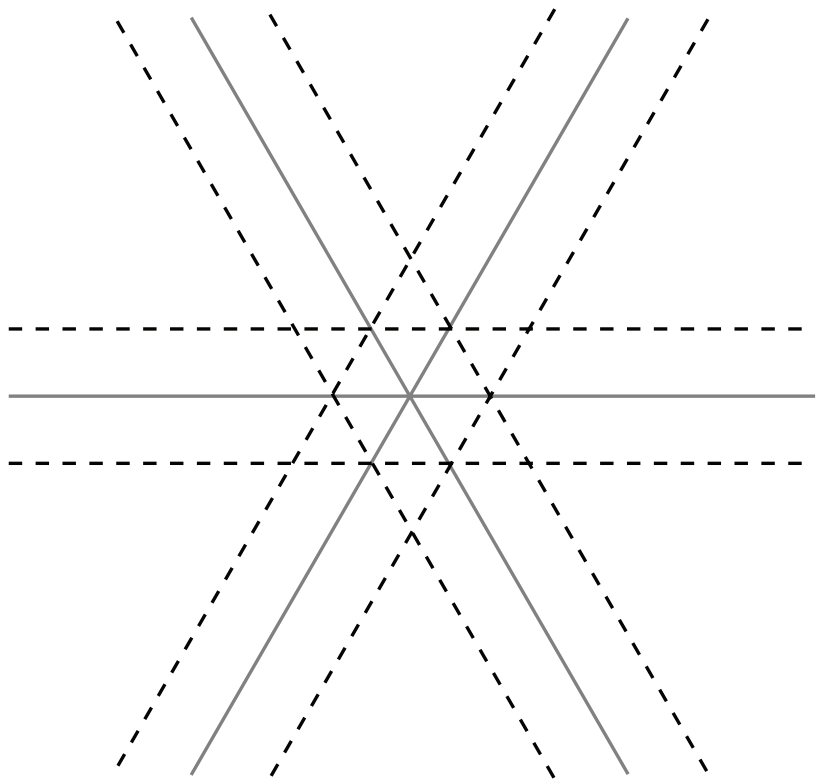


Figure 3. The Catalan arrangement $\mathcal{C}_3$ in $\ker(x_1 + x_2 + x_3)$

to the number of *antichains* A of strict intervals of $[n]$, i.e., sets A of intervals $[i, j]$, where $1 \leq i < j \leq n$, such that no interval in A is contained in another. ("Strict" means that $i = j$ is not allowed.) It is known (equivalent to [**32**, Exer. 6.19(bbb)]) that the number of such antichains is the *Catalan number* $C_n = \frac{1}{n+1}\binom{2n}{n}$. For the sake of completeness we give a bijection between these antichains and a standard combinatorial structure counted by Catalan numbers, viz., lattice paths from $(0, 0)$ to (n, n) with steps $(1, 0)$ and $(0, 1)$, never rising above the line $y = x$ ([**32**, Exer. 6.19(h)]). Given an antichain A of intervals of $[n]$, there is a unique lattice path of the claimed type whose "outer corners" (a step $(1, 0)$ followed by $(0, 1)$) consist of the points $(j, i - 1)$ where $[i, j] \in A$, together with the points $(i, i - 1)$ where no interval in A contains i. Figure 4 illustrates this bijection for $n = 8$ and $A = \{[1, 4], [3, 5], [7, 8]\}$.

We have therefore proved the following result. For a refinement, see Exercise 11.

Proposition 5.14. *The number of regions of the Catalan arrangement $\mathcal{C}_n$ is given by $r(\mathcal{C}_n) = n!C_n$. Each region of $\mathcal{B}_n$ contains C_n regions of $\mathcal{C}_n$.*

In fact, there is a simple formula for the characteristic polynomial $\chi_{\mathcal{C}_n}(t)$.

Theorem 5.18. *We have*

$$\chi_{\mathcal{C}_n}(t) = t(t - n - 1)(t - n - 2)(t - n - 3) \cdots (t - 2n + 1).$$

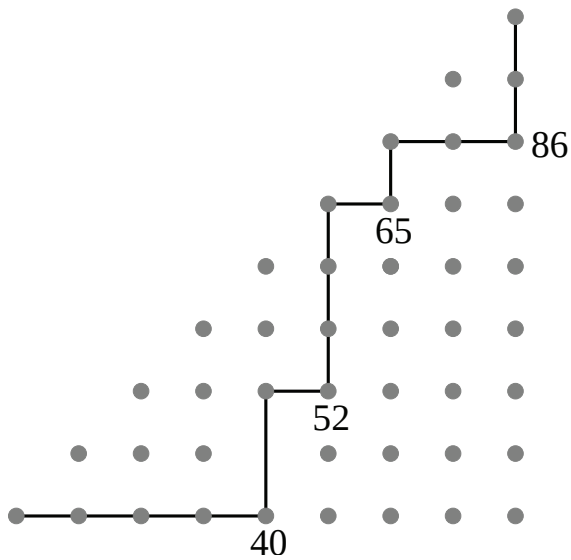


Figure 4. A bijection corresponding to $A = \{[1,4],[3,5],[7,8]\}$

Proof. Clearly the sequence $(\mathcal{C}_1, \mathcal{C}_2, \dots)$ is an ESA, so by Theorem 5.17 we have

$$\begin{aligned}\sum_{n\geq 0} \chi_{\mathcal{C}_n}(t)\frac{x^n}{n!} &= \left(\sum_{n\geq 0}(-1)^n n! C_n \frac{x^n}{n!}\right)^{-t} \\ &= \left(\sum_{n\geq 0}(-1)^n C_n x^n\right)^{-t}.\end{aligned}$$

One method for expanding this series is to use the *Lagrange inversion formula* [**32**, Thm. 5.4.2]. Let $F(x) = a_1 x + a_2 x^2 + \cdots$ be a formal power series over K, where $\mathrm{char}(K) = 0$ and $a_1 \neq 0$. Then there exists a unique formal power series $F^{\langle -1\rangle} = a_1^{-1}x + \cdots$ satisfying

$$F(F^{\langle -1\rangle}(x)) = F^{\langle -1\rangle}(F(x)) = x.$$

Let $k, t \in \mathbb{Z}$. The Lagrange inversion formula states that

$$t[x^t]F^{\langle -1\rangle}(x)^k = k[x^{t-k}]\left(\frac{x}{F(x)}\right)^t. \tag{42}$$

Let $y = \sum_{n\geq 0}(-1)^n C_n x^{n+1}$. By a fundamental property of Catalan numbers, $y^2 = -y + x$. Hence $y = (x + x^2)^{\langle -1\rangle}$. Substitute $t - n$ for k and apply equation (42) to $y = F(x)$, so $F^{\langle -1\rangle}(x) = x + x^2$:

$$t[x^t](x + x^2)^{t-n} = (t-n)[x^n]\left(\frac{x}{y}\right)^t. \tag{43}$$

The right-hand side of (43) is just

$$(t-n)[x^n]\left(\frac{y}{x}\right)^{-t} = \frac{(t-n)\chi_{\mathcal{C}_n}(t)}{n!}.$$

The left-hand side of (43) is given by

$$t[x^t]x^{t-n}(1+x)^{t-n} = t\binom{t-n}{n} = \frac{t(t-n)(t-n-1)\cdots(t-2n+1)}{n!}.$$

It follows that

$$\chi_{\mathcal{C}_n}(t) = t(t-n-1)(t-n-2)(t-n-3)\cdots(t-2n+1)$$

for all $t \in \mathbb{Z}$. It then follows easily (e.g., using the fact that a polynomial in one variable over a field of characteristic 0 is determined by its values on $\mathbb{Z}$) that this equation holds when t is an indeterminate. □

NOTE. It is not difficult to give an alternative proof of Theorem 5.18 based on the finite field method (Exercise 12).

5.5. Interval Orders

The subject of interval orders has a long history (see [**15**][**36**]), but only recently [**33**] was their connection with arrangements noticed. Let $P = \{I_1, \ldots, I_n\}$ be a finite set of closed intervals $I_i = [a_i, b_i]$, where $a_i, b_i \in \mathbb{R}$ and $a_i < b_i$. Partially order P by defining $I_i < I_j$ if $b_i < a_j$, i.e., I_i lies entirely to the left of I_j on the real number line. A poset isomorphic to P is called an *interval order*. Figure 5 gives an example of six intervals and the corresponding interval order. It is understood that the real line lies below and parallel to the line segments labelled $a, \ldots, f$, and that the actual intervals are the projections of these line segments to $\mathbb{R}$. If all the intervals I_i have length one, then P is called a *semiorder* or *unit interval order*.

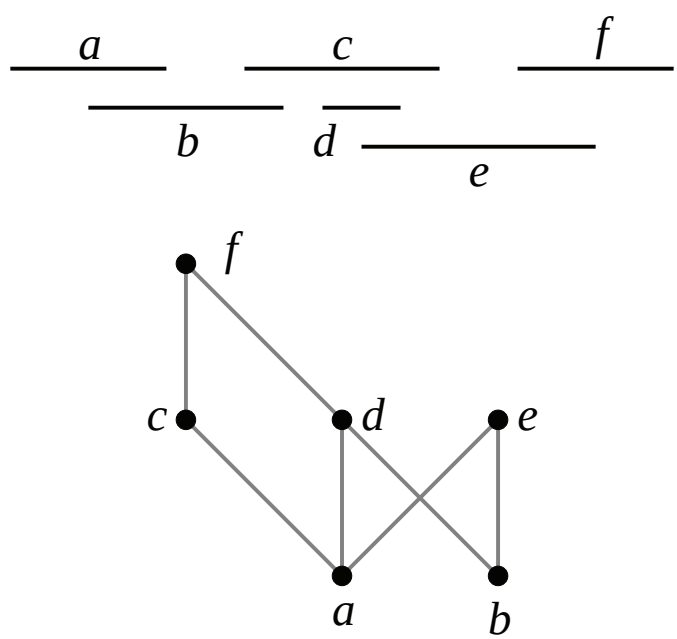


Figure 5. An example of an interval order

We will be considering both labelled and unlabelled interval orders. A *labelled* interval order is the same as an interval order on a set S, often taken to be $[n]$.

If an interval order P corresponds to intervals $I_1, \ldots, I_n$, then there is a natural labeling of P, viz., label the element corresponding to I_i by i. Thus the intervals $I_1 = [0,1]$ and $I_2 = [2,3]$ correspond to the labelled interval order P_1 defined by $1 < 2$, while the intervals $I_1 = [2,3]$ and $I_2 = [0,1]$ correspond to P_2 defined by $2 < 1$. Note that P_1 and P_2 are different labelled interval orders but are isomorphic as posets. As another example, consider the intervals $I_1 = [0,2]$ and $I_2 = [1,3]$. The corresponding labelled interval order P consists of the disjoint points 1 and 2. If we now let $I_1 = [1,3]$ and $I_2 = [0,2]$, then we obtain the *same* labelled interval order (or labelled poset) P, although the intervals themselves have been exchanged. An *unlabelled* interval order may be regarded as an isomorphism class of interval orders; two intervals orders P_1 and P_2 represent the same unlabelled interval order if and only if they are isomorphic. Of course our discussion of labelled and unlabelled interval orders applies equally well to semiorders.

Figure 6 shows the five nonisomorphic (or unlabelled) interval orders (which for three vertices coincides with semiorders) with three vertices, and below them the number of distinct labelings. (In general, the number of labelings of an n-element poset P is $n!/\#\mathrm{Aut}(P)$, where $\mathrm{Aut}(P)$ denotes the automorphism group of P.) It follows that there are 19 labelled interval orders or labelled semiorders on a 3-element set.

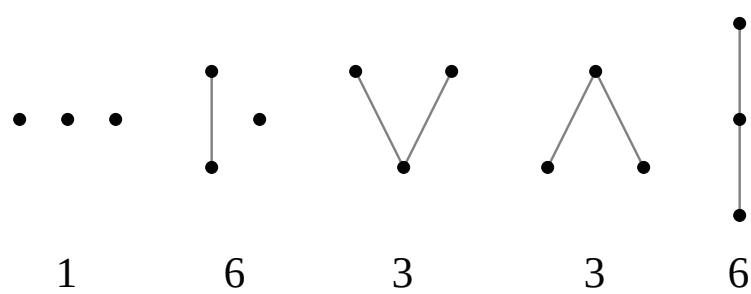


Figure 6. The number of labelings of semiorders with three elements

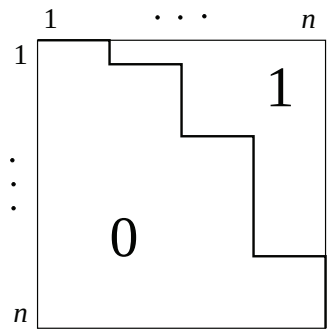
The following proposition collects some basic results on interval orders. We simply state them without proof. Only part (a) is needed in what follows (Lemma 5.6). We use the notation $\boldsymbol{i}$ to denote an i-element chain and $P+Q$ to denote the disjoint union of the posets P and Q.

Proposition 5.15.

(a) *A finite poset is an interval order if and only if it has no induced subposet isomorphic to* $\mathbf{2}+\mathbf{2}$.

(b) *A finite poset is a semiorder if and only if it has no induced subposet isomorphic to* $\mathbf{2}+\mathbf{2}$ *or* $\mathbf{3}+\mathbf{1}$.

(c) *A finite poset P is a semiorder if and only if its elements can be ordered as $I_1, \ldots, I_n$ so that the incidence matrix of P (i.e., the matrix $M = (m_{ij})$, where $m_{ij} = 1$ if $I_i < I_j$ and $m_{ij} = 0$ otherwise) has the form shown below. Moreover, all such semiorders are nonisomorphic.*



In (c) above, the southwest boundary of the positions of the 1's in M form a lattice path which by suitable indexing goes from $(0,0)$ to (n,n) with steps $(0,1)$ and $(1,0)$, never rising above $y = x$. Since the number of such lattice paths is the Catalan number C_n, it follows that the number of nonisomorphic n-element semiorders is C_n. Later (Proposition 5.17) we will give a proof based on properties of a certain arrangement. Figure 7 illustrates Proposition 5.15(c) when $n = 3$. It shows the matrices M, the corresponding set of unit intervals, and the associated semiorder.

$$\begin{bmatrix} 0&0&0\\0&0&0\\0&0&0 \end{bmatrix} \quad \begin{bmatrix} 0&0&1\\0&0&0\\0&0&0 \end{bmatrix} \quad \begin{bmatrix} 0&1&1\\0&0&0\\0&0&0 \end{bmatrix} \quad \begin{bmatrix} 0&0&1\\0&0&1\\0&0&0 \end{bmatrix} \quad \begin{bmatrix} 0&1&1\\0&0&1\\0&0&0 \end{bmatrix}$$

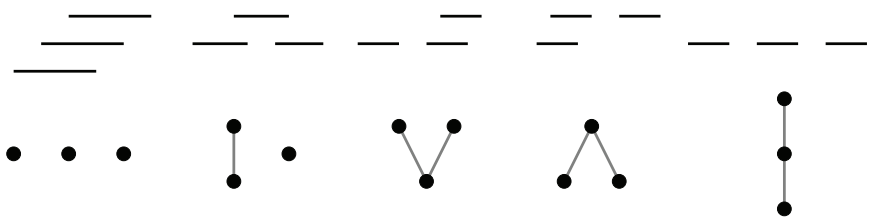


Figure 7. The semiorders with three elements

Let $\ell_1, \ldots, \ell_n > 0$ and set $\eta = (\ell_1, \ldots, \ell_n)$. Let $\mathcal{P}_\eta$ denote the set of all interval orders P on $[n]$ such that there exist a set $I_1, \ldots, I_n$ of intervals corresponding to P (with I_i corresponding to $i \in P$) such that $\ell(I_i) = \ell_i$. In other words, $i \overset{P}{<} j$ if and only if I_i lies entirely to the left of I_j. For instance, it follows from Figure 6 that $\#\mathcal{P}_{(1,1,1)} = 19$.

We now come to the connection with arrangements. Given $\eta = (\ell_1, \ldots, \ell_n)$ as above, define the arrangement $\mathcal{I}_\eta$ in $\mathbb{R}^n$ by letting its hyperplanes be given by

$$x_i - x_j = \ell_i, \quad i \neq j.$$

(Note the condition $i \neq j$, not $i < j$.) Thus $\mathcal{I}_\eta$ has rank $n-1$ and $n(n-1)$ hyperplanes (since $\ell_i > 0$). Figure 8 shows the arrangement $\mathcal{I}_{(1,1,1)}$ in the space $\ker(x_1 + x_2 + x_3)$.

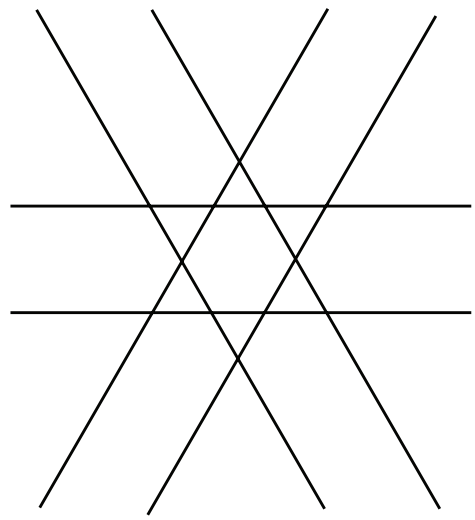


Figure 8. The arrangement $\mathcal{I}_{(1,1,1)}$ in the space $\ker(x_1 + x_2 + x_3)$

Proposition 5.16. *Let $\eta \in \mathbb{R}^n_+$. Then $r(\mathcal{I}_\eta) = \#\mathcal{P}_\eta$.*

Proof. Let $(x_1, \ldots, x_n)$ belong to some region R of $\mathcal{I}_\eta$. Define the interval $I_i = [x_i - \ell_i, x_i]$. The region R is determined by whether $x_i - x_j < \ell_i$ or $x_i - x_j > \ell_i$. Equivalently, $I_i \not> I_j$ or $I_i > I_j$ in the ordering on intervals that defines interval orders. Hence the number of possible interval orders corresponding to intervals $I_1, \ldots, I_n$ with $\ell(I_i) = \ell_i$ is just $r(\mathcal{I}_\eta)$. □

Consider the case $\ell_1 = \cdots = \ell_n = 1$, so we are looking at the *semiorder arrangement* $x_i - x_j = 1$ for $i \neq j$. We abbreviate $(1, 1, \ldots, 1)$ as 1^n and denote this arrangement by $\mathcal{I}_{1^n}$. By the proof of Proposition 5.16 the regions of $\mathcal{I}_{1^n}$ are in a natural bijection with semiorders on $[n]$.

Now note that $\mathcal{C}_n = \mathcal{I}_{1^n} \cup \mathcal{B}_n$, where $\mathcal{C}_n$ denotes the Catalan arrangement. Fix a region R of $\mathcal{B}_n$, say $x_1 < x_2 < \cdots < x_n$. Then the number of regions of $\mathcal{I}_{1^n}$ that intersect R is the number of semiorders on $[n]$ that correspond to (unit) intervals $I_1, \ldots, I_n$ with right endpoints $x_1 < x_2 < \cdots < x_n$. Another set $I'_1, \ldots, I'_n$ of unit intervals $I'_i = [x'_i - 1, x'_i]$ with $x'_1 < x'_2 < \cdots < x'_n$ defines a different region from that defined by $I_1, \ldots, I_n$ if and only if the corresponding semiorders are *nonisomorphic*. It follows that the number of nonisomorphic semiorders on $[n]$ is equal to the number of regions of $\mathcal{I}_{1^n}$ intersecting the region $x_1 < x_2 < \cdots < x_n$ of $\mathcal{B}_n$. Since $\mathcal{C}_n = \mathcal{I}_{1^n} \cup \mathcal{B}_n$, there follows from Proposition 5.14 the following result of Wine and Freunde [**38**].

Proposition 5.17. *The number $u(n)$ of nonisomorphic n-element semiorders is given by*

$$u(n) = \frac{1}{n!} r(\mathcal{C}_n) = C_n.$$

Figure 9 shows the nonisomorphic 3-element semiorders corresponding to the regions of $\mathcal{C}_n$ intersecting the region $x_1 < x_2 < \cdots < x_n$ of $\mathcal{B}_n$.

We now come to the problem of determining $r(\mathcal{I}_{1^n})$, the number of semiorders on $[n]$.

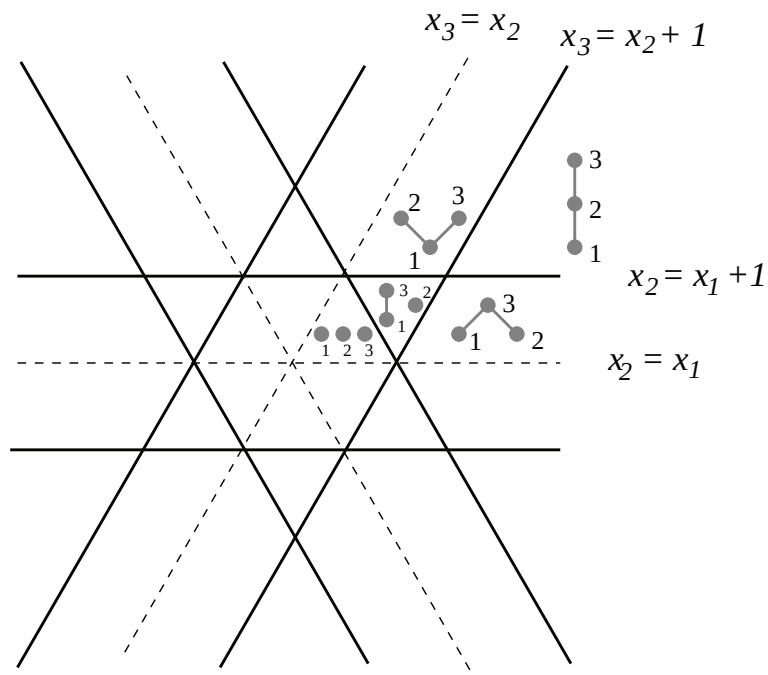


Figure 9. The nonisomorphic 3-element semiorders as regions of $\mathcal{C}_{1^n}$

Theorem 5.19. *Fix distinct real numbers $a_1, a_2, \ldots, a_m > 0$. Let $\mathcal{A}_n$ be the arrangement in $\mathbb{R}^n$ with hyperplanes*

$$\mathcal{A}_n : \quad x_i - x_j = a_1, \ldots, a_m, \quad i \neq j,$$

and let $\mathcal{A}_n^* = \mathcal{A}_n \cup \mathcal{B}_n$. Define

$$\begin{aligned} F(x) &= \sum_{n \geq 1} r(\mathcal{A}_n) \frac{x^n}{n!} \\ G(x) &= \sum_{n \geq 1} r(\mathcal{A}_n^*) \frac{x^n}{n!}. \end{aligned}$$

Then $F(x) = G(1 - e^{-x})$.

Proof. Let $c(n,k)$ denote the number of permutations w of n objects with k cycles (in the disjoint cycle decomposition of w). The integer $c(n,k)$ is known as a *signless Stirling number of the first kind* and for fixed k has the exponential generating function

$$\sum_{n \geq 0} c(n,k) \frac{x^n}{n!} = \frac{1}{k!} \left(\log(1-x)^{-1} \right)^k. \tag{44}$$

For futher information, see e.g. [**31**, pp. 17–20][**32**, (5.25)].

We have

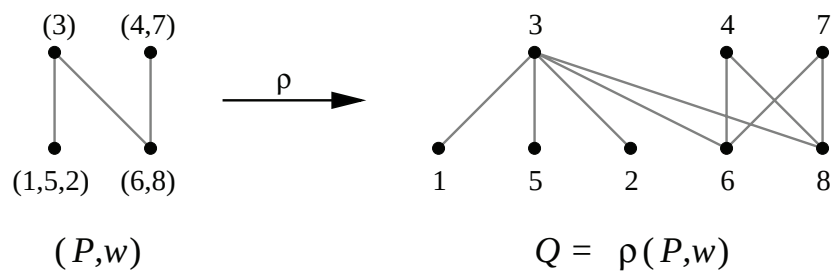
$$\begin{aligned} F(x) = G(1-e^{-x}) \Leftrightarrow G(x) &= F(\log(1-x)^{-1}) \\ &= \sum_{k\geq 1} r(\mathcal{A}_k)\frac{1}{k!}\left(\log(1-x)^{-1}\right)^k \\ &= \sum_{k\geq 1} r(\mathcal{A}_k)\sum_{n\geq 0} c(n,k)\frac{x^n}{n!}. \end{aligned}$$

It follows that we need to show that

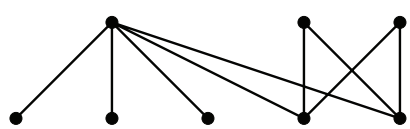
$$r(\mathcal{A}_n^*) = \sum_{k=1}^{n} c(n,k) r(\mathcal{A}_k). \tag{45}$$

For simplicity we consider only the case $m = 1$ and $a_1 = 1$, but the argument is completely analogous in the general case. When $m = 1$ and $a_1 = 1$ we have that $r(\mathcal{A}_n^*) = n!C_n$ and that $r(\mathcal{A}_n)$ is the number of semiorders on $[n]$. Thus it suffices to give a map $(P, w) \overset{\rho}{\mapsto} Q$, where $w \in \mathfrak{S}_k$ and P is a semiorder whose elements are labelled by the cycles of w, and where Q is an unlabelled n-element semiorder, such that ρ is $n!$-to-1, i.e., every Q appears exactly $n!$ times as an image of some (P, w).

Choose $w \in \mathfrak{S}_n$ with k cycles in $c(n,k)$ ways, and make these cycles the vertices of a semiorder P in $r(\mathcal{A}_k)$ ways. Define a new poset $\rho(P, w)$ as follows: if the cycle $(c_1, \ldots, c_j)$ is an element of P, then replace it with an antichain with elements $c_1, \ldots, c_j$. Given $1 \leq c \leq n$, let $C(c)$ be the cycle of w containing c. Define $c < d$ in $\rho(P, w)$ if $C(c) < C(d)$ in P. We illustrate this definition with $n = 8$ and $w = (1,5,2)(3)(6,8)(4,7)$:



Given an unlabelled n-element semiorder Q, such as

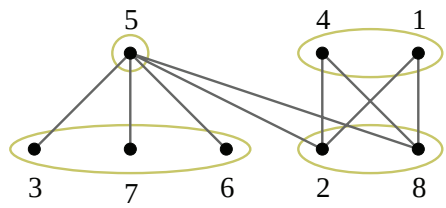


we now show that there are exactly $n!$ pairs (P, w) for which $\rho(P, w) \cong Q$. Call a pair of elements $x, y \in Q$ *autonomous* if for all $z \in Q$ we have

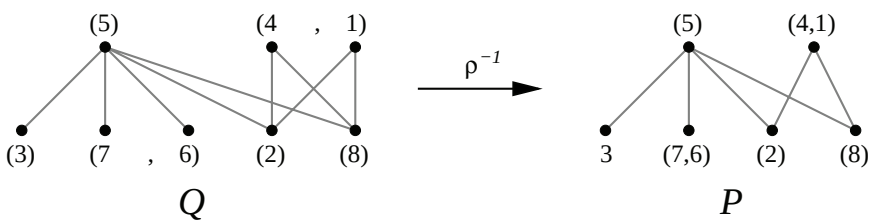
$$x < z \Leftrightarrow y < z, \qquad x > z \Leftrightarrow y > z.$$

Equivalently, the map $\tau : Q \to Q$ transposing x, y and fixing all other $z \in Q$ is an automorphism of Q. Clearly the relation of being autonomous is an equivalence relation. Partition Q into its autonomous equivalence classes. Regard the elements of Q as being distinguished, and choose a bijection (labeling) $\varphi : Q \to [n]$ (in $n!$ ways). Fix a linear ordering (independent of φ) of the elements in each equivalence

class. (The linear ordering of the elements in each equivalence class in the diagram below is left-to-right.)



In each class, place a left parenthesis before each left-to-right maximum, and place a right parenthesis before each left parenthesis and at the end. (This is the bijection $\mathfrak{S}_n \to \mathfrak{S}_n$, $\hat{w} \mapsto w$, in [**31**, p. 17].) Merge the elements $c_1, c_2, \ldots, c_j$ (appearing in that order) between each pair of parentheses into a single element labelled with the cycle $(c_1, c_2, \ldots, c_j)$.



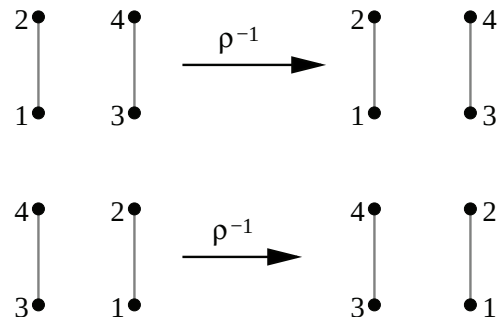
We have thus obtained a poset P whose elements are labelled by the cycles of a permutation $w \in \mathfrak{S}_n$, such that $\rho(P, w) = Q$. For each unlabelled Q, there are exactly $n!$ pairs (P, w) (where the poset P is labelled by the cycles of $w \in \mathfrak{S}_n$) for which $\rho(P, w) \cong Q$. Since by Proposition 5.17 there are C_n nonisomorphic n-element semiorders, we get

$$n!C_n = \sum_{k=1}^{n} c(n,k) r(\mathcal{A}_k).$$

□

NOTE. Theorem 5.19 can also be proved using Burnside's lemma (also called the Cauchy-Frobenius lemma) from group theory.

To test one's understanding of the proof of Theorem 5.19, consider why it doesn't work for *all* posets. In other words, let $f(n)$ denote the number of posets on $[n]$ and $g(n)$ the number of nonisomorphic n-element posets. Set $F(x) = \sum f(n)\frac{x^n}{n!}$ and $G(x) = \sum g(n)x^n$. Why doesn't the above argument show that $G(x) = F(1 - e^{-x})$? Let $Q = \mathbf{2} + \mathbf{2}$ (the unique obstruction to being an interval order, by Proposition 5.15(a)). The autonomous classes have one element each. Consider the two labelings $\varphi : Q \to [4]$ and the corresponding ρ^{-1}:



We obtain the same labelled posets in both cases, so the proof of Theorem 5.19 fails. The key property of interval orders that the proof of Theorem 5.19 uses implicitly is the following.

Lemma 5.6. *If $\sigma : P \to P$ is an automorphism of the interval order P and $\sigma(x) = \sigma(y)$, then x and y are autonomous.*

Proof. Assume not. Then there exists an element $s \in P$ satisfying $s > x$, $s \not> y$ (or dually). Since $\sigma(x) = y$, there must exist $t \in P$ satisfying $t > y$, $t \not> x$. But then $\{x, s, y, t\}$ form an induced $\mathbf{2} + \mathbf{2}$, so by Proposition 5.15(a) P is not an interval order. $\square$

Specializing $m = 1$ and $a_1 = 1$ in Theorem 5.19 yields the following corollary, due first (in an equivalent form) to Chandon, Lemaire and Pouget [**12**].

Corollary 5.12. *Let $f(n)$ denote the number of semiorders on $[n]$ (or n-element labelled semiorders). Then*

$$\sum_{n \geq 0} f(n) \frac{x^n}{n!} = C(1 - e^{-x}),$$

where

$$C(x) = \sum_{n \geq 0} C_n x^n = \frac{1 - \sqrt{1 - 4x}}{2x}.$$

5.6. Intervals with Generic Lengths

A particularly interesting class of interval orders are those corresponding to intervals with specified *generic lengths* $\eta = (\ell_1, \ldots, \ell_n)$. Intuitively, this means that the intersection poset $P(\mathcal{I}_\eta)$ is as "large as possible." One way to make this precise is to say that η is *generic* if $P(\mathcal{I}_\eta) \cong P(\mathcal{I}_{\eta'})$, where $\eta' = (\ell'_1, \ldots, \ell'_n)$ and the ℓ'_i's are linearly independent over $\mathbb{Q}$. Thus if η is generic, then the intersection poset $L(\mathcal{I}_\eta)$ does not depend on η, but rather only on n. In particular, $r(\mathcal{I}_\eta)$ does not depend on η (always assuming η is generic). Hence by Proposition 5.16, the number $\#\mathcal{P}_\eta$ of labelled interval orders corresponding to intervals $I_1, \ldots, I_n$ with $\ell(I_i) = \ell_i$ depends only on n. This fact is not at all obvious combinatorially, since the interval orders themselves do depend on η. For instance, it is easy to see that $\eta = (1, 1.0001, 1.001, 1.01, 1.1)$ is generic and that no corresponding interval order can be isomorphic to $\mathbf{4} + \mathbf{1}$. On the other hand, $\eta = (1, 10, 100, 1000, 10000)$ is also generic, but this time there is a corresponding interval order isomorphic to $\mathbf{4} + \mathbf{1}$. (See Exercise 17.)

The preceding discussion raises the question of computing $\#\mathcal{P}_n$ when η is generic. We write $\mathcal{G}_n$ for the corresponding interval order $x_i - x_j = \ell_i$, $i \neq j$, since the intersection poset depends only on n. The following result is a nice application of arrangements to "pure" enumeration; no proof is known except the one sketched here.

Theorem 5.20. *Let*

$$z = \sum_{n\geq 0} r(\mathcal{G}_n)\frac{x^n}{n!} = 1 + x + 3\frac{x^2}{2!} + 19\frac{x^3}{3!} + 195\frac{x^4}{4!} + 2831\frac{x^5}{5!} + \cdots .$$

Define a power series

$$y = 1 + x + 5\frac{x^2}{2!} + 46\frac{x^3}{3!} + 631\frac{x^4}{4} + \cdots$$

by $1 = y(2 - e^{xy})$. *Equivalently,*

$$y = 1 + \left(\frac{1}{1+x}\log\frac{1+2x}{1+x}\right)^{\langle -1\rangle}.$$

Then z is the unique power series satisfying $z'/z = y^2$, $z(0) = 1$.

NOTE. The condition $z'/z = y^2$ can be rewritten as $z = \exp\int y^2\,dx$.

Sketch of proof. Putting $t = -1$ in Theorem 2.4 gives

$$r(\mathcal{G}_n) = \sum_{\substack{\mathcal{B}\subseteq\mathcal{G}_n\\ \mathcal{B}\text{ central}}} (-1)^{\#\mathcal{B}-\mathrm{rk}(\mathcal{B})}. \tag{46}$$

Given a central subarrangement $\mathcal{B} \subseteq \mathcal{G}_n$, define a digraph (directed graph) $G_{\mathcal{B}}$ on $[n]$ by letting $i \to j$ be a (directed) edge if the hyperplane $x_i - x_j = \ell_i$ belongs to $\mathcal{B}$. One then shows that as an undirected graph $G_{\mathcal{B}}$ is bipartite, i.e., the vertices can be partitioned into two subsets U and V such that all edges connect a vertex in U to a vertex in V. The pair (U, V) is called a *vertex bipartition* of $G_{\mathcal{B}}$. Moreover, if B is a block of $G_{\mathcal{B}}$ (as defined preceding Proposition 4.11), say with vertex bipartition (U_B, V_B), then either all edges of B are directed from U_B to V_B, or all edges are directed from V_B to U_B. It can also be seen that all such directed bipartite graphs can arise in this way. It follows that equation (46) can be rewritten

$$r(\mathcal{G}_n) = (-1)^n \sum_G (-1)^{e(G)+c(G)} 2^{b(G)}, \tag{47}$$

where G ranges over all (undirected) bipartite graphs on $[n]$, $e(G)$ denotes the number of edges of G, and $b(G)$ denotes the number of blocks of G.

Equation (47) reduces the problem of determining $r(\mathcal{G})$ to a (rather difficult) problem in enumeration, whose solution may be found in [**25**, §6]. □

5.7. Other Examples

There are two additional arrangements related to the braid arrangement that involve nice enumerative combinatorics. We merely repeat the definitions here from Lecture 1 and assemble some of their basic properties in Exercises 19–28.

The *Linial arrangement* in K^n is given by the hyperplanes $x_i - x_j = 1$, $1 \leq i < j \leq n$. It consists of "half" of the semiorder arrangement $\mathcal{I}_{1^n}$. Despite its similarity to $\mathcal{I}_{1^n}$, it is considerably more difficult to obtain its characteristic polynomial and other enumerative invariants. Finally the *threshold arrangement* in K^n is given by

the hyperplanes $x_i + x_j = 0$, $1 \le i < j \le n$. It is a subarrangement of the Coxeter arrangements $\mathcal{A}(B_n)$ ($=\mathcal{A}(C_n)$) and $\mathcal{A}(D_n)$.

Exercises

(1) [2] Verify equation (37), viz.,

$$\chi_{\mathcal{A}(D_n)}(t) = (t-1)(t-3)\cdots(t-(2n-3))\cdot(t-n+1).$$

(2) [2] Draw a picture of the projectivization of the Coxeter arrangement $\mathcal{A}(B_3)$, similar to Figure 1 of Lecture 1.

(3) (a) [2] An *embroidered permutation* of $[n]$ consists of a permutation w of $[n]$ together with a collection $\mathcal{E}$ of ordered pairs (i,j) such that:
- $1 \le i < j \le n$ for all $(i,j) \in \mathcal{E}$.
- If (i,j) and (h,k) are distinct elements of $\mathcal{E}$, then it is false that $i \le h \le k \le j$.
- If $(i,j) \in \mathcal{E}$ then $w(i) < w(j)$.

For instance, the three embroidered permutations $(w,\mathcal{E})$ of [2] are given by $(12,\emptyset)$, $(12,\{(1,2)\})$, and $(21,\emptyset)$. Give a bijective proof that the number $r(\mathcal{S}_n)$ of regions of the Shi arrangement $\mathcal{S}_n$ is equal to the number of embroidered permutations of $[n]$.

(b) [2+] A *parking function* of length n is a sequence $(a_1,\ldots,a_n) \in \mathbb{P}^n$ whose increasing rearrangement $b_1 \le b_2 \le \cdots \le b_n$ satisfies $b_i \le i$. For instance, the parking functions of length three are 11, 12, 21. Give a bijective proof that the number of parking functions of length n is equal to the number of embroidered permutations of $[n]$.

(c) [3–] Give a combinatorial proof that the number of parking functions of length n is equal to $(n+1)^{n-1}$.

(4) [2+] Show that if $\mathcal{S}_n$ denotes the Shi arrangement, then the cone $c\mathcal{S}_n$ is not supersolvable for $n \ge 3$.

(5) [2] Show that if $f : \mathbb{P} \to R$ and $h : \mathbb{N} \to R$ are related by equation (40) (with $h(0) = 1$), then equation (39) holds.

(6) (a) [2] Compute the characteristic polynomial of the arrangement $\mathcal{B}'_n$ in $\mathbb{R}^n$ with defining polynomial

$$Q(x) = (x_1 - x_n - 1)\prod_{1\le i<j\le n}(x_i - x_j).$$

In other words, $\mathcal{B}'_n$ consists of the braid arrangement together with the hyperplane $x_1 - x_n = 1$.

(b) [5–] Is $c\mathcal{B}'_n$ (the cone over $\mathcal{B}'_n$) supersolvable?

(7) [2+] Let $1 \le k \le n$. Find the characteristic polynomial of the arrangement $\mathcal{S}_{n,k}$ in $\mathbb{R}^n$ defined by

$$\begin{array}{lll} x_i - x_j = 0 & \text{for} & 1 \le i < j \le n \\ x_i - x_j = 1 & \text{for} & 1 \le i < j \le k. \end{array}$$

(8) [2+] Let $1 \le k \le n$. Find the characteristic polynomial of the arrangement $\mathcal{C}_{n,k}$ in $\mathbb{R}^n$ defined by

$$\begin{array}{rll} x_i = 0 & \text{for} & 1 \le i \le n \\ x_i \pm x_j = 0 & \text{for} & 1 \le i < j \le n \\ x_i + x_j = 1 & \text{for} & 1 \le i < j \le k. \end{array}.$$

In particular, show that $r(\mathcal{C}_{n,k}) = 2^{n-k} n! \binom{2k}{k}$.

(9) (a) [2+] Let $\mathcal{A}_n$ be the arrangement in $\mathbb{R}^n$ with hyperplanes $x_i = 0$ for all i, $x_i = x_j$ for all $i < j$, and $x_i = 2x_j$ for all $i \neq j$. Show that

$$\chi_{\mathcal{A}_n}(t) = (t-1)(t-n-2)_{n-1},$$

where $(x)_m = x(x-1)\cdots(x-m+1)$. In particular, $r(\mathcal{A}_n) = 2(2n+1)!/(n+2)!$. Can this be seen combinatorially? (This last question has not been worked on.)

(b) [2+] Now let $\mathcal{A}_n$ be the arrangement in $\mathbb{R}^n$ with hyperplanes $x_i = x_j$ for all $i < j$ and $x_i = 2x_j$ for all $i \neq j$. Show that

$$\chi_{\mathcal{A}_n}(t) = (t-1)(t-n-2)_{n-3}(t^2 - (3n-1)t + 3n(n-1)).$$

In particular, $r(\mathcal{A}_n) = 6n^2(2n-1)!/(n+2)!$. Again, a combinatorial proof can be asked for.

(c) [5–] Modify. For instance, what about the arrangement with hyperplanes $x_i = 0$ for all i, $x_i = x_j$ for all $i < j$, and $x_i = 2x_j$ for all $i < j$? (This example is actually not difficult.) Or $x_i = 0$ for all i, $x_i = x_j$ for all $i < j$, $x_i = 2x_j$ for all $i \neq j$, and $x_i = 3x_j$ for all $i \neq j$?

(10) (a) [2+] For $n \geq 1$ let $\mathcal{A}_n$ be an arrangement in $\mathbb{R}^n$ such that every $H \in \mathcal{A}_n$ is parallel to a hyperplane of the form $x_i = cx_j$, where $c \in \mathbb{R}$. Just as in the definition of an exponential sequence of arrangements, define for every subset S of $[n]$ the arrangement

$$\mathcal{A}_n^S = \{H \in \mathcal{A}_n : H \text{ is parallel to some } x_i = cx_j, \text{ where } i, j \in S\}.$$

Suppose that for every such S we have $L_{\mathcal{A}_n^S} \cong L_{\mathcal{A}_k}$, where $k = \#S$. Let

$$\begin{aligned} F(x) &= \sum_{n\geq 0} (-1)^n r(\mathcal{A}_n) \frac{x^n}{n!} \\ G(x) &= \sum_{n\geq 0} (-1)^{\operatorname{rank}(\mathcal{A}_n)} b(\mathcal{A}_n) \frac{x^n}{n!}. \end{aligned}$$

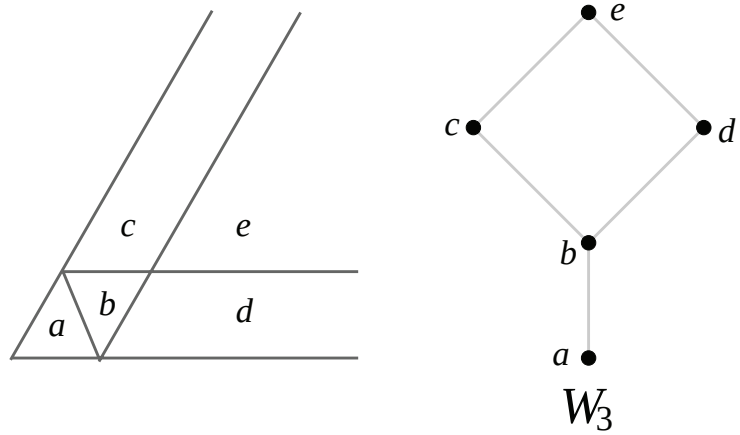
Show that

$$\sum_{n\geq 0} \chi_{\mathcal{A}_n}(t) \frac{x^n}{n!} = \frac{G(x)^{(t+1)/2}}{F(x)^{(t-1)/2}}. \tag{48}$$

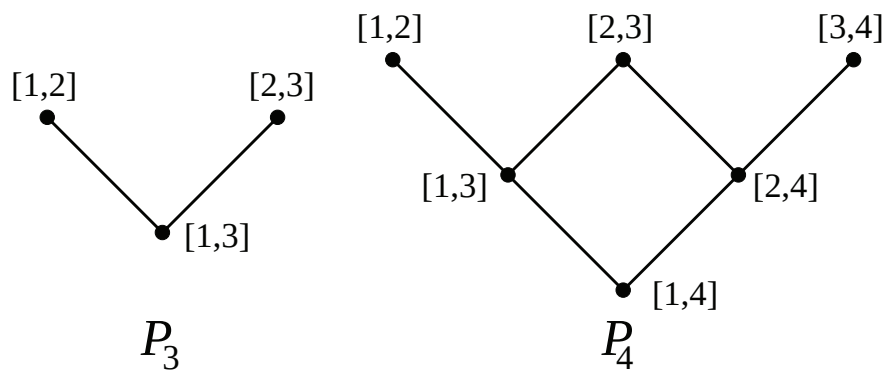
Verify that this formula is correct for the braid arrangement.

(b) [2] Simplify equation (48) when each $\mathcal{A}_n$, $n \geq 1$, is a central arrangement. Make sure that your simplification is valid for the coordinate hyperplane arrangement.

(11) [2+] Let $\mathcal{R}_0(\mathcal{C}_n)$ denote the set of regions of the Catalan arrangement $\mathcal{C}_n$ contained in the regions $x_1 > x_2 > \cdots > x_n$ of $\mathcal{B}_n$. Let $\hat{R}$ be the unique region in $\mathcal{R}_0(\mathcal{C}_n)$ whose closure contains the origin. For $R \in \mathcal{R}_0(\mathcal{C}_n)$, let X_R be the set of hyperplanes $H \in \mathcal{C}_n$ such that $\hat{R}$ and R lie on different sides of H. Let $W_n = \{X_R : R \in \mathcal{R}_0(\mathcal{C}_n)\}$, ordered by inclusion.



Let P_n be the poset of intervals $[i, j]$, $1 \le i < j \le n$, ordered by *reverse* inclusion.



Show that $W_n \cong J(P_n)$, the lattice of order ideals of P_n. (An *order ideal* of a poset P is a subset $I \subseteq P$ such that if $x \in I$ and $y \le x$, then $y \in I$. Define $J(P)$ to be the set of order ideals of P, ordered by inclusion. See [**31**, Thm. 3.4.1].)

(12) [2] Use the finite field method to prove that

$$\chi_{\mathcal{C}_n}(t) = t(t-n-1)(t-n-2)(t-n-3)\cdots(t-2n+1),$$

where $\mathcal{C}_n$ denotes the Catalan arrangement.

(13) [2+] Let $k \in \mathbb{P}$. Find the number of regions and characteristic polynomial of the *extended Catalan arrangement*

$$\mathcal{C}_n(k): \ x_i - x_j = 0, \pm 1, \pm 2, \ldots, \pm k, \text{ for } 1 \le i < j \le n.$$

Generalize Exercise 11 to the arrangements $\mathcal{C}_n(k)$.

(14) [3–] Let $\mathcal{S}_n^B$ denote the arrangement

$$\begin{aligned} x_i \pm x_j &= 0, 1, \qquad 1 \le i < j \le n \\ 2x_i &= 0, 1, \qquad 1 \le i \le n, \end{aligned}$$

called the *Shi arrangement of type B*. Find the characteristic polynomial and number of regions of $\mathcal{S}_n^B$. Is there a "nice" bijective proof of the formula for the number of regions?

(15) [5–] Let $1 \le k \le n$. Find the number of regions (or more generally the characteristic polynomial) of the arrangement (in $\mathbb{R}^n$)

$$x_i - x_j = \begin{cases} 1, & 1 \le i \le k \\ 2, & k+1 \le i \le n, \end{cases}$$

for all $i \neq j$. Thus we are counting interval orders on $[n]$ where the elements $1, 2, \ldots, k$ correspond to intervals of length one, while $k+1, \ldots, n$ correspond to intervals of length two. Is it possible to count such interval orders up to isomorphism (i.e., the unlabelled case)? What if the length 2 is replaced instead by a generic length a?

(16) [2+] A *double semiorder* on $[n]$ consists of two binary relations $<$ and $\ll$ on $[n]$ that arise from a set $x_1, \ldots, x_n$ of real numbers as follows:

$$\begin{aligned} i < j \quad &\text{if} \quad x_i < x_j - 1 \\ i \ll j \quad &\text{if} \quad x_i < x_j - 2. \end{aligned}$$

If we associate the interval $I_i = [x_i - 2, x_i]$ with the point x_i, then we are specifying whether I_i lies to the left of the midpoint of I_j, entirely to the left of I_j, or neither. It should be clear what is meant for two double semiorders to be *isomorphic*.

(a) [2] Draw interval diagrams of the 12 nonisomorphic double semiorders on $\{1, 2, 3\}$.

(b) [2] Let $\rho_2(n)$ denote the number of double semiorders on $[n]$. Find an arrangement $\mathcal{I}_n^{(2)}$ satisfying $r(\mathcal{I}_n^{(2)}) = \rho_2(n)$.

(c) [2+] Show that the number of nonisomorphic double semiorders on $[n]$ is given by $\frac{1}{2n+1}\binom{3n}{n}$.

(d) [2–] Let $F(x) = \sum_{n \ge 0} \frac{1}{2n+1}\binom{3n}{n} x^n$. Show that

$$\sum_{n \ge 0} \rho_2(n) \frac{x^n}{n!} = F(1 - e^{-x}).$$

(e) [2] Generalize to "k-semiorders," where ordinary semiorders (or unit interval orders) correspond to $k = 1$ and double semiorders to $k = 2$.

(17) [1+] Show that intervals of lengths $1, 1.0001, 1.001, 1.01, 1.1$ cannot form an interval order isomorphic to $\mathbf{4} + \mathbf{1}$, but that such an interval order *can* be formed if the lengths are $1, 10, 100, 1000, 10000$.

(18) [5–] What more can be said about interval orders with generic interval lengths? For instance, consider the two cases: (a) interval lengths very near each other (e.g., 1, 1.001, 1.01, 1.1), and (b) interval lengths superincreasing (e.g., 1, 10, 100, 1000). Are there finitely many obstructions to being such an interval order? Can the number of unlabelled interval orders of each type be determined? (Perhaps the numbers are the same, but this seems unlikely.)

(19) (a) [3] Let $\mathcal{L}_n$ denote the Linial arrangement, say in $\mathbb{R}^n$. Show that

$$\chi_{\mathcal{L}_n}(t) = \frac{t}{2^n} \sum_{k=1}^{n} \binom{n}{k} (t-k)^{n-1}.$$

(b) [1+] Deduce from (a) that

$$\frac{\chi_{\mathcal{L}_n}(t)}{t} = \frac{(-1)^n \chi_{\mathcal{L}_n}(-t+n)}{-t+n}.$$

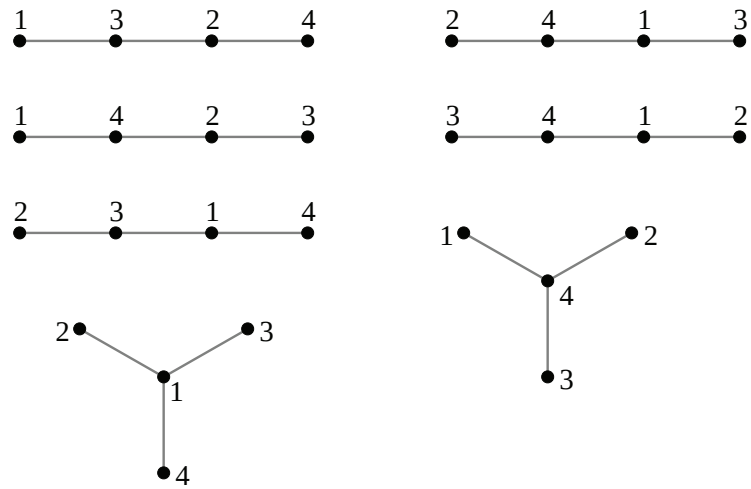


Figure 10. The seven alternating trees on the vertex set [4]

(20) (a) [3–] An *alternating tree* on the vertex set $[n]$ is a tree on $[n]$ such that every vertex is either less than all its neighbors or greater than all its neighbors. Figure 10 shows the seven alternating trees on [4]. Deduce from Exercise 19(a) that $r(\mathcal{L}_n)$ is equal to the number of alternating trees on $[n+1]$.

(b) [5] Find a bijective proof of (a), i.e., give an explicit bijection between the regions of $\mathcal{L}_n$ and the alternating trees on $[n+1]$.

(21) [3–] Let

$$\chi_{\mathcal{L}_n}(t) = a_n t^n - a_{n-1}t^{n-1} + \cdots + (-1)^{n-1}a_1 t.$$

Deduce from Exercise 19(a) that a_i is the number of alternating trees on the vertex set $0, 1, \ldots, n$ such that vertex 0 has degree (number of adjacent vertices) i.

(22) (a) [2+] Let $P(t) \in \mathbb{C}[t]$ have the property that every (complex) zero of $P(t)$ has real part a. Let $z \in \mathbb{C}$ satisfy $|z| = 1$. Show that every zero of the polynomial $P(t-1) + zP(t)$ has real part $a + \frac{1}{2}$.

(b) [2+] Deduce from (a) and Exercise 19(a) that every zero of the polynomial $\chi_{\mathcal{L}_n}(t)/t$ has real part $n/2$. This result is known as the "Riemann hypothesis for the Linial arrangement."

(23) (a) [2–] Compute $\lim_{n\to\infty} b(\mathcal{S}_n)/r(\mathcal{S}_n)$, where $\mathcal{S}_n$ denotes the Shi arrangement.

(b) [3] Do the same for the Linial arrangement $\mathcal{L}_n$.

(24) [2+] Let $\mathcal{L}_n$ denote the Linial arrangement in $\mathbb{R}^n$. Fix an integer $r \neq 0, \pm 1$, and let $\mathcal{M}_n(r)$ be the arrangement in $\mathbb{R}^n$ defined by $x_i = rx_j$, $1 \leq i < j \leq n$, together with the coordinate hyperplanes $x_i = 0$. Find a relationship between $\chi_{\mathcal{L}_n}(t)$ and $\chi_{\mathcal{M}_n(r)}(t)$ *without* explicitly computing these characteristic polynomials.

(25) (a) [3–] A *threshold graph* on $[n]$ may be defined recursively as follows: (i) the empty graph $\emptyset$ is a threshold graph, (ii) if G is a threshold graph, then so is the disjoint union of G and a single vertex, and (iii) if G is a threshold graph, then so is the graph obtained by adding a new vertex v and connecting it to every vertex of G. Let $\mathcal{T}_n$ denote the threshold arrangement. Show that $r(\mathcal{T}_n)$ is the number of threshold graphs on $[n]$.

(b) [2+] Deduce from (a) that

$$\sum_{n\geq 0} r(\mathcal{T}_n)\frac{x^n}{n!} = \frac{e^x(1-x)}{2-e^x}.$$

(c) [1+] Deduce from Exercise 10 that

$$\sum_{n\geq 0} \chi_{\mathcal{T}_n}(t)\frac{x^n}{n!} = (1+x)(2e^x-1)^{(t-1)/2}.$$

(26) [5–] Let

$$\chi_{\mathcal{T}_n}(t) = t^n - a_{n-1}t^{n-1} + \cdots + (-1)^n a_0.$$

For instance,

$$\begin{aligned}
\chi_{\mathcal{T}_3}(t) &= t^3 - 3t^2 + 3t - 1 \\
\chi_{\mathcal{T}_4}(t) &= t^4 - 6t^3 + 15t^2 - 17t + 7 \\
\chi_{\mathcal{T}_5}(t) &= t^5 - 10t^4 + 45t^3 - 105t^2 + 120t - 51.
\end{aligned}$$

By Exercise 25(a), $a_0 + a_1 + \cdots + a_{n-1} + 1$ is the number of threshold graphs on the vertex set $[n]$. Give a combinatorial interpretation of the numbers a_i as the number of threshold graphs with a certain property.

(27) (a) [1+] Find the number of regions of the "Linial threshold arrangement"

$$x_i + x_j = 1, \quad 1 \leq i < j \leq n.$$

(b) [5–] Find the number of regions, or even the characteristic polynomial, of the "Shi threshold arrangement"

$$x_i + x_j = 0, 1, \quad 1 \leq i < j \leq n.$$

(28) [3–] Let $\mathcal{A}_n$ denote the "generic threshold arrangement" (in $\mathbb{R}^n$) $x_i + x_j = a_{ij}$, $1 \leq i < j \leq n$, where the a_{ij}'s are generic. Let

$$T(x) = \sum_{n\geq 1} n^{n-2}\frac{x^n}{n!},$$

the generating function for labelled trees on n vertices. Let

$$R(x) = \sum_{n\geq 1} n^{n-1}\frac{x^n}{n!},$$

the generating function for *rooted* labelled trees on n vertices. Show that

$$\begin{aligned}
\sum_{n\geq 0} r(\mathcal{A}_n)\frac{x^n}{n!} &= e^{T(x)-\frac{1}{2}R(x)}\left(\frac{1+R(x)}{1-R(x)}\right)^{1/4} \\
&= 1 + x + 2\frac{x^2}{2!} + 8\frac{x^3}{3!} + 54\frac{x^4}{4!} + 533\frac{x^5}{5!} + 6934\frac{x^6}{6!} + \cdots.
\end{aligned}$$

(29) [2+] Fix $n \geq 1$. Let $f(k,n,r)$ be the number of $k \times n$ $(0,1)$-matrices A over the rationals such that all rows of A are distinct, every row has at least one 1, and $\operatorname{rank}(A) = r$. Let $g_n(q)$ be the number of n-tuples $(a_1, \ldots, a_n) \in \mathbb{F}_q^n$ such that no nonempty subset of the entries sums to 0 (in $\mathbb{F}_q$). Show that for $p \gg 0$, where $q = p^d$, we have

$$g_n(q) = \sum_{k,r} \frac{(-1)^k}{k!} f(k,n,r) q^{n-r}.$$

(The case $k = 0$ is included, corresponding to the empty matrix, which has rank 0.)

LECTURE 6
Separating Hyperplanes

6.1. The Distance Enumerator

Let $\mathcal{A}$ be a real arrangement, and let R and R' be regions of $\mathcal{A}$. A hyperplane $H \in \mathcal{A}$ *separates* R and R' if R and R' lie on opposite sides of H. In this chapter we will consider some results dealing with separating hyperplanes. To begin, let

$$\mathrm{sep}(R, R') = \{H \in \mathcal{A} : H \text{ separates } R \text{ and } R'\}.$$

Define the *distance* $d(R, R')$ between the regions R and R' to be the number of hyperplanes $H \in \mathcal{A}$ that separate R and R', i.e.,

$$d(R, R') = \#\mathrm{sep}(R, R').$$

It is easily seen that d is a metric on the set $\mathcal{R}(\mathcal{A})$ of regions of $\mathcal{A}$, i.e.,

- $d(R, R') \geq 0$ for all $R, R' \in \mathcal{R}(\mathcal{A})$, with equality if and only if $R = R'$
- $d(R, R') = d(R', R)$ for all $R, R' \in \mathcal{R}(\mathcal{A})$
- $d(R, R') + d(R', R'') \geq d(R, R'')$ for all $R, R', R'' \in \mathcal{R}(\mathcal{A})$.

Now fix a region $R_0 \in \mathcal{R}(\mathcal{A})$, called the *base region*. The *distance enumerator* of $\mathcal{A}$ (with respect to R_0) is the polynomial

$$D_{\mathcal{A},R_0}(t) = \sum_{R \in \mathcal{R}(\mathcal{A})} t^{d(R_0,R)}.$$

We simply write $D_{\mathcal{A}}(t)$ if no confusion will result. Also define the *weak order* (with respect to R_0) of $\mathcal{A}$ to be the partial order $W_{\mathcal{A}}$ on $\mathcal{R}(\mathcal{A})$ given by

$$R \leq R' \text{ if } \mathrm{sep}(R_0, R) \subseteq \mathrm{sep}(R_0, R').$$

It is easy to see that $W_{\mathcal{A}}$ is a partial ordering of $\mathcal{R}(\mathcal{A})$. The poset $W_{\mathcal{A}}$ is graded by distance from R_0, i.e., R_0 is the $\hat{0}$ element of $\mathcal{R}(\mathcal{A})$, and all saturated chains between R_0 and R have length $d(R_0, R)$.

Figure 1 shows three arrangements in $\mathbb{R}^2$, with R_0 labelled 0 and then each $R \neq R_0$ labelled $d(R_0, R)$. Under each arrangement is shown the corresponding weak order $W_{\mathcal{A}}$. The first arrangement is the braid arrangement $\mathcal{B}_3$ (essentialized). Here the choice of base region does not affect the distance enumerator $1+2t+2t^2+t^3 = (1+t)(1+t+t^2)$ nor the weak order. On the other hand, the second two

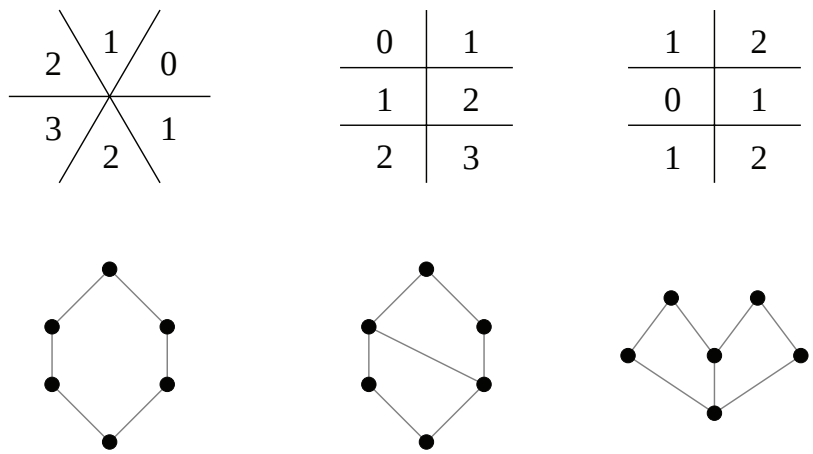


Figure 1. Examples of weak orders

arrangements of Figure 1 are identical, but the choice of R_0 leads to different weak orders and different distance enumerators, viz., $1 + 2t + 2t^2 + t^3$ and $1 + 3t + 2t^2$.

Consider now the braid arrangement $\mathcal{B}_n$. We know from Example 1.3 that the regions of $\mathcal{B}_n$ are in one-to-one correspondence with the permutations of $[n]$, viz.,

$$\begin{array}{rcl} \mathcal{R}(\mathcal{B}_n) & \leftrightarrow & \mathfrak{S}_n \\ x_{w(1)} > x_{w(2)} > \cdots > x_{w(n)} & \leftrightarrow & w. \end{array}$$

Given $w = a_1 a_2 \cdots a_n \in \mathfrak{S}_n$, define an *inversion* of w to be a pair (i, j) such that $i < j$ and $a_i > a_j$. Let $\ell(w)$ denote the number of inversions of w. The *inversion sequence* $\mathrm{IS}(w)$ of w is the vector $(c_1, \cdots, c_n)$, where

$$c_j = \#\{i \,:\, i < j,\ w^{-1}(j) < w^{-1}(i)\}.$$

Note that the condition $w^{-1}(j) < w^{-1}(i)$ is equivalent to i appearing to the right of j in w. For instance, $\mathrm{IS}(461352) = (0, 0, 1, 3, 1, 4)$. The inversion sequence is a modified form of the *inversion table* or of the *code* of w, as defined in the literature, e.g., [**31**, p. 21][**32**, solution to Exer. 6.19(x)]. For our purposes the inversion sequence is the most convenient. It is clear from the definition of $\mathrm{IS}(w)$ that if $\mathrm{IS}(w) = (c_1, \ldots, c_n)$ then $\ell(w) = c_1 + \cdots + c_n$. Moreover, is easy to see (Exercise 2) that a sequence $(c_1, \ldots, c_n) \in \mathbb{N}^n$ is the inversion sequence of a permutation $w \in \mathfrak{S}_n$ if and only if $c_i \le i - 1$ for $1 \le i \le n$. It follows that

$$\begin{aligned} \sum_{w \in \mathfrak{S}_n} t^{\ell(w)} &= \sum_{\substack{(c_1,\ldots,c_n) \\ 0 \le c_i \le i-1}} t^{c_1 + \cdots + c_n} \\ &= \left(\sum_{c_1=0}^{0} t^{c_1} \right) \cdots \left(\sum_{c_n=0}^{n-1} t^{c_n} \right) \\ &= 1 \cdot (1+t)(1+t+t^2) \cdots (1+t+\cdots+t^{n-1}), \end{aligned} \tag{49}$$

a standard result on permutation statistics [**31**, Cor. 1.3.10].

Denote by R_w the region of $\mathcal{B}_n$ corresponding to $w \in \mathfrak{S}_n$, and choose $R_0 = R_{\mathrm{id}}$, where id$= 12 \cdots n$, the identity permutation. Suppose that $R_u, R_v \in \mathcal{R}(\mathcal{B}_n)$ such that $\mathrm{sep}(R_0, R_v) = \{H\} \cup \mathrm{sep}(R_0, u)$ for some $H \in \mathcal{B}_n$, $H \notin \mathrm{sep}(R_0, R_u)$. Thus R_u

and R_v are separated by a single hyperplane H, and R_0 and R_u lie on the same side of H. Suppose that H is given by $x_i = x_j$ with $i < j$. Then i and j appear consecutively in u written as a word $a_1 \cdots a_n$ (since H is a bounding hyperplane of the region R_u) and i appears to the left of j (since R_0 and R_u lie on the same side of H). Thus v is obtained from u by transposing the adjacent pair ij of letters. It follows that $\ell(v) = \ell(u) + 1$. If $u(k) = i$ and we let $s_k = (k, k+1)$, the adjacent transposition interchanging k and $k+1$, then $v = us_k$.

The following result is an immediate consequence of equation (49) and mathematical induction.

Proposition 6.18. *Let $R_0 = R_{\mathrm{id}}$ as above. If $w \in \mathfrak{S}_n$ then $d(R_0, R_w) = \ell(w)$. Moreover,*

$$D_{\mathcal{B}_n}(t) = (1+t)(1+t+t^2)\cdots(1+t+\cdots+t^{n-1}).$$

There is a somewhat different approach to Proposition 6.18 which will be generalized to the Shi arrangement. We label each region R of $\mathcal{B}_n$ recursively by a vector $\lambda(R) = (c_1, \ldots, c_n) \in \mathbb{N}^n$ as follows.

- $\lambda(R_0) = (0, 0, \ldots, 0)$
- Let e_i denote the ith unit coordinate vector in $\mathbb{R}^n$. If the regions R and R' of $\mathcal{B}_n$ are separated by the single hyperplane H with the equation $x_i = x_j$, $i < j$, and if R and R_0 lie on the same side of H, then $\lambda(R') = \lambda(R) + e_j$.

Figure 2 shows the labels $\lambda(R)$ for $\mathcal{B}_3$.

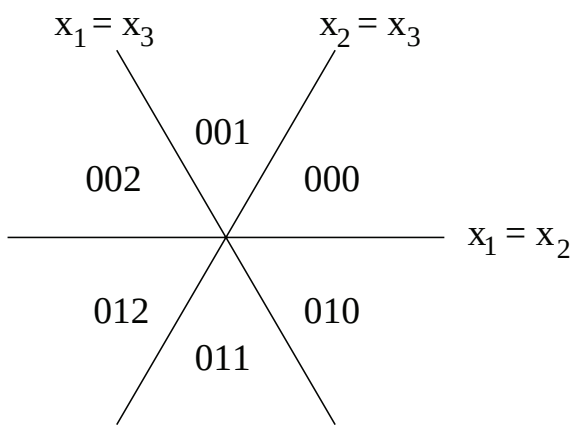


Figure 2. The inversion sequence labeling of the regions of $\mathcal{B}_3$

Proposition 6.19. *Let $w \in \mathfrak{S}_n$. Then $\lambda(R_w) = \mathrm{IS}(w)$, the inversion sequence of w.*

Proof. The proof is a straightforward induction on $\ell(w)$. If $\ell(w) = 0$, then $w =$id and

$$\lambda(R_{\mathrm{id}}) = \lambda(R_0) = (0, 0, \ldots, 0) = \mathrm{IS}(\mathrm{id}).$$

Suppose $w = a_1 \cdots a_n$ and $\ell(w) > 0$. For some $1 \le k \le n-1$ we must have $a_k = j > i = a_{k+1}$. Thus $\ell(ws_k) = \ell(w) - 1$. Hence by induction we may assume

$\lambda(ws_k) = \mathrm{IS}(ws_k)$. The hyperplane $x_i = x_j$ separates R_w from R_{ws_k}. Hence by the definition of λ we have

$$\lambda(R_w) = \lambda(R_{ws_k}) + e_j = \mathrm{IS}(ws_k) + e_j.$$

By the definition of the inversion sequence we have $\mathrm{IS}(ws_k) + e_j = \mathrm{IS}(w)$, and the proof follows. □

NOTE. The weak order $W_{\mathcal{B}_n}$ of the braid arrangement is an interesting poset, usually called the weak order or weak Bruhat order on $\mathfrak{S}_n$. For instance [**14**][**17**][**30**], the number of maximal chains of $W_{\mathcal{B}_n}$ is given by

$$\frac{\binom{n}{2}!}{1^{n-1}3^{n-2}5^{n-3}\cdots(2n-3)}.$$

For additional properties of $W_{\mathcal{B}_n}$, see [**5**].

6.2. Parking Functions and Tree Inversions

Some beautiful enumerative combinatorics is associated with the distance enumerator of the Shi arrangement $\mathcal{S}_n$ (for a suitable choice of R_0). The fundamental combinatorial object needed for this purpose is a parking function.

Definition 6.15. Let $n \in \mathbb{P}$. A *parking function* of length n is a sequence $(a_1, \ldots, a_n) \in \mathbb{Z}^n$ whose increasing rearrangment $b_1 \leq b_2 \leq \cdots \leq b_n$ satisfies $1 \leq b_i \leq i$ for $1 \leq i \leq n$. Equivalently, the sequence $(b_1 - 1, \ldots, b_n - 1)$ is the inversion sequence of some permutation $w \in \mathfrak{S}_n$.

The parking functions of length at most 3 are given as follows:

$$\begin{array}{cccccccc} & & 1 & & 11 & 12 & 21 & \\ 111 & 112 & 121 & 211 & 113 & 131 & 311 & 122 \\ 212 & 221 & 123 & 132 & 213 & 231 & 312 & 321 \end{array}.$$

The term "parking function" [**21**, §6] arises from the following scenario. A one-way street has parking spaces labelled $1, 2, \ldots, n$ in that order. There are n cars $C_1, \ldots, C_n$ which enter the street one at a time and try to park. Each car C_i has a preferred space $a_i \in [n]$. When it is C_i's turn to look for a space, it immediately drives to space a_i and then parks in the first available space. For instance, if $(a_1, a_2, a_3, a_4) = (2, 1, 2, 3)$, then C_1 parks in space 2, then C_2 parks in space 1, then C_3 goes to space 2 (which is occupied) parks in space 3 (the next available), and finally C_4 goes to space 3 and parks in space 4. On the other hand, if $(a_1, a_2, a_3, a_4) = (3, 1, 4, 3)$, then C_4 is unable to park, since its preferred space 3 and all subsequent spaces are already occupied. It is not hard to show (Exercise 3) that all the cars can park if and only if $(a_1, \ldots, a_n)$ is a parking function.

A basic question concerning parking functions (to be refined in Theorem 6.22) is their enumeration. The next result was first proved by Konheim and Weiss [**21**, §6]; we give an elegant proof due to Pollak (described in [**26**][**16**, p. 13]).

Proposition 6.20. *The number of parking functions of length n is $(n+1)^{n-1}$.*

Proof. Arrange $n+1$ (rather than n) parking spaces in a circle, labelled $1, \ldots, n+1$ in counterclockwise order. We still have n cars $C_1, \ldots, C_n$ with preferred spaces $(a_1, \ldots, a_n)$, but now we can have $1 \leq a_i \leq n+1$ (rather than $1 \leq a_i \leq n$). Each car enters the circle one at a time at their preferred space and then drives

counterclockwise until encountering an empty space, in which case the car parks there. Note the following:

- All the cars can always park, since they drive in a circle and will always find an empty space.
- After all cars have parked there will be one empty space.
- The sequence $(a_1, \ldots, a_n)$ is a parking function if and only if the empty space after all the cars have parked is $n+1$.
- If the preference sequence $(a_1, \ldots, a_n)$ produces the empty space i at the end, then the sequence $(a_1+k, \ldots, a_n+k)$ (taking entries modulo $n+1$ so they always lie in the set $[n+1]$) produces the empty space $i+k$ (modulo $n+1$).

It follows that exactly one of the sequences $(a_1+k, \ldots, a_n+k)$ (modulo $n+1$), where $1 \leq k \leq n+1$, is a parking function. There are $(n+1)^n$ sequences $(a_1, \ldots, a_n)$ in all, so exactly $(n+1)^n/(n+1) = (n+1)^{n-1}$ are parking functions. $\square$

Many readers will have recognized that the number $(n+1)^{n-1}$ is closely related to the enumeration of trees. Indeed, there is an intimate connection between trees and parking functions. We therefore now present some background material on trees. A *tree* on $[n]$ is a connected graph without cycles on the vertex set $[n]$. A *rooted* tree is a pair (T, i), where T is a tree and i is a vertex of T, called the *root*. We draw trees in the standard computer science manner with the root at the top and all edges emanating downwards. A *forest* on $[n]$ is a graph F on the vertex set $[n]$ for which every (connected) component is a tree. Equivalently, F has no cycles. A *rooted forest* (also called a *planted forest*) is a forest for which every component has a root, i.e., for each tree T of the forest select a vertex i_T of T to be the root of T. A standard result in enumerative combinatorics (e.g., [**32**, Prop. 5.3.2]) states that the number of rooted forests on $[n]$ is $(n+1)^{n-1}$.

An *inversion* of a rooted forest F on $[n]$ is a pair (i, j) of vertices such that $i < j$ and j appears on the (unique) path from i to the root of the tree in which i occurs. Write $\mathrm{inv}(F)$ for the number of inversions of F. For instance, the rooted forest F of Figure 3 has the inversions $(6, 7)$, $(1, 7)$, $(5, 7)$, $(1, 5)$, and $(2, 4)$, so $\mathrm{inv}(F) = 5$.

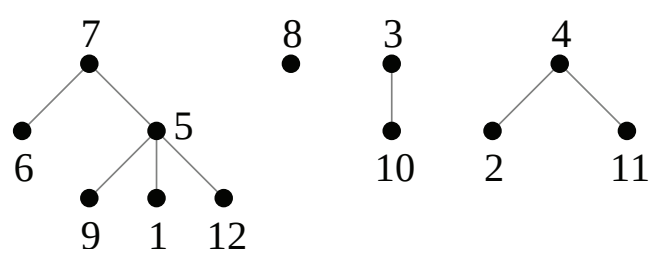


Figure 3. A rooted forest on [12]

Define the *inversion enumerator* $I_n(t)$ of rooted forests on $[n]$ by

$$I_n(t) = \sum_F t^{\mathrm{inv}(F)},$$

where F ranges over all rooted forests on $[n]$. Figure 4 shows the 16 rooted forests on [3] with their number of inversions written underneath, from which it follows that

$$I_3(t) = 6 + 6t + 3t^2 + t^3.$$

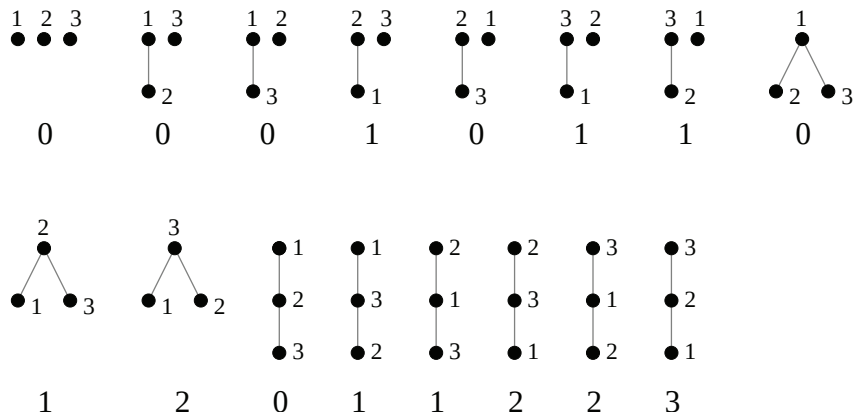


Figure 4. The 16 rooted forests on [3] and their number of inversions

We collect below the three main results on $I_n(t)$. They are theorems in "pure" enumeration and have no direct connection with arrangements. The first result, due to Mallows and Riordan [**23**], gives a remarkable connection with connected graphs.

Theorem 6.21. *We have*

$$I_n(1+t) = \sum_G t^{e(G)-n},$$

where G ranges over all connected (simple) graphs on the vertex set $[0,n] = \{0,1,\ldots,n\}$ and $e(G)$ denotes the number of edges of G.

For instance,

$$I_3(1+t) = 16 + 15t + 6t^2 + t^3.$$

Thus, for instance, there are 15 connected graphs on $[0,3]$ with four edges. Three of these are 4-cycles and twelve consist of a triangle with an incident edge. The enumeration of connected graphs is well-understood [**32**, Exam. 5.2.1]. In particular, if

$$C_n(t) = \sum_G t^{e(G)},$$

where G ranges over all connected (simple) graphs on $[n]$, then

$$\sum_{n\geq 0} C_n(t)\frac{x^n}{n!} = \log \sum_{n\geq 1} (1+t)^{\binom{n}{2}} \frac{x^n}{n!}. \tag{50}$$

Thus Theorem 6.21 "determines" $I_n(t)$. There is an alternative way to state this result that doesn't involve the logarithm function.

Corollary 6.13. *We have*

$$\sum_{n\geq 0} I_n(t)(t-1)^n \frac{x^n}{n!} = \frac{\displaystyle\sum_{n\geq 0} t^{\binom{n+1}{2}} \frac{x^n}{n!}}{\displaystyle\sum_{n\geq 0} t^{\binom{n}{2}} \frac{x^n}{n!}} \tag{51}$$

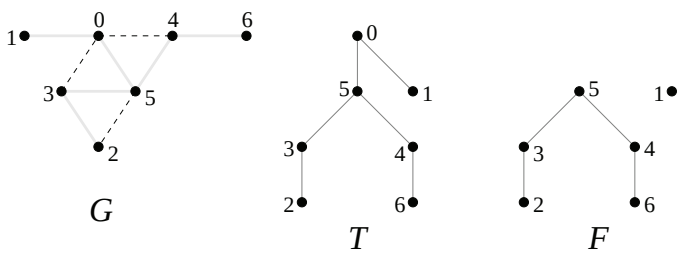
The third result, due to Kreweras [**22**], connects inversion enumerators with parking functions. Let PF_n denote the set of parking function of length n.

Theorem 6.22. *Let $n \geq 1$. Then*

$$t^{\binom{n}{2}} I_n(1/t) = \sum_{(a_1,\ldots,a_n)\in \mathrm{PF}_n} t^{a_1+\cdots+a_n-n}. \tag{52}$$

We now give proofs of Theorem 6.21, Corollary 6.13, and Theorem 6.22.

Proof of Theorem 6.21 (sketch). The following elegant proof is due to Gessel and Wang [**18**]. Let G be a connected graph on $[0,n]$. Start at vertex 0 and let T be the "depth-first spanning tree," i.e., move to the largest unvisited neighbor or else (if there is no unvisited neighbor) backtrack. The edges traversed when all vertices are visited are the edges of the spanning tree T. Remove the vertex 0 and root the trees that remain at the neighbors of 0. Denote this rooted forest by F_G.



Given a spanning forest F on $[n]$, what connected graphs G on $[0,n]$ satisfy $F = F_G$? The answer, whose straightforward verification we leave to the reader, is the following. Add the vertex 0 to F and connect it the roots of F, obtaining T. Clearly G consists of T with some added edges ij. The edge ij can be added to T if and only if the path from 0 to j contains i (or vice versa), and if i' is the next vertex after i on the path from i to j, then (j,i') is an inversion of F. Thus each inversion of F corresponds to a possible edge that can be added to T, and these edges can be added or not added independently. It follows that

$$\begin{aligned}\sum_{G\,:\,F=F_G} t^{e(G)} &= t^{e(T)}(1+t)^{\mathrm{inv}(F)}\\ &= t^n(1+t)^{\mathrm{inv}(F)}.\end{aligned}$$

Summing on all rooted forests F on $[n]$ gives

$$\begin{aligned}\sum_G t^{e(G)} &= t^n \sum_F (1+t)^{\mathrm{inv}(F)}\\ &= t^n I_n(1+t),\end{aligned}$$

where G ranges over all connected graphs on $[0,n]$. □

Proof of Corollary 6.13. By equation (50) and Theorem 6.21 we have

$$\sum_{n\geq 0} t^{n-1} I_{n-1}(1+t)\frac{x^n}{n!} = \log \sum_{n\geq 0}(1+t)^{\binom{n}{2}}\frac{x^n}{n!}.$$

Substituting $t-1$ for t gives

$$\sum_{n\geq 0}(t-1)^{n-1} I_{n-1}(t)\frac{x^n}{n!} = \log \sum_{n\geq 0} t^{\binom{n}{2}}\frac{x^n}{n!}.$$

Now differentiate both sides with respect to x to obtain equation (51). $\square$

Proof of Theorem 6.22. Let

$$\begin{aligned} J_n(t) &= \sum_{(a_1,\ldots,a_n)\in \mathrm{PF}_n} t^{\binom{n}{2}-\sum(a_i-1)} \\ &= \sum_{(a_1,\ldots,a_n)\in \mathrm{PF}_n} t^{\binom{n+1}{2}-\sum a_i}. \end{aligned}$$

Claim #1:

$$J_{n+1}(t) = \sum_{i=0}^{n} \binom{n}{i} (1+t+t^2+\cdots+t^i) J_i(t) J_{n-i}(t). \tag{53}$$

Proof of claim. Choose $0 \le i \le n$, and let S be an i-element subset of $[n]$. Choose also $\alpha \in \mathrm{PF}_i$, $\beta \in \mathrm{PF}_{n-i}$, and $0 \le j \le i$. Form a vector $\gamma = (\gamma_1, \ldots, \gamma_{n+1})$ by placing α at the positions indexed by S, placing $(\beta_1 + i + 1, \ldots, \beta_{n-i} + i + 1)$ at the positions indexed by $[n] - S$, and placing $j+1$ at position $n+1$. For instance, suppose $n = 7$, $i = 3$, $S = \{2,3,6\}$, $\alpha = (1,2,1)$, $\beta = (2,1,4,2)$, and $j = 1$. Then $\gamma = (6,1,2,5,8,1,6,2) \in \mathrm{PF}_8$. It is easy to check that in general $\gamma \in \mathrm{PF}_{n+1}$. Note that

$$\sum_{k=1}^{n+1} \gamma_k = \sum_{k=1}^{i} \alpha_k + \sum_{k=1}^{n-i} \beta_k + (n-i)(i+1) + j + 1,$$

so

$$\binom{n+2}{2} - \sum \gamma_k = \binom{i+1}{2} - \sum \alpha_k + \binom{n-i+1}{2} - \sum \beta_k + i - j.$$

Equation (53) then follows if the map $(i, S, \alpha, \beta, j) \mapsto \gamma$ is a bijection, i.e., given $\gamma \in \mathrm{PF}_{n+1}$, we can uniquely obtain (i, S, α, β, j) so that $(i, S, \alpha, \beta, j) \mapsto \gamma$. Now given γ, note that $i+1$ is the largest number that can replace γ_{n+1} so that we still have a parking function. Once i is determined, the rest of the argument is clear, proving the claim.

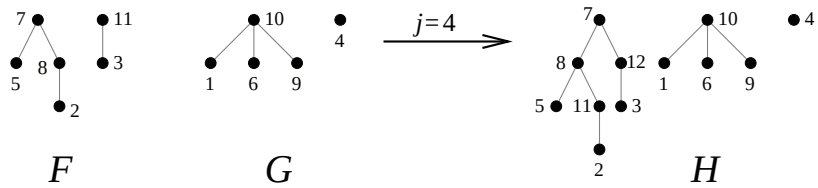
Note. Several bijections are known between the set of all rooted forests F on $[n]$ (or rooted trees on $[0,n]$) and the set PF_n of all parking functions $(a_1, \ldots, a_n)$ of length n, but none of them have the property that $\mathrm{inv}(F) = a_1 + \cdots + a_n - n$. Hence a direct bijective proof of Theorem 6.22 is not known. It would be interesting to find such a proof (Exercise 4).

Claim #2:

$$I_{n+1}(t) = \sum_{i=0}^{n} \binom{n}{i} (1+t+t^2+\cdots+t^i) I_i(t) I_{n-i}(t). \tag{54}$$

Proof of claim. We give a proof due to G. Kreweras [**22**]. Let F be a rooted forest on $S \subseteq [n]$, $\#S = i$, and let G be a rooted forest on $\bar{S} = [n] - S$. Let $u_1 < \cdots < u_i$ be the vertices of F, and set $u_{i+1} = n+1$. Choose $1 \le j \le i+1$. For all $m \ge j$ replace u_m by u_{m+1}. (If $j = i+1$, then do nothing.) This gives a labelled forest F' on $(S \cup \{n+1\}) - \{u_j\}$. Let T' be the labelled tree obtained from F' by adjoining the root u_j and connecting it to the roots of F'. Keep G the same. We obtain a rooted forest H on $[n+1]$ satisfying

$$\mathrm{inv}(H) = j - 1 + \mathrm{inv}(F) + \mathrm{inv}(G).$$



This process gives a bijection $(S, F, G, j) \mapsto H$, where $S \subseteq [n]$, F is a rooted forest on S, G is a rooted forest on $\bar{S}$, $1 \leq j \leq 1 + \#S$, and H is a rooted forest on $[n+1]$. Hence

$$\sum_{i=0}^{n} \sum_{\substack{S \subseteq [n] \\ \#S = i}} I_i(t) I_{n-i}(t)(1 + t + \cdots + t^i) = I_{n+1}(t),$$

and the claim follows.

The initial conditions $I_0(t) = J_0(t) = 1$ agree, so by the two claims we have $I_n(t) = J_n(t)$ for all $n \geq 0$. The proof of equation (52) follows by substituting $1/t$ for t. □

6.3. The Distance Enumerator of the Shi Arrangement

Recall that the Shi arrangement $\mathcal{S}_n$ is given by the defining polynomial

$$Q_{\mathcal{S}_n} = \prod_{1 \leq i < j \leq n} (x_i - x_j)(x_i - x_j - 1).$$

Let $K = \mathbb{R}$, and let R_0 denote the region

$$x_1 > x_2 > \cdots > x_n > x_1 - 1, \tag{55}$$

so $x \in R_0$ if and only if $0 \leq x_i - x_j \leq 1$ for all $i < j$. We define a labeling $\lambda : \mathcal{R}(\mathcal{S}_n) \to \mathbb{N}^n$ of the regions of $\mathcal{S}_n$ as follows.

- $\lambda(R_0) = (0, 0, \ldots, 0)$
- If the regions R and R' of $\mathcal{S}_n$ are separated by the single hyperplane H with the equation $x_i = x_j$, $i < j$, and if R and R_0 lie on the same side of H, then $\lambda(R') = \lambda(R) + e_j$ (exactly as for the braid arrangement).
- If the regions R and R' of $\mathcal{B}_n$ are separated by the single hyperplane H with the equation $x_i = x_j + 1$, $i < j$, and if R and R_0 lie on the same side of H, then $\lambda(R') = \lambda(R) + e_i$.

Note that the labeling λ is well-defined, since $\lambda(R)$ depends only on $\mathrm{sep}(R_0, R)$. Figure 5 shows the labeling λ for the case $n = 3$.

Theorem 6.23. *All labels $\lambda(R)$, $R \in \mathcal{R}(\mathcal{S}_n)$, are distinct, and*

$$\mathrm{PF}_n = \{(a_1 + 1, \ldots, a_n + 1) : (a_1, \ldots, a_n) = \lambda(R) \text{ for some } R \in \mathcal{R}(\mathcal{S}_n)\}.$$

In other words, the labels $\lambda(R)$ for $R \in \mathcal{R}(\mathcal{S}_n)$ are obtained from the labels $\lambda(R)$ for $R \in \mathcal{R}(\mathcal{B}_n)$ by permuting coordinates in all possible ways. This remarkable fact seems much more difficult to prove than the corresponding result for $\mathcal{B}_n$, viz., the labels $\lambda(R)$ for $\mathcal{B}_n$ consist of the sequences $(a_1, \ldots, a_n)$ with $0 \leq a_i \leq i - 1$ (an immediate consequence of Proposition 6.19 and Exercise 2.

Proof of Theorem 6.23 (sketch). An *antichain* $\mathfrak{I}$ of proper intervals of $[n]$ is a collection of intervals $[i, j] = \{i, i+1, \ldots, j\}$ with $1 \leq i < j \leq n$ such that if

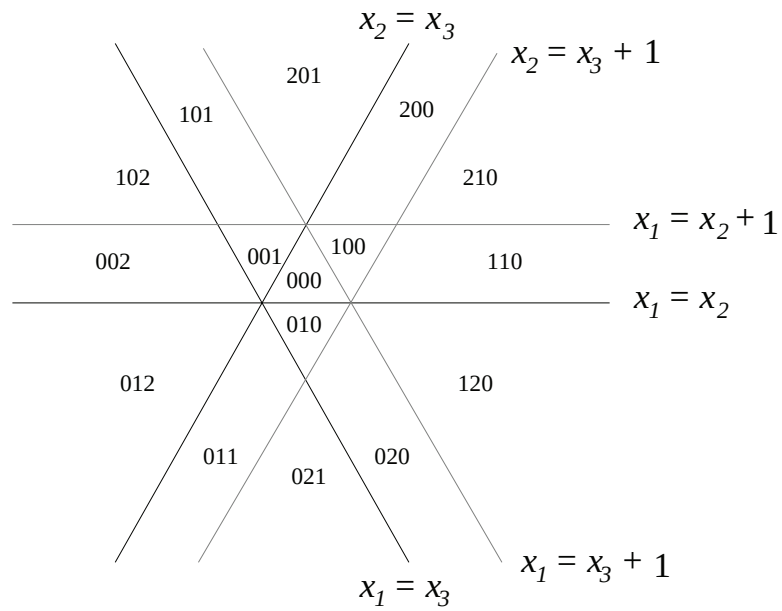


Figure 5. The labeling λ of the regions of $\mathcal{S}_3$

$I, I' \in \mathfrak{I}$ and $I \subseteq I'$, then $I = I'$. For instance, there are five antichains of proper intervals of [3], namely (writing ij for $[i,j]$)

$$\emptyset, \quad \{12\}, \quad \{23\}, \quad \{12,23\}, \quad \{13\}.$$

In general, the number of antichains of proper intervals of $[n]$ is the Catalan number C_n (immediate from [**32**, Exer. 6.19(bbb)], though this fact is not relevant here.

Every region $R \in \mathcal{R}(\mathcal{S}_n)$ corresponds bijectively to a pair $(w, \mathfrak{I})$, where $w \in \mathfrak{S}_n$ and $\mathfrak{I}$ is an antichain of proper intervals such that if $[i,j] \in \mathfrak{I}$ then $w(i) < w(j)$. Namely, the pair $(w, \mathfrak{I})$ corresponds to the region

$$x_{w(1)} > x_{w(2)} > \cdots > x_{w(n)}$$

$$x_{w(r)} - x_{w(s)} < 1 \text{ if } [r,s] \in \mathfrak{I}$$

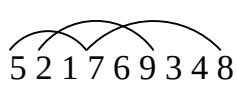
$$x_{w(r)} - x_{w(s)} > 1 \text{ if } r < s, w(r) < w(s), \text{ and } \nexists [i,j] \in \mathfrak{I} \text{ such that } i \le r < s \le j.$$

We call $(w, \mathfrak{I})$ a *valid pair*. Given a valid pair $(w, \mathfrak{I})$ corresponding to a region R, write $d(w, \mathfrak{I}) = d(R_0, R)$. It is easy to see that

$$\begin{aligned} d(w,\mathfrak{I}) &= \#\{(i,j) : i < j,\ w(i) > w(j)\} \\ &\quad + \#\{(i,j) : i < j,\ w(i) < w(j),\ \text{no } I \in \mathfrak{I} \text{ satisfies } i,j \in I\}. \end{aligned} \tag{56}$$

We say that the pair (i,j) is of *type 1* if $i < j$ and $w(i) > w(j)$, and is of *type 2* if $i < j$, $w(i) < w(j)$, and no $I \in \mathfrak{I}$ satisfies $i, j \in I$. Thus $d(w, \mathfrak{I})$ is the number of pairs (i,j) that are either of type 1 or type 2.

Example. Let $w = 521769348$ and $\mathfrak{I} = \{14, 27, 49\}$. We can represent the pair $(w, \mathfrak{I})$ by the diagram



This corresponds to the region

$$x_5 > x_2 > x_1 > x_7 > x_6 > x_9 > x_3 > x_4 > x_8$$

$$x_5 - x_7 < 1,\ x_2 - x_3 < 1,\ x_7 - x_8 < 1.$$

This region is separated from R_0 by the hyperplanes

$$x_5 = x_2, x_5 = x_1, \ldots \text{ (13 in all)}$$

$$x_5 = x_6 + 1, x_5 = x_9 + 1, \ldots \text{ (7 in all)}.$$

Let $\lambda(w, \mathfrak{I}, w(i))$ be the number of integers j such that (i, j) is either of type 1 or type 2. Thus

$$\lambda(R) = (\lambda(w, \mathfrak{I}, 1), \ldots, \lambda(w, \mathfrak{I}, n)).$$

For the example above we have $\lambda(R) = (2, 3, 0, 0, 7, 2, 3, 0, 3)$. For instance, the entry $\lambda(w, \mathfrak{I}, 5) = 7$ corresponds to the seven pairs 12, 13, 17, 18 (type 1) and 15, 16, 19 (type 2).

Clearly $\lambda(R) + (1, 1, \ldots, 1) \in \mathrm{PF}_n$, since $\lambda(w, \mathfrak{I}, w(i)) \leq n - i$ (the number of elements to the right of $w(i)$ in w).

Key lemma. Let X be an r-element subset of $[n]$, and let $v = v_1 \cdots v_r$ be a permutation of X. Let $\mathfrak{J}$ be an antichain of proper intervals $[a, b]$, where $v_a < v_b$. Suppose that the pair (i, j) is either of type 1 or type 2. Then

$$\lambda(v, \mathfrak{J}, v_i) > \lambda(v, \mathfrak{J}, v_j).$$

The proof of this lemma is straightforward and is left to the reader. For the example above, writing $\lambda(R) = (\lambda_1, \ldots, \lambda_9) = (5, 2, 1, 7, 6, 9, 3, 4, 8)$, the above lemma implies that

(a) $\lambda_5 > \lambda_2,\ \lambda_5 > \lambda_1,\ \lambda_5 > \lambda_3,\ \lambda_5 > \lambda_4,\ \lambda_2 > \lambda_1,\ \lambda_7 > \lambda_6,\ \lambda_7 > \lambda_3,\ \lambda_7 > \lambda_4,$
$\lambda_6 > \lambda_3,\ \lambda_6 > \lambda_4,\ \lambda_9 > \lambda_3,\ \lambda_9 > \lambda_4,\ \lambda_9 > \lambda_8$

(b) $\lambda_5 > \lambda_6,\ \lambda_5 > \lambda_9,\ \lambda_5 > \lambda_8,\ \lambda_2 > \lambda_4,\ \lambda_2 > \lambda_8,\ \lambda_1 > \lambda_4,\ \lambda_1 > \lambda_8.$

The crux of the proof of Theorem 6.23 is to show that given $\alpha + (1, 1, \ldots, 1) \in \mathrm{PF}_n$, there is a unique region $R \in \mathcal{R}(\mathcal{S}_n)$ satisfying $\lambda(R) = \alpha$. We will illustrate the construction of R from α with the example $\alpha = (2, 3, 0, 0, 7, 2, 3, 0, 3)$. We build up the pair $(w, \mathfrak{I})$ representing R one step at a time. First let v be the permutation of $[n]$ obtained from "standardizing" α from right-to-left. This means replacing the 0's in α with $1, 2, \ldots, m_1$ from right-to-left, then replacing the 1's in α with $m_1 + 1, m_1 + 2, \ldots, m_2$ from right-to-left, etc. Let $v^{-1} = (t_1, \ldots, t_n)$. For our example, we have

$$\begin{array}{ccccccccccc} \alpha & = & 2 & 3 & 0 & 0 & 7 & 2 & 3 & 0 & 3 \\ v & = & 5 & 8 & 3 & 2 & 9 & 4 & 7 & 1 & 6 \\ v^{-1} & = & 8 & 4 & 3 & 6 & 1 & 9 & 7 & 2 & 5 \end{array}.$$

Next we insert $t_1, \ldots, t_n$ from left-to-right into w. From α we can read off where t_i is inserted. After inserting t_i, we also record which of the positions of the elements so far inserted belong to some interval $I \in \mathfrak{I}$. We can also determine from α the unique way to do this. The best way to understand this insertion technique is to practice with some examples. Figure 6 illustrates the steps in the insertion process for our current example. These steps are explained as follows.

(1) First insert 8.

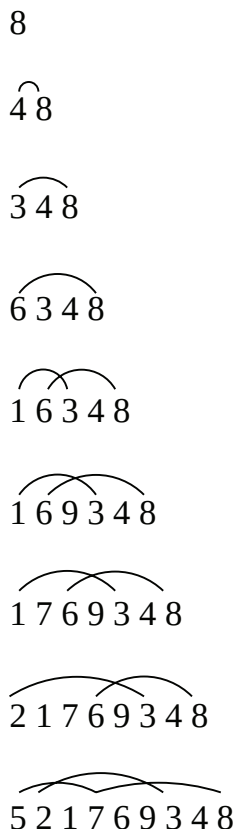


Figure 6. Constructing a valid pair $(w, \mathfrak{I})$ from the parking function $\alpha = (2, 3, 0, 0, 7, 2, 3, 0, 3)$

(2) Insert 4. Since $\alpha_8 = 0$, 4 appears to the left of 8, so we have the partial permutation 48. We now must decide whether the positions of 4 and 8 belong to some interval $I \in \mathfrak{I}$. (In other words, in the pictorial representation of $(w, \mathfrak{I})$, will 4 and 8 lie under some arc?) By the first term on the right-hand side of (56), we would have $\alpha_4 \geq 1$ if there were no such I. Since $\alpha_4 = 0$, we obtain the second row of Figure 6.

(3) Insert 3. As in the previous step, we obtain 348 with a single arc over all three terms.

(4) Insert 6. Suppose we inserted it after the 3, obtaining 3648, with a single arc over all four terms (since 3 and 8 have already been determined to lie under a single arc). We have $\alpha_6 = 2$, but the contribution so far (of 3648 with an arc over all four terms) to α_6 is 1. Thus later we must insert some j to the right of 6 so that the pair $(6, j)$ is of type 1 or type 2. By the lemma, we would have $\lambda(w, \mathfrak{I}, 6) > \lambda(w, \mathfrak{I}, j)$, contradicting that we are inserting elements in order of increasing α_i's. Similarly 3468 and 3486 are excluded, so 6 must be inserted at the left, yielding 6348. If the arc over 4,6,8 is not extended to 6, then we would have $\alpha_6 \geq 3$. Hence we obtain the fourth row of Figure 6.

(5) Insert 1. Using the lemma we obtain 16348. Since $\alpha_1 = 2$, there is an arc over 1 and two other elements to the right to 1. This gives the fifth row of Figure 6.
(6) Insert 9. Placing 9 before 1 or 6 yields $\alpha_9 \geq 4$, contradicting $\alpha_9 = 3$. Placing 9 after 3,4, or 8 is excluded by the lemma. Hence we get the sixth row of Figure 6.
(7) Insert 7. Placing 7 at the beginning yields four terms $j < 7$ appearing to the right of 7, giving $\alpha_7 \geq 4$, a contradiction. Placing 7 after 6,9,3,4,8 will violate the lemma, so we get the partial permutation 1769348. In order that $\alpha_7 = 3$, we must have 7 and 8 appearing under the same arc. Hence the arc from 6 to 8 must be extended to 7, yielding row seven of Figure 6.
(8) Insert 2 and 5. By now we hope it is clear that there is always a unique way to proceed.

The uniqueness of the above procedure shows that the map from the regions R of $\mathcal{S}_n$ (or the valid pairs $(w, \mathfrak{I})$ that index the regions) to parking functions α is injective. Since the number of valid pairs and number of parking functions are both $(n+1)^{n-1}$, the map is bijective, completing the (sketched) proof. In fact, it's not hard to show surjectivity directly, i.e., that the above procedure produces a valid pair $(w, \mathfrak{I})$ for *any* parking function, circumventing the need to know that $r(\mathcal{S}_n) = \#\mathrm{PF}_n$ in advance. □

Corollary 6.14. *The distance enumerator of $\mathcal{S}_n$ is given by*

$$D_{\mathcal{S}_n}(t) = \sum_{(a_1,\ldots,a_n)\in \mathrm{PF}_n} t^{a_1+\cdots+a_n-n}. \tag{57}$$

Proof. It is immediate from the definition of the labeling $\lambda : \mathcal{R}(\mathcal{S}_n) \to \mathbb{N}^n$ that if $\lambda(R) = (a_1, \ldots, a_n)$, then $d(R_0, R) = a_1 + \cdots + a_n$. Now use Theorem 6.23. □

Note. An alternative proof of Corollary 6.14 is given by Athanasiadis [**3**].

6.4. The Distance Enumerator of a Supersolvable Arrangement

The goal of this section is a formula for the distance enumerator of a supersolvable (central) arrangement with respect to a "canonical" base region R_0. The proof will be by induction, based on the following lemma of Björner, Edelman, and Ziegler [**8**].

Lemma 6.7. *Every central arrangement of rank 2 is supersolvable. A central arrangement $\mathcal{A}$ of rank $d \geq 3$ is supersolvable if and only if $\mathcal{A} = \mathcal{A}_0 \uplus \mathcal{A}_1$ (disjoint union), where $\mathcal{A}_0$ is supersolvable of rank $d-1$ (so $\mathcal{A}_1 \neq \emptyset$) and for all $H', H'' \in \mathcal{A}_1$ with $H' \neq H''$, there exists $H \in \mathcal{A}_0$ such that $H' \cap H'' \subseteq H$.*

Proof. Every geometric lattice of rank 2 is modular, hence supersolvable, so let $\mathcal{A}$ be supersolvable of rank $d \geq 3$. Let $\hat{0} = x_0 \lessdot x_1 \lessdot \cdots \lessdot x_{d-1} \lessdot x_d = \hat{1}$ be a modular maximal chain in $L_{\mathcal{A}}$. Define

$$\mathcal{A}_0 = \mathcal{A}_{x_{d-1}} = \{H \in \mathcal{A} \,:\, x_{d-1} \subseteq H\},$$

so $L(\mathcal{A}_0) \cong [\hat{0}, x_{d+1}]$. Clearly $\mathcal{A}_0$ is supersolvable of rank $d-1$. Let $\mathcal{A}_1 = \mathcal{A} - \mathcal{A}_0$. Let $H', H'' \in \mathcal{A}_1$, $H' \neq H''$. Since $x_{d-1} \not\subseteq H'$ we have $x_{d-1} \vee (H' \vee H'') = \hat{1}$ in $L(\mathcal{A})$. Now $\mathrm{rk}(x_{d-1}) = d-1$, and $\mathrm{rk}(H' \vee H'') = 2$ by semimodularity. Since x_{d-1} is modular we obtain

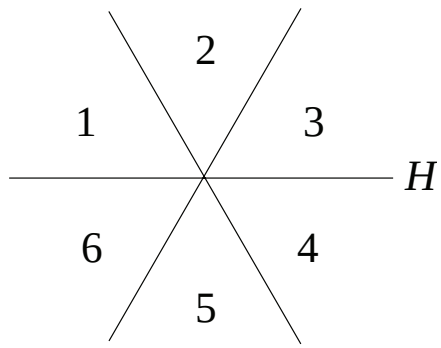
$$\mathrm{rk}(x_{d-1} \wedge (H' \vee H'')) = (d-1) + 2 - d = 1,$$

i.e., $x_{d-1} \wedge (H' \vee H'') = H \in \mathcal{A}$. Since $H \leq x_{d-1}$ it follows that $H \in \mathcal{A}_0$. Moreover, $H' \cap H'' \subseteq H$ since $H \leq H' \vee H''$. This proves the "only if" part of the lemma. The "if" part is straightforward and not needed here, so we omit the proof. $\square$

Given $\mathcal{A}_0 = \mathcal{A}_{x_{d-1}}$ as above, define a map $\pi : \mathcal{R}(\mathcal{A}) \twoheadrightarrow \mathcal{R}(\mathcal{A}_0)$ (the symbol $\twoheadrightarrow$ denotes surjectivity) by $\pi(R) = R'$ if $R \subseteq R'$. For $R \in \mathcal{R}(\mathcal{A})$ let

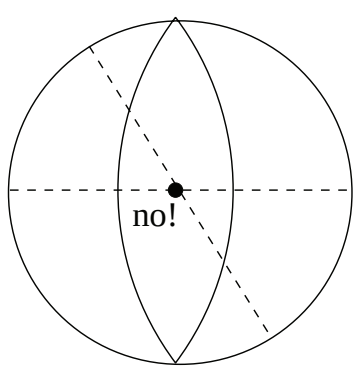
$$\mathcal{F}(R) = \{R_1 \in \mathcal{R}(\mathcal{A}) \,:\, \pi(R) = \pi(R_1)\} = \pi^{-1}(\pi(R)).$$

For example, let $\mathcal{A}$ be the arrangement

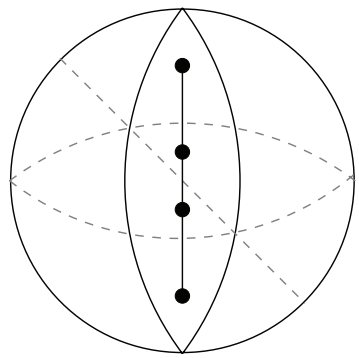


Let $\mathcal{A}_0 = \{H\}$. Then $\mathcal{F}(1) = \{1, 2, 3\}$ and $\mathcal{F}(5) = \{4, 5, 6\}$.

Now let $R' \in \mathcal{R}(\mathcal{A}_0)$. By Lemma 6.7 no $H', H'' \in \mathcal{A}$ can intersect inside R'. The illustration below is a projective diagram of a bad intersection. The solid lines define $\mathcal{A}_0$ and the dashed lines $\mathcal{A}_1$.



Thus $\pi^{-1}(R')$ must be arranged "linearly" in R', i.e., there is a straight line intersecting all $R \in \pi^{-1}(R')$.



Since $\mathrm{rank}(\mathcal{A}) > \mathrm{rank}(\mathcal{A}_0)$, we have $\#\pi^{-1}(R') > 1$ (for $H \in \mathcal{A}$ does not bisect R' if and only if $\mathrm{rank}(\mathcal{A}_0 \cup H) = \mathrm{rank}(\mathcal{A}_0)$). Thus there are two distinct regions $R_1, R_2 \in \pi^{-1}(R')$ that are endpoints of the "chain of regions."

Let e_d have the meaning of equation (34), i.e.,

$$e_d = \#\{H \in \mathcal{A} : H \notin \mathcal{A}_0\} = \#\mathcal{A}_1.$$

Then $\pi^{-1}(R')$ is a chain of regions of length e_d, so $\#\pi^{-1}(R') = 1 + e_d$. We now come to the key definition of this section. The definition is recursive by rank, the base case being rank at most 2.

Definition 6.16. Let $\mathcal{A}$ be a real supersolvable central arrangement of rank d, and let $\mathcal{A}_0$ be a supersolvable subarrangement of rank $d-1$ (which always exists by the definition of supersolvability). A region $R_0 \in \mathcal{R}(\mathcal{A})$ is called *canonical* if either (1) $d \leq 2$, or else (2) $d \geq 3$, $\pi(R_0) \in \mathcal{R}(\mathcal{A}_0)$ is canonical, and R_0 is an endpoint of the chain $\mathcal{F}(R_0)$.

Since every chain has two endpoints and a central arrangement of rank 1 has two (canonical) regions, it follows that there are at least 2^d canonical regions.

The main result on distance enumerators of supersolvable arrangements is the following, due to Björner, Edelman, and Ziegler [**8**, Thm. 6.11].

Theorem 6.24. *Let $\mathcal{A}$ be a supersolvable central arrangement of rank d in $\mathbb{R}^n$. Let $R_0 \in \mathcal{R}(\mathcal{A})$ be canonical, and suppose that*

$$\chi_{\mathcal{A}}(t) = (t - e_1)(t - e_2) \cdots (t - e_d)t^{n-d}.$$

(There always exist such positive integers e_i by Corollary 4.9.) Then

$$D_{\mathcal{A},R_0}(t) = \prod_{i=1}^{d}(1 + t + t^2 + \cdots + t^{e_i}).$$

Proof. Let $W_{\mathcal{A}}$ be the weak order on $\mathcal{A}$ with respect to R_0, i.e.,

$$W_{\mathcal{A}} = \{\mathrm{sep}(R_0, R) : R \in \mathcal{R}(\mathcal{A})\},$$

ordered by inclusion. Thus $W_{\mathcal{A}}$ is graded with rank function given by $\mathrm{rk}(R) = d(R_0, R)$ and rank generating function

$$\sum_{R \in W_{\mathcal{A}}} t^{\mathrm{rk}(R)} = D_{\mathcal{A}}(t).$$

Since R_0 is canonical, for all $R' \in \mathcal{R}(\mathcal{A}_0)$ we have that $\pi^{-1}(R')$ is a chain of length e_d. Hence if $R \in \mathcal{R}(\mathcal{A})$ and $h(R)$ denotes the rank of R in the chain $\mathcal{F}(R)$, then

$$d_{\mathcal{A}}(R_0, R) = d_{\mathcal{A}_0}(\pi(R)) + h(R).$$

Therefore

$$D_{\mathcal{A}}(t) = D_{\mathcal{A}_0}(t)(1 + t + \cdots + t^{e_d}),$$

and the proof follows by induction. □

NOTE. The following two results were also proved in [**8**]. We simply state them here without proof.

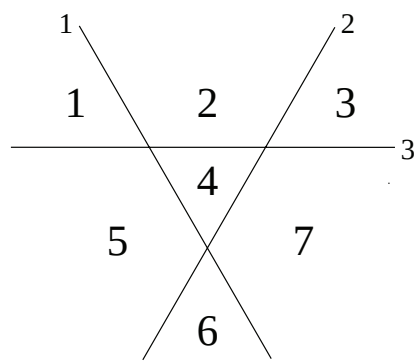
- If $\mathcal{A}$ is a real supersolvable central arrangement and R_0 is canonical, then $W_{\mathcal{A}}$ is a lattice (Exercise 7).
- If $\mathcal{A}$ is any real central arrangement and $W_{\mathcal{A}}$ is a lattice, then R_0 is simplicial (bounded by exactly $\mathrm{rk}(\mathcal{A})$ hyperplanes, the minimum possible). In other words, the closure $\bar{R}_0$ is a simplex. As a partial converse, if *every* region R is simplicial, then $W_{\mathcal{A}}$ is a lattice (Exercise 8).

6.5. The Varchenko Matrix

Let $\mathcal{A}$ be a real arrangement. For each $H \in \mathcal{A}$ let a_H be an indeterminate. Define a matrix $V = V(\mathcal{A})$ with rows and columns indexed by $\mathcal{R}(\mathcal{A})$ by

$$V_{RR'} = \prod_{H \in \mathrm{sep}(R,R')} a_H.$$

For instance, let $\mathcal{A}$ be given as follows:



Then

$$V = \begin{array}{c|ccccccc} & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ \hline 1 & 1 & a_1 & a_1a_2 & a_1a_3 & a_3 & a_2a_3 & a_1a_2a_3 \\ 2 & a_1 & 1 & a_2 & a_3 & a_1a_3 & a_1a_2a_3 & a_2a_3 \\ 3 & a_1a_2 & a_2 & 1 & a_2a_3 & a_1a_2a_3 & a_1a_3 & a_3 \\ 4 & a_1a_3 & a_3 & a_2a_3 & 1 & a_1 & a_1a_2 & a_2 \\ 5 & a_3 & a_1a_3 & a_1a_2a_3 & a_1 & 1 & a_2 & a_1a_2 \\ 6 & a_2a_3 & a_aa_2a_3 & a_1a_3 & a_1a_2 & a_2 & 1 & a_1 \\ 7 & a_1a_2a_3 & a_1a_3 & a_3 & a_2 & a_1a_2 & a_1 & 1 \end{array}$$

The determinant of this matrix happens to be given by

$$\det(V) = \left(1 - a_1^2\right)^3 \left(1 - a_2^2\right)^3 \left(1 - a_3^2\right)^3.$$

In order to state the general result, define for $x \in L(\mathcal{A})$,

$$\begin{aligned} a_x &= \prod_{H \supseteq x} a_H \\ n(x) &= r(\mathcal{A}^x) \\ p(x) &= b(c^{-1}\mathcal{A}_x) = \beta(\mathcal{A}_x), \end{aligned}$$

where as usual $\mathcal{A}^x = \{x \cap H \neq \emptyset : x \not\subseteq H\}$ and $\mathcal{A}_x = \{H \in \mathcal{A} : H \supseteq x\}$, and where c^{-1} denotes deconing and β is defined in Exercise 4.22. Thus

$$n(x) = |\chi_{\mathcal{A}^x}(-1)| = \sum_{y \geq x} |\mu(x,y)|$$

$$p(x) = \left|\chi'_{\mathcal{A}_x}(1)\right|.$$

Example 6.14. The arrangement of three lines illustrated above has two types of intersections (other than $\hat{0}$): a line x and a point y. For a line x, $\mathcal{A}^x$ consists of two points on a line, so $n(x) = r(\mathcal{A}^x) = 3$. Moreover, $\mathcal{A}_x$ consists of the single hyperplane x in $\mathbb{R}^2$, so $c^{-1}\mathcal{A}_x = \emptyset$ and $p(x) = b(\emptyset) = 1$. Hence we obtain the factor $(1-a_x)^3$ in the determinant. On the other hand, $\mathcal{A}^y = \emptyset$ so $n(y) = r(\emptyset) = 1$. Moreover, $\mathcal{A}_y$ consists of two intersecting lines in $\mathbb{R}^2$, with characteristic polynomial $\chi_{\mathcal{A}_y}(t) = (t-1)^2$. Hence $p(y) = |\chi'_{\mathcal{A}_y}(1)| = 0$. Equivalently, $c^{-1}\mathcal{A}_y$ consists of a single point on a line, so again $p(y) = b(c^{-1}\mathcal{A}_y) = 0$. Thus y contributes a factor $(1-a_y^2)^0 = 1$ to $\det(V)$.

We can now state the remarkable result of Varchenko [**37**], generalized to "weighted matroids" by Brylawski and Varchenko [**11**].

Theorem 6.25. *Let $\mathcal{A}$ be a real arrangement. Then*

$$\det V(\mathcal{A}) = \prod_{\hat{0} \neq x \in L(\mathcal{A})} (1-a_x^2)^{n(x)p(x)}.$$

Proof. Omitted.

Exercises

(1) Let $\mathcal{A}$ be a central arrangement in $\mathbb{R}^n$ with distance enumerator $D_{\mathcal{A}}(t)$ (with respect to some base region R_0). Define a graph $G_{\mathcal{A}}$ on the vertex set $\mathcal{R}(\mathcal{A})$ by putting an edge between R and R' if $\#\mathrm{sep}(R,R') = 1$ (i.e., R and R' are separated by a unique hyperplane).
   (a) [2–] Show that $G_{\mathcal{A}}$ is a bipartite graph.
   (b) [2] Show that if $\#\mathcal{A}$ is odd, then $D_{\mathcal{A}}(-1) = 0$.
   (c) [2] Show that if $\#\mathcal{A}$ is even and $r(\mathcal{A}) \equiv 2 \pmod 4$, then $D_{\mathcal{A}}(-1) \equiv 2 \pmod 4$ (so $D_{\mathcal{A}}(-1) \neq 0$).
   (d) [2] Give an example of (c), i.e., find $\mathcal{A}$ so that $\#\mathcal{A}$ is even and $r(\mathcal{A}) \equiv 2 \pmod 4$.
   (e) [2] Show that (c) cannot hold if $\mathcal{A}$ is supersolvable. (It is *not* assumed that the base region R_0 is canonical. Try to avoid the use of Section 6.0.4.)
   (f) [2+] Show that if $\#\mathcal{A}$ is even and $r(\mathcal{A}) \equiv 0 \pmod 4$, then it is possible for $D_{\mathcal{A}}(-1) = 0$ and for $D_{\mathcal{A}}(-1) \neq 0$. Can examples be found for $\mathrm{rank}(\mathcal{A}) \leq 3$?

(2) [2–] Show that a sequence $(c_1, \ldots, c_n) \in \mathbb{N}^n$ is the inversion sequence of a permutation $w \in \mathfrak{S}_n$ if and only if $c_i \leq i-1$ for $1 \leq i \leq n$.

(3) [2] Show that all cars can park under the scenario following Definition 6.15 if and only if the sequence $(a_1, \ldots, a_n)$ of preferred parking spaces is a parking function.

(4) [5] Find a bijective proof of Theorem 6.22, i.e., find a bijection φ between the set of all rooted forests on $[n]$ and the set PF_n of all parking functions of length n satisfying $\mathrm{inv}(F) = \binom{n+1}{2} - a_1 - \cdots - a_n$ when $\varphi(F) = (a_1, \ldots, a_n)$. NOTE. In principle a bijection φ can be obtained by carefully analyzing the proof of Theorem 6.22. However, this bijection will be of a messy recursive nature. A "nonrecursive" bijection would be greatly preferred.

(5) [3] There is a natural two-variable refinement of the distance enumerator (57) of $\mathcal{S}_n$. Given $R \in \mathcal{R}(\mathcal{S}_n)$, define $d_0(R_0, R)$ to be the number of hyperplanes $x_i = x_j$ separating R_0 from R, and $d_1(R_0, R)$ to be the number of hyperplanes $x_i = x_j + 1$ separating R_0 from R. (Here R_0 is given by (55) as usual.) Set

$$D_n(q,t) = \sum_{R \in \mathcal{R}(\mathcal{S}_n)} q^{d_0(R_0,R)} t^{d_1(R_0,R)}.$$

What can be said about the polynomial $D_n(q,t)$? Can its coefficients be interpreted in a simple way in terms of tree or forest inversions? Are there formulas or recurrences for $D_n(q,t)$ generalizing Theorem 6.21, Corollary 6.13, or equation (53)? The table below give the coefficients of $q^i t^j$ in $D_n(q,t)$ for $2 \le n \le 4$.

$t \backslash q$	0	1
0	1	1
1	1	

$t \backslash q$	0	1	2	3
0	1	1	2	1
1	2	2	2	
2	2	2		
3	1			

$t \backslash q$	0	1	2	3	4	5	6
0	1	1	2	3	3	3	1
1	3	3	6	7	6	3	
2	5	5	8	9	5		
3	6	7	9	6			
4	5	6	5				
5	3	3					
6	1						

(6) [5–] Let $\mathcal{G}_n$ denote the generic braid arrangement

$$x_i - x_j = a_{ij}, \quad 1 \le i < j \le n,$$

in $\mathbb{R}^n$. Can anything interesting be said about the distance enumerator $D_{\mathcal{G}_n}(t)$ (which depends on the choice of base region R_0 and possibly on the a_{ij}'s)? Generalize if possible to generic graphical arrangments, especially for supersolvable (or chordal) graphs.

(7) [3–] Let $\mathcal{A}$ be a real supersolvable arrangement and R_0 a canonical region of $\mathcal{A}$. Show that the weak order $W_{\mathcal{A}}$ (with respect to R_0) is a lattice.

(8) (a) [2+] let $\mathcal{A}$ be a real central arrangement of rank d. Suppose that the weak order $W_{\mathcal{A}}$ (with respect to some region $R_0 \in \mathcal{R}(\mathcal{A})$) is a lattice. Show that R_0 is simplicial, i.e., bounded by exactly d hyperplanes.

(b) [3–] Let $\mathcal{A}$ be a real central arrangement. Show that if every region $R \in \mathcal{R}(\mathcal{A})$ is simplicial, then $W_{\mathcal{A}}$ is a lattice.

(9) (a) [2] Set each $a_H = q$ in the Varchenko matrix V of an arrangement R in $\mathbb{R}^n$, obtaining a matrix $V(q)$. Let $r = r(\mathcal{A})$. The entries of $V(q)$ belong to the principal ideal domain $\mathbb{Q}[q]$, so $V(q)$ has a Smith normal form $AV(q)B = \mathrm{diag}(p_1, \ldots, p_r)$, where A, B are $r \times r$ matrices whose entries belong to $\mathbb{Q}[q]$ and whose determinants are nonzero elements of $\mathbb{Q}$, and where $p_1, \ldots, p_r \in$

$\mathbb{Q}[q]$ such that $p_i \mid p_{i+1}$ for $1 \le i \le r-1$. The Smith normal form is unique up to multiplication of the p_i's by nonzero elements of $\mathbb{Q}$. For instance, if $\mathcal{A} = \mathcal{B}_3$, then

$$AV(q)B = \operatorname{diag}(1, q^2-1, q^2-1, q^2-1, (q^2-1)^2, (q^2-1)^2(q^4+q^2+1)).$$

Show that each p_i is a polynomial in q^2.

(b) [3+] Let a_i be the number of j's for which $(q^2-1)^i \mid p_j$ but $(q^2-1)^{i+1} \nmid p_j$. Show that

$$\chi_{\mathcal{A}}(t) = \sum_{i \ge 0} (-1)^{n-i} a_i q^{n-i}.$$

(c) [5] What more can be said about the polynomials p_i? By Theorem 6.25 they are products of cyclotomic polynomials, so one could begin by asking for the largest powers of q^2+1 or q^4+q^2+1 dividing each p_i.

BIBLIOGRAPHY

1. C. A. Athanasiadis, *Algebraic Combinatorics of Graph Spectra, Subspace Arrangements and Tutte Polynomials*, Ph.D. thesis, MIT, 1996.
2. C. A. Athanasiadis, Characteristic polynomials of subspace arrangements and finite fields, *Advances in Math.* **122** (1996), 193–233.
3. C. A. Athanasiadis, A class of labeled posets and the Shi arrangement of hyperplanes, *J. Combinatorial Theory (A)* **80** (1997), 158–162.
4. H. Barcelo and E. Ihrig, Lattices of parabolic subgroups in connection with hyperplane arrangements, *J. Algebraic Combinatorics* **9** (1999), 5–24.
5. A. Björner, Orderings of Coxeter groups, in Combinatorics and algebra (Boulder, Colo., 1983), *Contemp. Math.* **34**, American Mathematical Society, Providence, RI, 1984, pp. 175–195.
6. A. Björner and F. Brenti, *Combinatorics of Coxeter Groups*, Springer-Verlag, New York, 2005.
7. A. Björner, M. Las Vergnas, B. Sturmfels, N. White, and G. Ziegler, *Oriented Matroids*, second ed., Encyclopedia of Mathematics and Its Applications **46**, Cambridge University Press, Cambridge, 1999.
8. A. Björner, P. Edelman, and G. Ziegler, Hyperplane arrangements with a lattice of regions, *Discrete Comput. Geom.* **5** (1990), 263–288.
9. A. Blass and B. Sagan, Characteristic and Ehrhart polynomials, *J. Algebraic Combinatorics* **7** (1998), 115–126.
10. N. Bourbaki, *Groupes et algèbres de Lie*, Éléments de Mathématique, Fasc. XXXIV, Hermann, Paris, 1968.
11. T. Brylawski and A. Varchenko, The determinant formula for a matroid bilinear form, *Advances in Math.* **129** (1997), 1–24.
12. J. L. Chandon, J. Lemaire, and J. Pouget, Dénombrement des quasi-ordres sur un ensemble fini, *Math. Inform. Sci. Humaines* **62** (1978), 61–80, 83.
13. H. Crapo and G.-C. Rota, *On the Foundations of Combinatorial Theory: Combinatorial Geometries*, preliminary edition, MIT Press, Cambridge, MA, 1970.
14. P. Edelman and C. Greene, Balanced tableaux, *Advances in Math.* **63** (1987), 42–99.
15. P. C. Fishburn, *Interval Orders and Interval Graphs*, Wiley-Interscience, New York, 1985.
16. D. Foata and J. Riordan, Mappings of acyclic and parking functions, *Aequationes Math.* **10** (1974), 10–22.

17. A. Garsia, The saga of reduced factorizations of elements of the symmetric group, *Publ. LACIM* **29**, Univerité du Québec à Montréal, 2002.
18. I. Gessel and D. L. Wang, Depth-first search as a combinatorial correspondence, *J. Combinatorial Theory (A)* **26** (1979), 308–313.
19. C. Greene, On the Möbius algebra of a partially ordered set, *Advances in Math.* **10** (1973), 177–187.
20. J. E. Humphreys, *Reflection Groups and Coxeter Groups*, Cambridge University Press, Cambridge, 1990.
21. A. G. Konheim and B. Weiss, An occupancy discipline and applications, *SIAM J. Applied Math.* **14** (1966), 1266–1274.
22. G. Kreweras, Une famille de polynômes ayant plusieurs propriétés énumerative, *Per. Math. Hung.* **11** (1980), 309–320.
23. C. L. Mallows and J. Riordan, The inversion enumerator for labeled trees, *Bull. Amer. Math. Soc.* **74**, (1968), 92–94.
24. P. Orlik and H. Terao, *Arrangements of Hyperplanes*, Springer-Verlag, Berlin, 1992.
25. A. Postnikov and R. Stanley, Deformations of Coxeter hyperplane arrangements, *J. Combinatorial Theory (A)* **91** (2000), 544–597.
26. J. Riordan, Ballots and trees, *J. Combinatorial Theory* **6** (1969), 408–411
27. J.-Y. Shi, The Kazhdan-Lusztig cells in certain affine Weyl groups, *Lecture Notes in Mathematics*, vol. 1179, Springer-Verlag, Berlin, 1986.
28. R. Stanley, Modular elements of geometric lattices, *Algebra Universalis* **1** (1971), 214–217.
29. R. Stanley, Supersolvable lattices, *Algebra Universalis* **2** (1972), 197–217.
30. R. Stanley, On the number of reduced decompositions of elements of Coxeter groups, *European J. Combinatorics* **5** (1984), 359–372.
31. R. Stanley, *Enumerative Combinatorics*, vol. 1, Wadsworth and Brooks/Cole, Pacific Grove, CA, 1986; second printing, Cambridge University Press, Cambridge, 1996.
32. R. Stanley, *Enumerative Combinatorics*, vol. 2, Cambridge University Press, Cambridge, 1999
33. R. Stanley, Hyperplane arrangements, interval orders, and trees, *Proc. Nat. Acad. Sci.* **93** (1996), 2620–2625.
34. H. Terao, Free arrangements of hyperplanes and unitary reflection groups, *Proc. Japan Acac., Ser. A* **56** (1980), 389–392.
35. H. Terao, Generalized exponents of a free arrangement of hyperplanes and Shepherd[sic]-Todd-Brieskorn formula, *Invent. math.* **63** (1981), 159–179.
36. W. T. Trotter, *Combinatorics and Partially Ordered Sets*, The Johns Hopkins Univ. Press, Baltimore/London, 1992.
37. A. Varchenko, Bilinear form of real configuration of hyperplanes, *Advances in Math.* **97** (1993), 110–144.
38. R. L. Wine and J. E. Freund, On the enumeration of decision patterns involving n means, *Ann. Math. Stat.* **28**, 256–259.

Poset Topology: Tools and Applications

Michelle L. Wachs

IAS/Park City Mathematics Series
Volume 14, 2004

Poset Topology: Tools and Applications

Michelle L. Wachs

Introduction

The theory of poset topology evolved from the seminal 1964 paper of Gian-Carlo Rota on the Möbius function of a partially ordered set. This theory provides a deep and fundamental link between combinatorics and other branches of mathematics. Early impetus for this theory came from diverse fields such as

- commutative algebra (Stanley's 1975 proof of the upper bound conjecture)
- group theory (the work of Brown (1974) and Quillen (1978) on p-subgroup posets)
- combinatorics (Björner's 1980 paper on poset shellability)
- representation theory (Stanley's 1982 paper on group actions on the homology of posets)
- topology (the Orlik-Solomon theory of hyperplane arrangements (1980))
- complexity theory (the 1984 paper of Kahn, Saks, and Sturtevant on the evasiveness conjecture).

Later developments have kept the theory vital. I mention just a few examples: Goresky-MacPherson formula for subspace arrangements, Björner-Lovász-Yao complexity theory results, Björner-Wachs extension of shellability to nonpure complexes, Forman's discrete version of Morse theory, and Vassiliev's work on knot invariants and graph connectivity.

So, what is poset topology? By the topology of a partially ordered set (poset) we mean the topology of a certain simplicial complex associated with the poset, called the order complex of the poset. In these lectures I will present some of the techniques that have been developed over the years to study the topology of a poset, and discuss some of the applications of poset topology to the fields mentioned above as well as to other fields. In particular, I will discuss tools for computing homotopy type and (co)homology of posets, with an emphasis on group equivariant (co)homology. Although posets and simplicial complexes can be viewed as essentially the same topological object, we will narrow our focus, for the most part,

[1]Department of Mathematics, University of Miami, Coral Gables, Fl 33124.
E-mail address: `wachs@math.miami.edu`.
This work was partially supported by NSF grant DMS 0302310.

©2007 Michelle L. Wachs

to tools that were developed specifically for posets; for example, lexicographical shellability, recursive atom orderings, Whitney homology techniques, (co)homology bases/generating set techniques, and fiber theorems.

Research in poset topology is very much driven by the study of concrete examples that arise in various contexts both inside and outside of combinatorics. These examples often turn out to have a rich and interesting topological structure, whose analysis leads to the development of new techniques in poset topology. These lecture notes are organized according to techniques rather than applications. A recurring theme is the use of original examples in demonstrating a technique, where by original example I mean the example that led to the development of the technique in the first place. More recent examples will be discussed as well.

With regard to the choice of topics, I was primarily motivated by my own research interests and the desire to provide the students at the PCMI graduate school with concrete skills in this subject. Due to space and time constraints and my decision to focus on techniques specific to posets, there are a number of very important tools for general simplicial complexes that I have only been able to mention in passing (or not at all). I point out, in particular, discrete Morse theory (which is a major part of the lecture series of Robin Forman, its originator) and basic techniques from algebraic topology such as long exact sequences and spectral sequences. For further techniques and applications, still of current interest, we strongly recommend the influential 1995 book chapter of Anders Björner [**29**].

The exercises vary in difficulty and are there to reinforce and supplement the material treated in these notes. There are many open problems (simply referred to as problems) and conjectures sprinkled throughout the text.

I would like to thank the organizers (Ezra Miller, Vic Reiner and Bernd Sturmfels) of the 2004 PCMI Graduate Summer School for inviting me to deliver these lectures. I am very grateful to Vic Reiner for his encouragement and support. I would also like to thank Tricia Hersh for the help and support she provided as my overqualified teaching assistant. Finally, I would like to express my gratitude to the graduate students at the summer school for their interest and inspiration.

LECTURE 1
Basic Definitions, Results, and Examples

1.1. Order Complexes and Face Posets

We begin by defining the order complex of a poset and the face poset of a simplicial complex. These constructions enable us to view posets and simplicial complexes as essentially the same topological object. We shall assume throughout these lectures that all posets and simplicial complexes are finite, unless otherwise stated.

An *abstract simplicial complex* Δ on finite vertex set V is a nonempty collection of subsets of V such that

- $\{v\} \in \Delta$ for all $v \in V$
- if $G \in \Delta$ and $F \subseteq G$ then $F \in \Delta$.

The elements of Δ are called *faces* (or simplices) of Δ and the maximal faces are called *facets*. We say that a face F has dimension d and write $\dim F = d$ if $d = |F| - 1$. Faces of dimension d are referred to as d-faces. The *dimension* $\dim \Delta$ of Δ is defined to be $\max_{F \in \Delta} \dim F$. We also allow the (-1)-dimensional complex $\{\emptyset\}$, which we refer to as the *empty simplicial complex*. It will be convenient to refer to the empty set $\emptyset$, as the *degenerate empty complex* and say that it has dimension -2, even though we don't really consider it to be a simplicial complex. If all facets of Δ have the same dimension then Δ is said to be *pure*.

A d-dimensional *geometric simplex* in $\mathbb{R}^n$ is defined to be the convex hull of $d+1$ affinely independent points in $\mathbb{R}^n$ called *vertices*. The convex hull of any subset of the vertices is called a *face* of the geometric simplex. A *geometric simplicial complex* K in $\mathbb{R}^n$ is a nonempty collection of geometric simplices in $\mathbb{R}^n$ such that

- Every face of a simplex in K is in K.
- The intersection of any two simplices of K is a face of both of them.

From a geometric simplicial complex K, one gets an abstract simplicial complex $\Delta(K)$ by letting the faces of $\Delta(K)$ be the vertex sets of the simplices of K. Every abstract simplicial complex Δ can be obtained in this way, i.e., there is a geometric simplicial complex K such that $\Delta(K) = \Delta$. Although K is not unique, the underlying topological space, obtained by taking the union of the simplices of K under the usual topology on $\mathbb{R}^n$, is unique up to homeomorphism. We refer to this space as the *geometric realization* of Δ and denote it by $\|\Delta\|$. We will usually

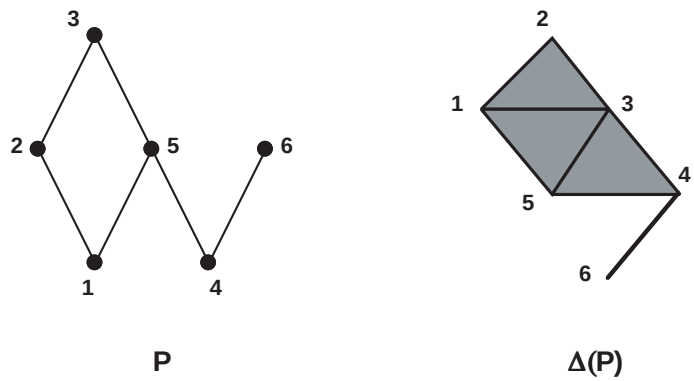


Figure 1.1.1. Order complex of a poset

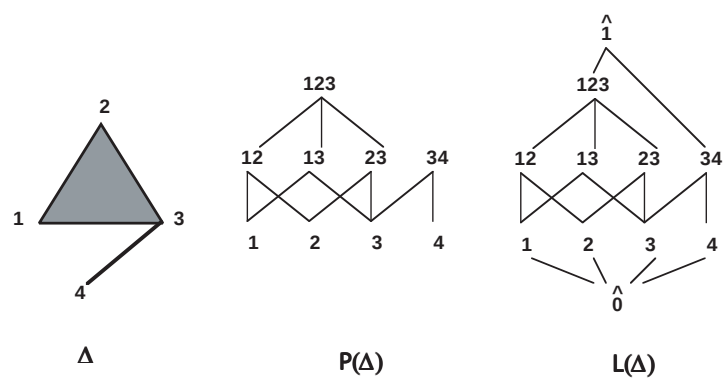


Figure 1.1.2. Face poset and face lattice of a simplicial complex

drop the $\| \ \|$ and let Δ denote an abstract simplicial complex as well as its geometric realization.

To every poset P, one can associate an abstract simplicial complex $\Delta(P)$ called the *order complex* of P. The vertices of $\Delta(P)$ are the elements of P and the faces of $\Delta(P)$ are the chains (i.e., totally ordered subsets) of P. (The order complex of the empty poset is the empty simplicial complex $\{\emptyset\}$.) For example, the Hasse diagram of a poset P and the geometric realization of its order complex are given in Figure 1.1.1.

To every simplicial complex Δ, one can associate a poset $P(\Delta)$ called the *face poset* of Δ, which is defined to be the poset of nonempty faces ordered by inclusion. The *face lattice* $L(\Delta)$ is $P(\Delta)$ with a smallest element $\hat{0}$ and a largest element $\hat{1}$ attached. An example is given in Figure 1.1.2.

If we start with a simplicial complex Δ, take its face poset $P(\Delta)$, and then take the order complex $\Delta(P(\Delta))$, we get a simplicial complex known as the *barycentric*

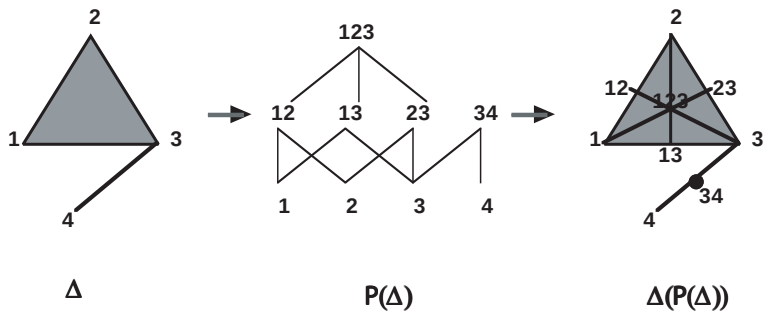


Figure 1.1.3. Barycentric subdivision

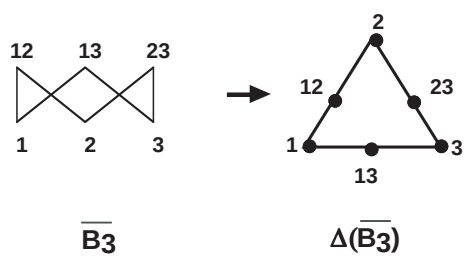


Figure 1.1.4. Order complex of the subset lattice (Boolean algebra)

subdivision of Δ; see Figure 1.1.3. The geometric realizations are always homeomorphic,

$$\Delta \cong \Delta(P(\Delta)).$$

When we attribute a topological property to a poset, we mean that the geometric realization of the order complex of the poset has that property. For instance, if we say that the poset P is homeomorphic to the n-sphere $\mathbb{S}^n$ we mean that $\|\Delta(P)\|$ is homeomorphic to $\mathbb{S}^n$.

Example 1.1.1. The Boolean algebra. Let B_n denote the lattice of subsets of $[n] := \{1, 2, \ldots, n\}$ ordered by containment, and let $\bar{B}_n := B_n - \{\emptyset, [n]\}$. Then

$$\bar{B}_n \cong \mathbb{S}^{n-2}$$

because $\Delta(\bar{B}_n)$ is the barycentric subdivision of the boundary of the $(n-1)$-simplex. See Figure 1.1.4.

We now review some basic poset terminology. An m-chain of a poset P is a totally ordered subset $c = \{x_1 < x_2 < \cdots < x_{m+1}\}$ of P. We say the length $l(c)$ of c is m. We consider the empty chain to be a (-1)-chain. The length $l(P)$ of P is defined to be

$$l(P) := \max\{l(c) : c \text{ is a chain of } P\}.$$

Thus, $l(P) = \dim \Delta(P)$ and $l(P(\Delta)) = \dim \Delta$.

A chain of P is said to be *maximal* if it is inclusionwise maximal. Thus, the set $\mathcal{M}(P)$ of maximal chains of P is the set of facets of $\Delta(P)$. A poset P is said to be

pure (also known as ranked or graded) if all maximal chains have the same length. Thus, P is pure if and only if $\Delta(P)$ is pure. Also a simplicial complex Δ is pure if and only if its face poset $P(\Delta)$ is pure. The posets and simplicial complexes of Figures 1.1.1 and 1.1.2 are all nonpure, while the poset and simplicial complex of Figure 1.1.4 are both pure.

For $x \leq y$ in P, let (x, y) denote the open interval $\{z \in P : x < z < y\}$ and let $[x, y]$ denote the closed interval $\{z \in P : x \leq z \leq y\}$. Half open intervals $(x, y]$ and $[x, y)$ are defined similarly.

If P has a unique minimum element, it is usual to denote it by $\hat{0}$ and refer to it as the *bottom* element. Similarly, the unique maximum element, if it exists, is denoted $\hat{1}$ and is referred to as the *top* element. Note that if P has a bottom element $\hat{0}$ or top element $\hat{1}$ then $\Delta(P)$ is contractible since it is a cone. We usually remove the top and bottom elements and study the more interesting topology of the remaining poset. Define the *proper part* of a poset P, for which $|P| > 1$, to be

$$\bar{P} := P - \{\hat{0}, \hat{1}\}.$$

In the case that $|P| = 1$, it will be convenient to define $\Delta(\bar{P})$ to be the degenerate empty complex $\emptyset$. We will also say $\Delta((x, y)) = \emptyset$ and $l((x, y)) = -2$ if $x = y$.

For posets with a bottom element $\hat{0}$, the elements that cover $\hat{0}$ are called *atoms*. For posets with a top element $\hat{1}$, the elements that are covered by $\hat{1}$ are called *coatoms*.

A poset P is said to be *bounded* if it has a top element $\hat{1}$ and a bottom element $\hat{0}$. Given a poset P, we define the bounded extension

$$\hat{P} := P \cup \{\hat{0}, \hat{1}\},$$

where new elements $\hat{0}$ and $\hat{1}$ are adjoined (even if P already has a bottom or top element).

A poset P is said to be a *meet semilattice* if every pair of elements $x, y \in P$ has a meet $x \wedge y$, i.e. an element less than or equal to both x and y that is greater than all other such elements. A poset P is said to be a *join semilattice* if every pair of elements $x, y \in P$ has a join $x \vee y$, i.e. a unique element greater than or equal to both x and y that is less than all other such elements. If P is both a join semilattice and a meet semilattice then P is said to be a *lattice*. It is a basic fact of lattice theory that any finite meet (join) semilattice with a top (bottom) element is a lattice.

The *dual* of a poset P is the poset P^* on the same underlying set with the order relation reversed. Topologically there is no difference between a poset and its dual since $\Delta(P)$ and $\Delta(P^*)$ are identical simplicial complexes. The direct product $P \times Q$ of two posets P and Q is the poset whose underlying set is the cartesian product $\{(p, q) : p \in P,\ q \in Q\}$ and whose order relation is given by

$$(p_1, q_1) \leq_{P \times Q} (p_2, q_2) \ \text{ if } \ p_1 \leq_P p_2 \ \text{ and } \ q_1 \leq_Q q_2.$$

Define the *join* of two simplicial complexes Δ and Γ on disjoint vertex sets to be the simplicial complex given by

$$\Delta * \Gamma := \{A \cup B : A \in \Delta, B \in \Gamma\}. \tag{1.1.1}$$

The *join* (or ordinal sum) $P * Q$ of posets P and Q is the poset whose underlying set is the disjoint union of P and Q and whose order relation is given by $x < y$ if

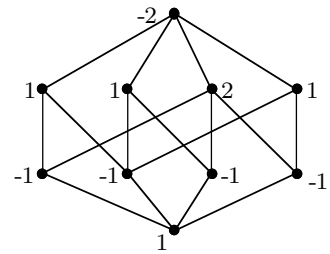


Figure 1.2.1. $\mu(\hat{0}, x)$

either (i) $x <_P y$, (ii) $x <_Q y$, or (iii) $x \in P$ and $y \in Q$. Clearly

$$\Delta(P * Q) = \Delta(P) * \Delta(Q).$$

There are topological relationships between the join and product of posets, which are discussed in Section 5.1.

1.2. The Möbius Function

The story of poset topology begins with the Möbius function $\mu(= \mu_P)$ of a poset P defined recursively on closed intervals of P as follows:

$$\begin{aligned} \mu(x,x) &= 1, \quad \text{for all } x \in P \\ \mu(x,y) &= -\sum_{x \le z < y} \mu(x,z), \quad \text{for all } x < y \in P. \end{aligned}$$

For a bounded poset P, define the Möbius invariant

$$\mu(P) := \mu_P(\hat{0}, \hat{1}).$$

In Figure 1.2.1, the values of $\mu(\hat{0}, x)$ are shown for each element x of the poset.

There are various techniques for computing the Möbius function of a poset; see [**171**]. Perhaps the most basic technique is given by the product formula.

Proposition 1.2.1. *Let P and Q be posets. Then for $(p_1, q_1) \le (p_2, q_2) \in P \times Q$,*

$$\mu_{P \times Q}((p_1, q_1), (p_2, q_2)) = \mu_P(p_1, p_2)\mu_Q(q_1, q_2).$$

Exercise 1.2.2. Prove Proposition 1.2.1.

Exercise 1.2.3. Use the product formula to show that the Möbius function for the subset lattice B_n is given by

$$\mu(X, Y) = (-1)^{|Y - X|} \quad \text{for all } X \subseteq Y.$$

Exercise 1.2.4. For positive integer n, the lattice D_n of divisors of n is the set of positive divisors of n ordered by $a \le b$ if a divides b. Show that the Möbius function for D_n is given by

$$\mu(d, m) = \mu(m/d),$$

where $\mu(\cdot)$ is the classical Möbius function of number theory, which is defined on the set of positive integers by

$$\mu(n) = \begin{cases} (-1)^k & \text{if } n \text{ is the product of } k \text{ distinct primes} \\ 0 & \text{if } n \text{ is divisible by a square.} \end{cases}$$

This example is the reason for the name *Möbius function of a poset*.

The combinatorial significance of the Möbius function was first demonstrated by Rota in 1964 in his Steele-prize winning paper [**149**]. The Möbius function of a poset is used in enumerative combinatorics to obtain inversion formulas.

Proposition 1.2.5 (Möbius inversion)**.** *Let P be a poset and let $f, g : P \to \mathbb{C}$. Then*

$$g(y) = \sum_{x \leq y} f(x)$$

if and only if

$$f(y) = \sum_{x \leq y} \mu(x, y)\, g(x).$$

Three examples of Möbius inversion are classical Möbius inversion ($P = D_n$), inclusion-exclusion ($P = B_n$), and Gaussian inversion ($P = B_n(q)$, the lattice of subspaces of an n-dimensional vector space over the field with q elements); see [**171**] for details.

Our interest in the Möbius function stems from its connection to the Euler characteristic. The reduced Euler characteristic $\tilde{\chi}(\Delta)$ of a simplicial complex Δ is defined to be

$$\tilde{\chi}(\Delta) := \sum_{i=-1}^{\dim \Delta} (-1)^i f_i(\Delta),$$

where $f_i(\Delta)$ is the number of i-faces of Δ.

Proposition 1.2.6 (Philip Hall Theorem)**.** *For any poset P,*

$$\mu(\hat{P}) = \tilde{\chi}(\Delta(P)).$$

Exercise 1.2.7. Prove Proposition 1.2.6.

It follows from the Euler-Poincaré formula below that the Euler characteristic is a topological invariant. Hence by Proposition 1.2.6, $\mu_P(x, y)$ depends only on the topology of the open interval (x, y) of P.

Theorem 1.2.8 (Euler-Poincaré formula)**.** *For any simplicial complex Δ,*

$$\tilde{\chi}(\Delta) := \sum_{i=-1}^{\dim \Delta} (-1)^i \tilde{\beta}_i(\Delta),$$

where $\tilde{\beta}_i(\Delta)$ is the ith reduced Betti number of Δ, i.e., the rank, as an abelian group, of the ith reduced homology of Δ over $\mathbb{Z}$.

The Möbius function of a poset plays a fundamental role in the theory of hyperplane arrangements and the homology of a poset plays a fundamental role in the theory of subspace arrangements. We discuss the connection with arrangements in the next section.

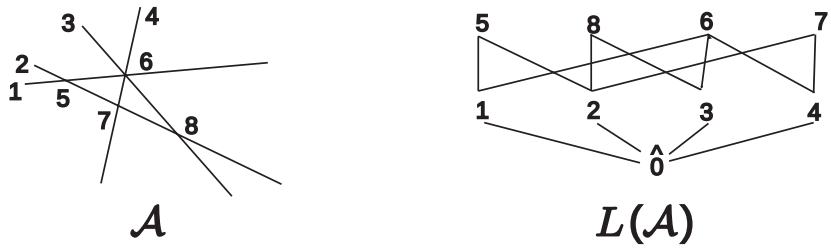


Figure 1.3.1. Intersection semilattice of a hyperplane arrangement

1.3. Hyperplane and Subspace Arrangements

A *hyperplane arrangement* $\mathcal{A}$ is a finite collection of (affine) hyperplanes in some vector space V. We will consider only real hyperplane arrangements ($V = \mathbb{R}^n$) and complex hyperplane arrangements ($V = \mathbb{C}^n$) here.

Real hyperplane arrangements divide $\mathbb{R}^n$ into regions. A remarkable formula for the number of regions was given by Zaslavsky [**215**] in 1975. This formula involves the notion of intersection semilattice of a hyperplane arrangement.

The *intersection semilattice* $L(\mathcal{A})$ of a hyperplane arrangement $\mathcal{A}$ is defined to be the meet semilattice of nonempty intersections of hyperplanes in $\mathcal{A}$ ordered by reverse inclusion. Note that we include the intersection over the empty set which is the bottom element $\hat{0}$ of $L(\mathcal{A})$. Note also that $L(\mathcal{A})$ has a top element if and only if $\cap\mathcal{A} \neq \emptyset$. Such an arrangement is called a *central arrangement*. Hence for central arrangements $\mathcal{A}$, the intersection semilattice $L(\mathcal{A})$ is actually a lattice. An example of a hyperplane arrangement in $\mathbb{R}^2$ and its intersection semilattice are given in Figure 1.3.1.

Before stating Zaslavsky's formula, we discuss four fundamental examples of real hyperplane arrangements and their intersection lattices, to which we refer throughout these lectures.

Example 1.3.1. *The (type A) coordinate hyperplane arrangement and the Boolean algebra B_n.* The coordinate hyperplane arrangement is the central hyperplane arrangement consisting of the coordinate hyperplanes $x_i = 0$ in $\mathbb{R}^n$. It is easy to see that the intersection lattice of this arrangement is isomorphic to the subset lattice B_n. Indeed, the intersection

$$\{\mathbf{x} \in \mathbb{R}^n : x_{i_1} = x_{i_2} = \cdots = x_{i_k} = 0\},$$

where $1 \leq i_1 < i_2 < \cdots < i_k \leq n$, corresponds to the subset $\{i_1, i_2, \ldots, i_k\}$. This correspondence is an isomorphism from the intersection lattice to B_n.

Example 1.3.2. *The type B coordinate hyperplane arrangement and the face lattice of the n-cross-polytope C_n.* The type B coordinate hyperplane arrangement is the affine hyperplane arrangement consisting of the hyperplanes $x_i = \pm 1$ in $\mathbb{R}^n$. One can see that if we attach a top element $\hat{1}$ to the intersection semilattice of this arrangement we have a lattice that is isomorphic to the lattice of faces of the n-cross-polytope, which we denote by C_n. (This is dual to the lattice of faces of the

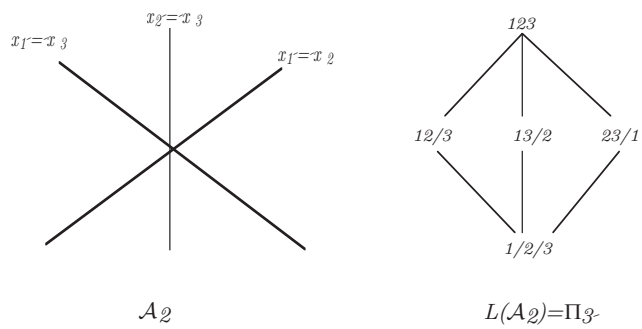


Figure 1.3.2. Intersection lattice of braid arrangement

n-cube.) Indeed, the intersection

$$\{\mathbf{x} \in \mathbb{R}^n : \epsilon_1 x_{i_1} = \epsilon_2 x_{i_2} = \cdots = \epsilon_k x_{i_k} = 1\},$$

where $1 \leq i_1 < i_2 < \cdots < i_k \leq n$ and $\epsilon_i \in \{-1, 1\}$, maps to the $(n-k)$-face

$$\{\mathbf{x} \in [-1, 1]^n : \epsilon_1 x_{i_1} = \epsilon_2 x_{i_2} = \cdots = \epsilon_k x_{i_k} = 1\}$$

of the n-cube. This correspondence is an isomorphism from the intersection lattice to the dual of the face lattice of the cube.

Example 1.3.3. *The (type A) braid arrangement and the partition lattice* Π_n. For $1 \leq i < j \leq n$, let

$$H_{i,j} = \{\mathbf{x} \in \mathbb{R}^n : x_i = x_j\}.$$

The hyperplane arrangement

$$\mathcal{A}_{n-1} := \{H_{i,j} : 1 \leq i < j \leq n\}$$

is known as the *braid arrangement* or the *type A Coxeter arrangement*. The intersection lattice $L(\mathcal{A}_{n-1})$ is isomorphic to Π_n, the lattice of partitions of the set $[n]$ ordered by refinement. Indeed, for each partition $\pi \in \Pi_n$, let ℓ_π be the linear subspace of $\mathbb{R}^n$ consisting of all points $(x_1, \ldots, x_n)$ such that $x_i = x_j$ whenever i and j are in the same block of π. The map $\pi \mapsto \ell_\pi$ is a poset isomorphism from Π_n to $L(\mathcal{A}_{n-1})$. The braid arrangement $\mathcal{A}_2$ intersected with the plane $x_1 + x_2 + x_3 = 0$ and the partition lattice Π_3 are shown in Figure 1.3.2.

Example 1.3.4. *The type B braid arrangement and the type B partition lattice* Π_n^B. For $1 \leq i < j \leq n$, let

$$H_{i,j}^+ = \{\mathbf{x} \in \mathbb{R}^n : x_i = x_j\} \quad \text{and} \quad H_{i,j}^- = \{\mathbf{x} \in \mathbb{R}^n : x_i = -x_j\}.$$

For $i = 1, \ldots n$, let

$$H_i = \{\mathbf{x} \in \mathbb{R}^n : x_i = 0\}.$$

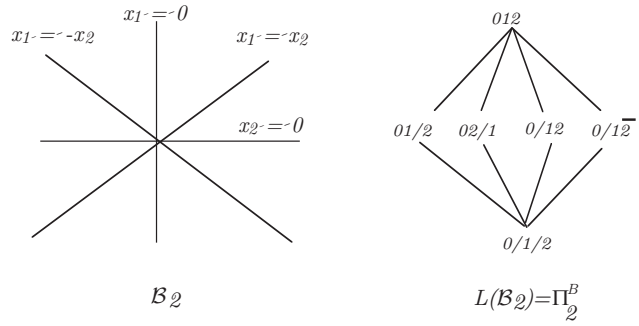


Figure 1.3.3. Intersection lattice of type B braid arrangement

The hyperplane arrangement

$$\mathcal{B}_n := \{H^+_{i,j} : 1 \le i < j \le n\} \cup \{H^-_{i,j} : 1 \le i < j \le n\} \cup \{H_i : 1 \le i \le n\}$$

is called the *type B braid arrangement* or the *type B Coxeter arrangement*. The intersection lattice $L(\mathcal{B}_n)$ is isomorphic to the *type B (or signed) partition lattice*. The elements of the type B partition lattice Π^B_n are partitions of $\{0, \ldots, n\}$ for which any of the elements of $[n]$ can have a bar except for the elements of the block that contains 0 and the smallest element of each block. For example, $025/1\bar{7}9/34\bar{6}\bar{8}$ is an element of Π^B_9, while $02\bar{5}/1\bar{7}9/\bar{3}4\bar{6}\bar{8}$ is not because 5 and 3 are not allowed to be barred. The covering relation is given by $\pi_1 \lessdot \pi_2$ if π_2 is obtained from π_1 by merging two blocks B_1 and B_2 into a single block B in the following manner: Suppose $\min B_1 < \min B_2$. Then

- if $0 \in B_1$, let B be the union of B_1 and B_2 with all bars removed,
- if $0 \notin B_1$, let B be the union of either
  - B_1 and B_2 with all bars intact or
  - B_1 and $\bar{B}_2$, where $\bar{B}_2$ is obtained from B_2 by barring all unbarred elements and unbarring all barred elements.

For example, the type B partitions that cover $025/1\bar{7}9/34\bar{6}\bar{8}$ in Π^B_9 are

$$012579/34\bar{6}\bar{8}, \qquad 0234568/1\bar{7}9, \qquad 025/134\bar{6}\bar{7}\bar{8}9, \qquad 025/1\bar{3}\bar{4}\bar{6}\bar{7}89.$$

The isomorphism from Π^B_n to $L(\mathcal{B}_n)$ is quite natural. Take a typical type B partition

$$025/1\bar{7}9/34\bar{6}\bar{8}.$$

It maps to the subspace

$$\{\mathbf{x} \in \mathbb{R}^9 : x_2 = x_5 = 0,\, x_1 = -x_7 = x_9,\, x_3 = x_4 = -x_6 = -x_8\},$$

in $L(\mathcal{B}_n)$. The type B braid arrangement $\mathcal{B}_2$ and the type B partition lattice Π^B_2 are shown in Figure 1.3.3.

Examples 1.3.1 and 1.3.3 are referred to as type A examples, and Examples 1.3.2 and 1.3.4 are referred to as type B examples because of their connection with Coxeter groups. Indeed, associated with every finite Coxeter group (i.e., finite group generated by Euclidean reflections) is a simplicial complex called its Coxeter complex. The order complex of the Boolean algebra B_n is the Coxeter complex of the symmetric group $\mathfrak{S}_n$, which is the type A Coxeter group, and the order complex of the face lattice of the cross-polytope C_n is the Coxeter complex of the hyperoctahedral group, which is the type B Coxeter group. (The use of notation is unfortunate here; B_n is type A and C_n is type B.)

Also associated with every finite Coxeter group is a hyperplane arrangement, called its Coxeter arrangement, which consists of all its reflecting hyperplanes. The group generated by the reflections about hyperplanes in the Coxeter arrangement is the Coxeter group. The braid arrangement is the Coxeter arrangement of the symmetric group (type A Coxeter group) and type B braid arrangement is the Coxeter arrangement of the hyperoctahedral group (type B Coxeter group). Coxeter groups are discussed further in Section 3.3. See the chapters in this volume by Fomin and Reading [**71**] and Stanley [**176**] for further discussion of Coxeter arrangements.

The types A and B partition lattices belong to another family of well-studied lattices, namely the Dowling lattices. We will not define Dowling lattices, but we will occasionally refer to them; see [**83**] for the definition. A broad class of Dowling lattices arise as intersection lattices of complex hyperplane arrangements $\mathcal{A}_{m,n}$ consisting of hyperplanes of the forms $z_j = 0$, where $j = 1, \ldots, n$, and $z_j = \omega^h z_i$, where ω is the mth primitive root of unity $e^{\frac{2\pi i}{m}}$, $1 \le i < j \le n$, and $h \in [m]$. This class includes the types A and B partition lattices.

We now state Zaslavsky's seminal result.

Theorem 1.3.5 (Zaslavsky [**215**])**.** *Suppose $\mathcal{A}$ is a hyperplane arrangement in $\mathbb{R}^n$. Let $r(\mathcal{A})$ be the number of regions into which $\mathcal{A}$ divides $\mathbb{R}^n$ and let $b(\mathcal{A})$ be the number of these regions that are bounded. Then*

$$r(\mathcal{A}) = \sum_{x \in L(\mathcal{A})} |\mu(\hat{0}, x)| \tag{1.3.1}$$

and

$$b(\mathcal{A}) = |\mu(L(\mathcal{A}) \cup \hat{1})|. \tag{1.3.2}$$

The arrangement of Figure 1.3.1 has a total of 10 regions with 2 of them bounded. One can use the values of the Möbius function given in Figure 1.2.1 to confirm (1.3.1) and (1.3.2) for the arrangement of Figure 1.3.1. Note that if $\mathcal{A}$ is a central arrangement, $L(\mathcal{A}) \cup \hat{1}$ has an artificial top element above the top element of $L(\mathcal{A})$. In other words $L(\mathcal{A}) \cup \hat{1}$ has exactly one coatom. It is easy to see that posets with only one coatom have Möbius invariant 0. Since central arrangements clearly have no bounded regions, (1.3.2) is trivial for central arrangements.

Exercise 1.3.6. Suppose we have a hyperplane H of $\mathbb{R}^n$ which is generic with respect to a central hyperplane arrangement $\mathcal{A}$ in $\mathbb{R}^n$. This means that $\dim(H \cap X) = \dim(X) - 1$ for all $X \in L(\mathcal{A})$. Let $\mathcal{A}_H = \{H \cap K : K \in \mathcal{A}\}$. This is a hyperplane arrangement induced in $H \cong R^{n-1}$. Show that the number of bounded regions of $\mathcal{A}_H$ is independent of the choice of generic hyperplane H (see [**42**]).

The next major development in the combinatorial theory of hyperplane arrangements is a 1980 formula of Orlik and Solomon [**131**], which can be viewed as a

complex analog of (1.3.1). The number of regions in a real hyperplane arrangement $\mathcal{A}$ is equal to the sum of all the Betti numbers of the complement $\mathbb{R}^n - \cup\mathcal{A}$. Indeed, since each region is contractible, all the Betti numbers are 0 except for the degree 0 Betti number, which equals the number of regions. Hence (1.3.1) can be interpreted as a formula for the sum of the Betti numbers of the complement $\mathbb{R}^n - \cup\mathcal{A}$. The analog for complex arrangements is given by following result.

Theorem 1.3.7 (Orlik and Solomon [**131**]). *Let $\mathcal{A}$ be a hyperplane arrangement in $\mathbb{C}^n$. The complement $M_{\mathcal{A}} := \mathbb{C}^n - \cup\mathcal{A}$ has torsion-free integral cohomology and has Betti numbers given by,*

$$\beta_i(M_{\mathcal{A}}) = \sum_{\substack{x \in L(\mathcal{A}) \\ \dim^{\mathbb{C}}(x) = n-i}} |\mu(\hat{0}, x)|,$$

for all i.

There is a striking common generalization of the Zaslavsky formula (1.3.1) and the Orlik-Solomon formula, obtained by Goresky and MacPherson in 1988, which involves subspace arrangements. A *real subspace arrangement* is a finite collection of (affine) subspaces in $\mathbb{R}^n$. Real hyperplane arrangements and complex hyperplane arrangements are both examples of real subspace arrangements. Indeed, hyperplanes in $\mathbb{C}^n$ can be viewed as codimension 2 subspaces of $\mathbb{R}^{2n}$. Again the intersection semilattice $L(\mathcal{A})$ is defined to be the semilattice of nonempty intersections of subspaces in the subspace arrangement $\mathcal{A}$.

Theorem 1.3.8 (Goresky and MacPherson [**80**]). *Let $\mathcal{A}$ be a subspace arrangement in $\mathbb{R}^n$. The reduced integral cohomology of the complement $M_{\mathcal{A}} := \mathbb{R}^n - \cup\mathcal{A}$ is given by the group isomorphism*

$$\tilde{H}^i(M_{\mathcal{A}}; \mathbb{Z}) \cong \bigoplus_{x \in L(\mathcal{A}) \setminus \{\hat{0}\}} \tilde{H}_{n - \dim x - 2 - i}((\hat{0}, x); \mathbb{Z}),$$

for all i.

To see that the Goresky-MacPherson formula reduces to the Zaslavsky formula and to the Orlik-Solomon formula, one needs to understand the homology of the intersection lattice of a central hyperplane arrangement. The intersection lattice belongs to a well-understood class of lattices called geometric lattices. A fundamental result due to Folkman [**70**] states that the proper part of any geometric lattice L has vanishing reduced homology in every dimension except the top dimension (i.e. dimension equal to $l(L) - 2$). In fact, the homotopy type is that of a wedge of spheres of top dimension. The intersection lattice of an affine hyperplane arrangement belongs to a more general class of lattices called geometric semilattices, which were introduced and studied by Wachs and Walker [**205**]. The proper part of a geometric semilattice also has the homotopy type of a wedge of spheres of top dimension. Topology of geometric (semi)lattices is discussed further in Sections 3.2.3 and 4.2.

Exercise 1.3.9. Use Folkman's result to show that the Goresky-MacPherson formula reduces to both the Zaslavsky formula and the Orlik-Solomon formula.

The intersection lattice of a hyperplane arrangement determines more than the additive group structure of the integral cohomology of the complement. Orlik and

Solomon show that it determines the ring structure as well. Ziegler [**218**] showed that, in general, for subspace arrangements the combinatorial data (intersection lattice and dimension information) does not determine ring structure. However in certain special cases the combinatorial data does determine the cohomology algebra, see [**69**], [**214**], [**58**]. In Section 5.4 we discuss some stronger versions of the Goresky-MacPherson formula, namely a homotopy version due to Ziegler and Živaljević [**222**], and an equivariant version due to Sundaram and Welker [**186**]. For further reading on hyperplane arrangements, see the chapter by Stanley in this volume [**176**] and the text by Orlik and Terao [**132**]. Further information on subspace arrangements can be found in Björner [**28**] and Ziegler [**216**].

1.4. Some Connections with Graphs, Groups and Lattices

In this section we briefly discuss some results and questions in which poset topology plays a role. We start with a old conjecture of Karp in graph complexity theory. An algorithm for deciding whether a graph with n nodes has a certain property checks the entries of the graph's adjacency matrix until a determination can be made. A graph property is said to be evasive if the best algorithm needs to check all $\binom{n}{2}$ entries (in the worst case). Here are some examples of evasive graph properties:

- property of being connected
- property of containing a perfect matching
- property of having degree at most b for some fixed b.

A monotone graph property is a property of graphs that is isomorphism invariant and closed under addition of edges or closed under removal of edges. The graph properties listed above are clearly monotone graph properties. We say that a graph property is trivial if every graph has the property or every graph lacks the property.

Conjecture 1.4.1 (Karp's Evasiveness Conjecture)**.** *Every nontrivial monotone graph property is evasive.*

Kahn, Saks, and Sturtevant [**108**] proved the evasiveness conjecture for n a prime power by using topological techniques and group actions. Since determining whether a graph has a certain property is equivalent to determining whether the graph lacks the property, one can require without loss of generality that a monotone graph property be closed under removal of edges. Given such a monotone graph property, $\mathcal{P}$, let $\Delta^n_{\mathcal{P}}$ be the simplicial complex whose vertex set is $\binom{[n]}{2}$ and whose faces are the edge sets of graphs on node set $[n]$ that have the property. Alternatively, $\Delta^n_{\mathcal{P}}$ is the simplicial complex whose face poset is the poset of graphs on node set $[n]$ that have property $\mathcal{P}$, ordered by edge set inclusion. Kahn, Saks, and Sturtevant show that

- nonvanishing reduced simplicial homology of $\Delta^n_{\mathcal{P}}$ implies $\mathcal{P}$ is evasive
- $\Delta^n_{\mathcal{P}}$ has nonvanishing reduced simplicial homology when n is a prime power, using a topological fixed point theorem.

Although this connection between evasiveness and topology doesn't really involve posets directly, we mention it here because posets and simplicial complexes can be viewed as the same object, and as we will see in later in these lectures, the tools of poset topology are useful in the study of the topology of graph complexes. For other significant results on evasiveness and topology of graph complexes, see

eg., [**55**], [**213**], [**73**]. This topic is discussed in greater depth in Forman's chapter of this volume [**74**]. Applications of graph complexes in knot theory and group theory are discussed in Section 5.2. There are also connections between the topology of graph complexes and commutative algebra, which are explored in the work of Reiner and Roberts [**140**] and Dong [**61**]. A direct application of poset topology in a different complexity theory problem is discussed in Section 3.2.4.

Representability questions in lattice theory deal with whether an arbitrary lattice can be represented as a sublattice, subposet or interval in a given class of lattices. We briefly discuss three examples that have connections to poset topology.

A result of Pudlák and Tuma [**136**] states that every lattice is isomorphic to a sublattice of some partition lattice Π_n. This implies that every lattice can be represented as the intersection lattice of a subspace arrangement embedded in the braid arrangement. There is another representability result that is much easier to prove; namely that every meet semilattice can be represented as the intersection semilattice of some subspace arrangement, see [**216**]. From either of these representability results, we see that, in contrast to the situation with hyperplane arrangements, where the topology of the proper part of the intersection semilattice is rather special (a wedge of spheres), any topology is possible for the intersection semilattice of a general subspace arrangement. Indeed, given any simplicial complex Δ, there is a linear subspace arrangement $\mathcal{A}$ such that $\overline{L(\mathcal{A})}$ is homeomorphic to Δ; namely $\mathcal{A}$ is the linear subspace arrangement whose intersection lattice $L(\mathcal{A})$ is isomorphic to the face lattice $L(\Delta)$.

An open representability question is whether every lattice can be represented as an interval in the lattice of subgroups of some group ordered by inclusion. An approach to obtaining a negative answer to this question, proposed by Shareshian [**155**], is to establish restrictions on the topology of intervals in the subgroup lattice.

Conjecture 1.4.2 (Shareshian [**155**]). *Let G be a finite group. Then every open interval in the lattice of subgroups of G has the homotopy type of a wedge of spheres.*

This conjecture was shown to hold for solvable groups by Kratzer and Thévenaz [**114**] (see Theorem 3.1.13 which strengthens the Kratzer-Thévenaz result). Further discussion of connections between poset topology and group theory can be found in Section 5.2

Our last example deals with the *order dimension* of a poset P, which is defined to be the smallest integer n such that P can be represented as an induced subposet of a product of n chains. Order dimension is an important and extensively studied poset invariant, see [**191**]. Reiner and Welker give a lower bound on order dimension of a lattice in terms of its homology.

Theorem 1.4.3 (Reiner and Welker [**143**]). *Let L be a lattice and let d be the largest dimension for which the reduced integral simplicial homology of the proper part of L is nonvanishing. Then the order dimension of L is at least $d+2$.*

1.5. Poset Homology and Cohomology

By (co)homology of a poset, we usually mean the reduced simplicial (co)homology of its order complex. On rare occasions, we will deal with nonreduced simplicial homology. Although it is presumed that the reader is familiar with simplicial homology and cohomology, we review these concepts for posets in terms of chains of

the poset. For each poset P and integer j, define the chain space

$$C_j(P;\mathbf{k}) := \mathbf{k}\text{-module freely generated by } j\text{-chains of } P,$$

where $\mathbf{k}$ is a field or the ring of integers.

The boundary map $\partial_j : C_j(P;\mathbf{k}) \to C_{j-1}(P;\mathbf{k})$ is defined by

$$\partial_j(x_1 < \cdots < x_{j+1}) = \sum_{i=1}^{j+1} (-1)^i \, (x_1 < \cdots < \hat{x}_i < \cdots < x_{j+1}),$$

where the $\hat{\cdot}$ denotes deletion. We have that $\partial_{j-1}\partial_j = 0$, which makes $(C_j(P;\mathbf{k}), \partial_j)$ an algebraic complex. Define the cycle space $Z_j(P;\mathbf{k}) := \ker \partial_j$ and the boundary space $B_j(P;\mathbf{k}) := \mathrm{im}\partial_{j+1}$. Homology of the poset P in dimension j is defined by

$$\tilde{H}_j(P;\mathbf{k}) := Z_j(P;\mathbf{k})/B_j(P;\mathbf{k}).$$

The coboundary map $\delta_j : C_j(P;\mathbf{k}) \to C_{j+1}(P;\mathbf{k})$ is defined by

$$\langle \delta_j(\alpha), \beta \rangle = \langle \alpha, \partial_{j+1}(\beta) \rangle \tag{1.5.1}$$

where $\alpha \in C_j(P;\mathbf{k})$, $\beta \in C_{j+1}(P;\mathbf{k})$, and $\langle \cdot, \cdot \rangle$ is the bilinear form on $\oplus_{j \geq -1} C_j(P;\mathbf{k})$ for which the chains of P form an orthonormal basis. This is equivalent to saying

$$\delta_j(x_1 < \cdots < x_j) = \sum_{i=1}^{j+1} (-1)^i \sum_{x \in (x_{i-1}, x_i)} (x_1 < \cdots < x_{i-1} < x < x_i < \cdots < x_j),$$

for all chains $x_1 < \cdots < x_j$, where x_0 is the bottom element of $\hat{P}$ and x_{j+1} is the top element of $\hat{P}$. Define the cocycle space to be $Z^j(P;\mathbf{k}) := \ker \delta_j$ and the coboundary space to be $B^j(P;\mathbf{k}) := \mathrm{im}\delta_{j-1}$. Cohomology of the poset P in dimension j is defined to be

$$\tilde{H}^j(P;\mathbf{k}) := Z^j(P;\mathbf{k})/B^j(P;\mathbf{k}).$$

When $\mathbf{k}$ is a field, $\tilde{H}^j(P;\mathbf{k})$ and $\tilde{H}_j(P;\mathbf{k})$ are isomorphic vector spaces. The jth (reduced) Betti number of P is given by

$$\tilde{\beta}_j(P) := \dim \tilde{H}_j(P;\mathbb{C}),$$

which is the same as the rank of the free part of $\tilde{H}_j(P;\mathbb{Z})$.

We will work primarily with homology over $\mathbb{C}$ and $\mathbb{Z}$. For $x < y$ in P, we write $\tilde{H}_j(x,y)$ for the complex homology of the open interval (x,y) of P, and $\tilde{\beta}_j(x,y)$ for the jth Betti number of the open interval (x,y). When $x = y$, define $\tilde{H}_j(x,y)$ to be $\mathbb{C}$ and $\tilde{\beta}_j(x,y)$ to be 1 if $j = -2$, and to be 0 for all other j.

Many of the posets that arise have the homotopy type of a wedge of spheres. We review a basic fact pertaining to wedges of spheres and a partial converse.

Theorem 1.5.1. *Suppose Δ has the homotopy type of a wedge of spheres of various dimensions, where r_i is the number of spheres of dimension i. Then for each $i = 0, 1, \ldots, \dim \Delta$,*

$$\tilde{H}_i(\Delta;\mathbb{Z}) \cong \tilde{H}^i(\Delta;\mathbb{Z}) \cong \mathbb{Z}^{r_i}. \tag{1.5.2}$$

Theorem 1.5.2. *If Δ is simply connected and has vanishing reduced integral homology in all dimensions but dimension n, where homology is free of rank r, then Δ has the homotopy type of a wedge of r spheres of dimension n.*

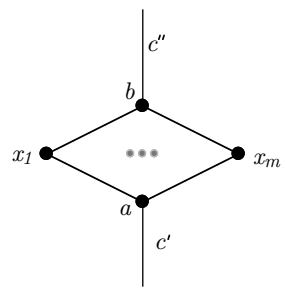


Figure 1.6.1. Top coboundary relations: $\delta(c'c'')$

The first tool that we mention for computing homology of posets and simplicial complexes is a very efficient computer software package called "SimplicialHomology", developed by Dumas, Heckenbach, Sauders, and Welker [**63**]. One can run it interactively or download the source file at the web site:

`http://www.cis.udel.edu/~dumas/Homology`,

where a manual can also be found. This package has been responsible for many of the more recent conjectures in the field. Its output was also part of the proofs of (at least) three results on integral homology appearing in the literature; see [**156, 157, 202**].

1.6. Top Cohomology of the Partition Lattice

The top dimensional cohomology of a poset has a particularly simple description. For the sake of simplicity assume P is pure of length d. Let $\mathcal{M}(P)$ be the set of maximal chains of P and let $\mathcal{M}'(P)$ be the set of chains of length $d-1$. Since $\ker \delta_d = C_d(P;\mathbf{k})$, we have the following presentation of top cohomology as a quotient of $C_d(P;\mathbf{k})$:

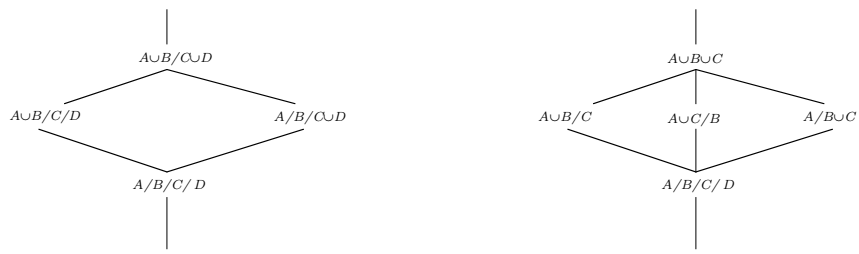
$$\tilde{H}^d(P;\mathbf{k}) = \langle \mathcal{M}(P) \mid \text{coboundary relations} \rangle,$$

where the coboundary relations have the form $\delta_{d-1}(c)$ for $c \in \mathcal{M}'(P)$. Each chain c in $\mathcal{M}'(P)$ is the concatenation $c'c''$ of two unrefinable chains c' and c''. If c' is not empty, let a be the maximum element of c_1, and if c' is empty, let a be $\hat{0}$ of $\hat{P}$. If c'' is not empty, let b be the minimum element of c'', and if c'' is empty, let b be $\hat{1}$ of $\hat{P}$. Clearly, $[a,b]$ is an interval of length 2 in $\hat{P}$. Let $\{x_1,\ldots,x_m\}$ be the set of elements in the open interval (a,b). We have

$$\delta_{d-1}(c) = \pm(c'x_1c'' + \cdots + c'x_mc'').$$

Hence the cohomology relations can be associated with the intervals of length 2 in $\hat{P}$. See Figure 1.6.1.

We demonstrate the use of intervals of length 2 by deriving a presentation for the top cohomology of the proper part of the partition lattice $\bar{\Pi}_n$. There are two types of length 2 closed intervals in Π_n; see Figure 1.6.2. In Type I intervals, there are 2 pairs of blocks $\{A,B\}$ and $\{C,D\}$ which are separately merged resulting in blocks $A \cup B$ and $C \cup D$. In Type II intervals, there are 3 blocks A, B, C, which



Type I interval Type II interval

Figure 1.6.2.

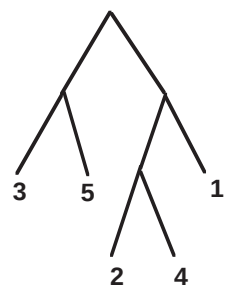


Figure 1.6.3.

are merged into one block $A \cup B \cup C$. Type I intervals have 4 elements and type II intervals have 5 elements.

The two types of intervals induce two types of cohomology relations, Type I and Type II cohomology relations on maximal chains. It is convenient to use binary trees on leaf set $[n]$ to describe these relations. A maximal chain of $\bar{\Pi}_n$ is just a sequence of merges of pairs of blocks. The binary tree given in Figure 1.6.3 corresponds to the sequence of merges:

(1) merge blocks $\{3\}$ and $\{5\}$
(2) merge blocks $\{2\}$ and $\{4\}$
(3) merge blocks $\{2, 4\}$ and $\{1\}$.

This corresponds to the maximal chain

$$1/2/35/4 \lessdot 1/24/35 \lessdot 124/35$$

of $\bar{\Pi}_5$ The internal nodes of the tree represent the merges, and the leaf sets of the left and right subtrees of the internal nodes are the blocks that are merged. The sequence of merges follows the postorder traversal of the internal nodes, i.e. first traverse the left subtree in postorder, then the right subtree in postorder, then the root.

Given a binary tree T on leaf set $[n]$, let $c(T)$ be the maximal chain of Π_n obtained by the procedure described above. Although not all maximal chains can be obtained in this way, it can be seen that every maximal chain is equal, modulo the cohomology relations of Type I, to $\pm c(T)$ for some T. So the set

$$\{c(T) : T \text{ is a binary tree on leaf set } [n]\}$$

generates top cohomology $\tilde{H}^{n-3}(\bar{\Pi}_n;\mathbf{k})$. The Type I cohomology relations induce the following relations

$$c(\cdots(A\wedge B)\cdots) = (-1)^{|A||B|}\, c(\cdots(B\wedge A)\cdots), \tag{1.6.1}$$

where $X\wedge Y$ denotes the binary tree whose left subtree is X and whose right subtree is Y, and $|X|$ denotes the number of internal nodes of X. The Type II cohomology relations induce the following relations

$$\begin{aligned} c(\cdots(A\wedge(B\wedge C))\cdots) \quad &+ \quad (-1)^{|C|}c(\cdots((A\wedge B)\wedge C)\cdots) \\ &+ \quad (-1)^{|A||B|}c(\cdots(B\wedge(A\wedge C))\cdots) = 0. \end{aligned} \tag{1.6.2}$$

Exercise 1.6.1. Show that the Type I cohomology relations yield (1.6.1) and the Type II cohomology relations yield (1.6.2).

The relations (1.6.1) and (1.6.2) resemble the relations satisfied by the bracket operation of a Lie algebra. The Type I relation (1.6.1) corresponds to the anticommuting relation and the Type II relation (1.6.2) corresponds to the Jacobi relation. Indeed there is a well-known connection between the top homology of the partition lattice and the free Lie algebra which involves representations of the symmetric group. In the next lecture, we discuss representation theory.

Theorem 1.6.2 (Stanley [**169**], Klyachko [**110**], Joyal [**106**])**.** *The representation of the symmetric group $\mathfrak{S}_n$ on $\tilde{H}_{n-3}(\bar{\Pi}_n;\mathbb{C})$ is isomorphic to the representation of $\mathfrak{S}_n$ on the multilinear component of the free Lie algebra over $\mathbb{C}$ on n generators tensored with the sign representation.*

This result follows from a formula of Stanley for the representation of the symmetric group on homology of the partition lattice (Theorem 4.4.7) and an earlier similar formula of Klyachko for the free Lie algebra. The first purely combinatorial proof was obtained by Barcelo [**14**]. The presentation of top cohomology discussed above appeared in an alternative combinatorial proof of Wachs [**199**]. It also appeared in the proof of a superalgebra version of this result obtained by Hanlon and Wachs [**91**]. A k-analog of the Lie superalgebra result was also obtained by Hanlon and Wachs [**91**]. A type B version (Example 1.3.4) was obtained by Bergeron [**16**] and a generalization to Dowling lattices was obtained by Gottlieb and Wachs [**83**].

LECTURE 2
Group Actions on Posets

In this lecture we give a crash course on the representation theory of the symmetric group and then discuss some representations on homology that are induced by symmetric group actions on posets. For further details on the representation theory of the symmetric group and symmetric functions, we refer the reader to the following excellent standard references [**76, 126, 151, 174**].

There are various reasons that we are interested in understanding how a group acts on the homology of a poset. One is that this can be a useful tool in computing the homology of the poset. Another is that interesting representations often arise. We limit our discussion to the symmetric group, but point out there are often interesting analogous results for other groups such as the hyperoctahedral group, wreath product groups, and the general linear group.

2.1. Group Representations

We restrict our discussion to finite groups G and finite dimensional vector spaces over the field $\mathbb{C}$. A finite dimensional vector space V over $\mathbb{C}$ is said to be a *representation* of G if there is a group homomorphism

$$\phi : G \to GL(V).$$

For $g \in G$ and $v \in V$, we write gv instead of $\phi(g)(v)$ and view V as a module over the ring $\mathbb{C}G$ (G-module for short). The *dimension* of the representation V is defined to be the dimension of V as a vector space.

There are two particular representations of every group that are very important; the trivial representation and the regular representation. The *trivial representation*, denoted 1_G, is the 1-dimensional representation $V = \mathbb{C}$, where $gz = z$ for all $g \in G$ and $z \in \mathbb{C}$. The *(left) regular representation* is the G-module $\mathbb{C}G$ where G acts on itself by left multiplication, i.e., the action of $g \in G$ on generator $h \in G$ is gh.

We say that V_1 and V_2 are *isomorphic* representations of G and write $V_1 \cong_G V_2$, if there is a vector space isomorphism $\psi : V_1 \to V_2$ such that

$$\psi(gv) = g\psi(v)$$

for all $g \in G$ and $v \in V_1$. In other words, V_1 and V_2 are isomorphic representations of G means that they are isomorphic G-modules.

The *character* of a G-module V is a function $\chi^V : G \to \mathbb{C}$ defined by

$$\chi^V(g) = \mathrm{trace}(\phi(g)).$$

One basic fact of representation theory is that the character of a representation completely determines the representation. Another is that $\chi^V(g)$ depends only on the conjugacy class of g.

A G-module V is said to be *irreducible* if its only submodules are the trivial submodule 0 and V itself. A basic result of representation theory is that the number of irreducible representations of G is the same as the number of conjugacy classes of G. Another very important fact is that every G-module decomposes into a direct sum of irreducible submodules,

$$V \cong_G V_1 \oplus \cdots \oplus V_m.$$

The decomposition is unique (up to order and up to isomorphism). Hence it makes sense to talk about the multiplicity of an irreducible in a representation. We have the following fundamental fact.

Theorem 2.1.1. *The multiplicity of any irreducible representation of G in the regular representation of G is equal to the dimension of the irreducible.*

There are two operations on representations that are quite useful. The first is called *restriction*. For H a subgroup of G and V a representation of G, the restriction of V to H, denoted $V \downarrow_H^G$, is the representation of H obtained by restricting ϕ to H. Thus the restriction has the same underlying vector space with a smaller group action. The other operation, which is called induction, is a bit more complicated. For H a subgroup of G and V a representation of H, the induction of V to G is given by

$$V \uparrow_H^G := \mathbb{C}G \otimes_{\mathbb{C}H} V,$$

where the tensor product $A \otimes_S B$ denotes the usual tensor product of an (R, S)-bimodule A and a left S-module B resulting in a left R-module. Now the underlying vector space of the induction is larger than V.

Exercise 2.1.2. Show

$$\dim V \uparrow_H^G = \frac{|G|}{|H|} \dim V.$$

Although restriction and induction are not inverse operations, they are related by a formula called Frobenius reciprocity. We will state an important special case, which is, in fact, equivalent to Frobenius reciprocity.

Theorem 2.1.3. *Let U be an irreducible representation of H and let V be an irreducible representation of G, where H is a subgroup of G. Then the multiplicity of U in $V \downarrow_H^G$ is equal to the multiplicity of V in $U \uparrow_H^G$.*

There are two types of tensor products of representations. Given a representation U of G and a representation V of H, the (outer) tensor product $U \otimes V$ is a representation of $G \times H$ defined by $(g, h)(u, v) = (gu, hv)$. Given two representations U and V of G, the (inner) tensor product, also denoted $U \otimes V$, is the representation of G defined by $g(u, v) = (gu, gv)$.

In these lectures we will describe representations in any of the following ways:

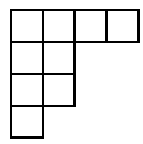
- giving the character
- giving an isomorphic representation

- giving the multiplicity of each irreducible
- using operations such as restriction, induction and tensor product.

2.2. Representations of the Symmetric Group

In this section, we construct the irreducible representations of the symmetric group $\mathfrak{S}_n$, which are called Specht modules and are denoted by S^λ, where λ is a partition of n. We also discuss skew shaped Specht modules.

Let λ be a partition of n, i.e., a weakly decreasing sequence of positive integers $\lambda = (\lambda_1 \geq \cdots \geq \lambda_k)$ whose sum is n. We write $\lambda \vdash n$ (or $|\lambda| = n$) and say that the length $l(\lambda)$ is k. We will also write $\lambda = 1^{m_1}2^{m_2}\cdots n^{m_n}$, if λ has m_i parts of size i for each i. Each partition λ is identified with a *Young (or Ferrers) diagram* whose ith row has λ_i cells. For example, the partition $(4,2,2,1) \vdash 9$ is identified with the Young diagram



A *Young tableau* of shape $\lambda \vdash n$ is a filling of the Young diagram corresponding to λ, with distinct positive integers in [n]. A Young tableau is said to be *standard* if the entries increase along each row and column. For example, the Young tableau on the left is not standard, while the one on the right is.

$$\begin{array}{|c|c|c|c|}\hline 8 & 2 & 4 & 1 \\ \hline 7 & 5 \\ \cline{1-2} 9 & 3 \\ \cline{1-2} 6 \\ \cline{1-1}\end{array} \qquad\qquad \begin{array}{|c|c|c|c|}\hline 1 & 2 & 4 & 7 \\ \hline 3 & 6 \\ \cline{1-2} 5 & 8 \\ \cline{1-2} 9 \\ \cline{1-1}\end{array}$$

Let $\mathcal{T}_\lambda$ be the set of Young tableaux of shape λ and let M^λ be the complex vector space generated by elements of $\mathcal{T}_\lambda$. The symmetric group $\mathfrak{S}_n$ acts on M^λ by permuting entries of the Young tableaux. That is, for transposition $\sigma = (i,j) \in \mathfrak{S}_n$, the tableau σT is obtained from T by switching entries i and j. For example,

$$(2,3)\ \begin{array}{|c|c|c|c|}\hline 8 & \mathbf{2} & 4 & 1 \\ \hline 7 & 5 \\ \cline{1-2} 9 & \mathbf{3} \\ \cline{1-2} 6 \\ \cline{1-1}\end{array} \quad = \quad \begin{array}{|c|c|c|c|}\hline 8 & \mathbf{3} & 4 & 1 \\ \hline 7 & 5 \\ \cline{1-2} 9 & \mathbf{2} \\ \cline{1-2} 6 \\ \cline{1-1}\end{array}\ .$$

The representation that we have described is clearly the left regular representation of $\mathfrak{S}_n$. One can also let $\mathfrak{S}_n$ act as the right regular representation on M^λ. That is for transposition $\sigma = (i,j) \in \mathfrak{S}_n$ and $T \in \mathcal{T}_\lambda$, the tableau $T\sigma$ is obtained from T by switching the contents of the ith and jth cell under some fixed ordering of the cells of λ.

We will say that two tableaux in $\mathcal{T}_\lambda$ are *row-equivalent* if they have the same sequence of row sets. For example, the tableaux

$$\begin{array}{|c|c|c|c|}\hline 8 & 2 & 4 & 1 \\ \hline 7 & 5 \\ \cline{1-2} 9 & 3 \\ \cline{1-2} 6 \\ \cline{1-1}\end{array} \qquad \text{and} \qquad \begin{array}{|c|c|c|c|}\hline 1 & 2 & 4 & 8 \\ \hline 5 & 7 \\ \cline{1-2} 3 & 9 \\ \cline{1-2} 6 \\ \cline{1-1}\end{array}$$

are row-equivalent. *Column-equivalent* is defined similarly. For shape λ, the row stabilizer R_λ is defined to be the subgroup

$$R_\lambda := \{\sigma \in \mathfrak{S}_n : T\sigma \text{ and } T \text{ are row-equivalent for all } T \in \mathcal{T}_\lambda\}.$$

Similarly, the column stabilizer C_λ is defined to be the subgroup

$$C_\lambda := \{\sigma \in \mathfrak{S}_n : T\sigma \text{ and } T \text{ are column-equivalent for all } T \in \mathcal{T}_\lambda\}.$$

We now give two characterizations of the Specht module S^λ; one as a subspace of M^λ generated by certain signed sums of Young tableaux called polytabloids; and the other as a quotient of M^λ by certain relations called row relations and Garnir relations. We caution the reader that our notions of polytabloids and Garnir relations are dual to the usual notions given in standard texts such as [**151**].

We begin with the submodule characterization. For each $T \in \mathcal{T}_\lambda$, define the *polytabloid* of shape λ,

$$e_T := \sum_{\alpha \in R_\lambda} \sum_{\beta \in C_\lambda} \operatorname{sgn}(\beta)\, T\alpha\beta.$$

For example if

$$T = \begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 \\ \cline{1-1}\end{array} \quad \text{then} \quad e_T = \left(\begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 \\ \cline{1-1}\end{array} - \begin{array}{|c|c|}\hline 3 & 2 \\ \hline 1 \\ \cline{1-1}\end{array} \right) + \left(\begin{array}{|c|c|}\hline 2 & 1 \\ \hline 3 \\ \cline{1-1}\end{array} - \begin{array}{|c|c|}\hline 3 & 1 \\ \hline 2 \\ \cline{1-1}\end{array} \right).$$

Since the left and right action of $\mathfrak{S}_n$ on $\mathcal{T}_\lambda$ commute, we have

$$\pi e_T = e_{\pi T}, \tag{2.2.1}$$

for all $T \in \mathcal{T}_\lambda$ and $\pi \in \mathfrak{S}_n$. We can now define the *Specht module* S^λ to be the subspace of M^λ given by

$$S^\lambda := \langle e_T : T \in \mathcal{T}_\lambda \rangle.$$

It follows from (2.2.1) that S^λ is an $\mathfrak{S}_n$-submodule of M^λ (under the left action).

Theorem 2.2.1. *The Specht modules S^λ for all $\lambda \vdash n$ form a complete set of irreducible $\mathfrak{S}_n$-modules.*

A polytabloid e_T is said to be a *standard polytabloid* if T is a standard Young tableau. We will see shortly that the standard polytabloids of shape λ form a basis for the Specht module S^λ.

Now we give the quotient characterization. The *row relations* are defined for all $T \in T_\lambda$ and $\sigma \in R_\lambda$ by

$$r_\sigma(T) := T\sigma - T. \tag{2.2.2}$$

For all i, j such that $1 \le j \le \lambda_i$, let $C_{i,j}(\lambda)$ be the set of cells in columns j through λ_i of row i and in columns 1 through j of row $i+1$. Let $G_{i,j}(\lambda)$ be the subgroup of $\mathfrak{S}_n$ consisting of permutations σ that fix all entries of the cells that are not in $C_{i,j}(\lambda)$ under the right action of σ on tableaux of shape λ. The *Garnir relations* are defined for all i, j such that $1 \le j \le \lambda_i$ and for all $T \in \mathcal{T}_\lambda$ by

$$g_{i,j}(T) := \sum_{\sigma \in G_{i,j}(\lambda)} T\sigma. \tag{2.2.3}$$

For example if

$$T = \begin{array}{|c|c|c|c|}\hline 7 & \mathbf{1} & \mathbf{5} & \mathbf{10} \\ \hline \mathbf{3} & \mathbf{4} & 2 \\ \cline{1-3} 9 & 8 \\ \cline{1-2} 11 & 6 \\ \cline{1-2}\end{array} \quad \text{and} \quad (i,j) = (1,2)$$

then the entries $1, 5, 10, 3, 4$ are permuted while the remaining entries are fixed. So

$$g_{1,2}(T) = \begin{array}{|c|c|c|c|}\hline 7 & \mathbf{1} & \mathbf{3} & \mathbf{4} \\ \hline \mathbf{5} & \mathbf{10} & 2 \\ \cline{1-3} 9 & 8 \\ \cline{1-2} 11 & 6 \\ \cline{1-2}\end{array} + \begin{array}{|c|c|c|c|}\hline 7 & \mathbf{1} & \mathbf{3} & \mathbf{4} \\ \hline \mathbf{10} & \mathbf{5} & 2 \\ \cline{1-3} 9 & 8 \\ \cline{1-2} 11 & 6 \\ \cline{1-2}\end{array} + \begin{array}{|c|c|c|c|}\hline 7 & \mathbf{1} & \mathbf{3} & \mathbf{5} \\ \hline \mathbf{4} & \mathbf{10} & 2 \\ \cline{1-3} 9 & 8 \\ \cline{1-2} 11 & 6 \\ \cline{1-2}\end{array} + \begin{array}{|c|c|c|c|}\hline 7 & \mathbf{1} & \mathbf{3} & \mathbf{5} \\ \hline \mathbf{10} & \mathbf{4} & 2 \\ \cline{1-3} 9 & 8 \\ \cline{1-2} 11 & 6 \\ \cline{1-2}\end{array} + \dots$$

Again, since the left and right action of $\mathfrak{S}_n$ on $\mathcal{T}_\lambda$ commute, we have

$$\begin{aligned} \pi r_\sigma(T) &= r_\sigma(\pi T) \\ \pi g_{i,j}(T) &= g_{i,j}(\pi T), \end{aligned}$$

for all $\pi \in \mathfrak{S}_n$. Consequently, the subspace U^λ of M^λ generated by the row relations (2.2.2) and the Garnir relations (2.2.3) is an $\mathfrak{S}_n$-submodule of M^λ.

Theorem 2.2.2. *For all $\lambda \vdash n$,*

$$S^\lambda \cong_{\mathfrak{S}_n} M^\lambda / U^\lambda.$$

Now we can view the Specht module S^λ as the module generated by tableaux of shape λ subject to the row and Garnir relations.

Exercise 2.2.3. Prove Theorem 2.2.2 by first showing that,
(a) $U^\lambda \subseteq \ker \psi$, where $\psi : M^\lambda \to S^\lambda$ is defined by $\psi(T) = e_T$,
(b) the standard polytabloids e_T are linearly independent,
(c) the standard tableaux span M^λ / U^λ.

We have the following consequence of Exercise 2.2.3.

Corollary 2.2.4. *The standard polytabloids of shape λ form a basis for S^λ. The standard tableaux of shape λ form a basis for M^λ / U^λ. Consequently* $\dim S^\lambda$ *is equal to the number of standard tableaux of shape λ.*

There is a remarkable formula for the number of standard tableaux of a fixed shape λ.

Theorem 2.2.5 (Frame-Robinson-Thrall hook length formula)**.** *For all $\lambda \vdash n$,*

$$\dim S^\lambda = \frac{n!}{\prod_{x \in \lambda} h_x},$$

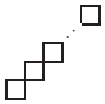
where the product is taken over all cells x in the Young diagram λ, and h_x is the number of cells in the hook formed by x, which consists of x, the cells that are below x in the same column, and the cells to the right of x in the same row.

One can generalize Specht modules to skew shapes. By removing a smaller skew diagram μ from the northwest corner of a skew diagram λ, one gets a *skew* diagram denoted by λ/μ. For example if $\lambda = (4, 3, 3)$ and $\mu = (2, 1)$ then

$$\lambda/\mu = \begin{array}{|c|c|c|c|}\hline & & & \\ \hline & & \\ \cline{1-3} & & \\ \cline{1-3}\end{array}.$$

Skew Specht modules $S^{\lambda/\mu}$ are defined analogously to "straight" Specht modules. There is a submodule characterization and a quotient characterization. Theorem 2.2.2 and Corollary 2.2.4 hold in the skew setting. There is a classical combinatorial rule for decomposing Specht modules of skew shape into irreducible straight shape Specht modules called the Littlewood-Richardson rule, which we will not present here.

Example 2.2.6. Some important classes of skew and straight Specht modules are listed below.

$\lambda/\mu =$  $S^{\lambda/\mu} =$ regular representation.

$\lambda \quad =$ $S^{\lambda} =$ trivial representation

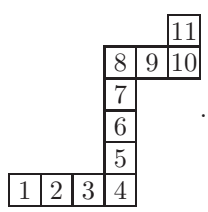
$\lambda \quad =$ $S^{\lambda} =$ sign representation

where the sign representation is the 1-dimensional representation $V = \mathbb{C}$ whose character is $\mathrm{sgn}(\sigma)$, that is $\sigma z = \mathrm{sgn}(\sigma)z$ for all $\sigma \in \mathfrak{S}_n$ and $z \in \mathbb{C}$. We denote the sign representation by sgn_n or $S^{(1^n)}$ and we denote the trivial representation by $1_{\mathfrak{S}_n}$ or $S^{(n)}$.

Exercise 2.2.7. For skew or straight shape λ/μ, let $\chi^{\lambda/\mu}$ denote the character of the representation $S^{\lambda/\mu}$, and for $\sigma \in \mathfrak{S}_n$, let $f(\sigma)$ denote the number of fixed points of σ.

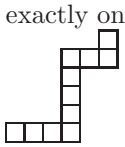
(a) Show $\chi^{(n-1,1)}(\sigma) = f(\sigma) - 1$ for all $\sigma \in \mathfrak{S}_n$.
(b) Show $\chi^{(n,1)/(1)}(\sigma) = f(\sigma)$ for all $\sigma \in \mathfrak{S}_n$.
(c) Find the character of each of the representations in Example 2.2.6.

A *skew hook* is a connected skew diagram that does not contain the subdiagram $(2,2)$. Each cell of a skew hook, except for the southwestern most and northeastern most end cells, has exactly two cells adjacent to it. Each of the end cells has only one cell adjacent to it. Let H be a skew hook with n cells. We label the cells of H with numbers 1 through n, starting at the southwestern end cell, moving through the adjacent cells, and ending at the northeastern end cell. For example, we have the labeled skew hook



.

If cell $i+1$ is above cell i in H then we say that the skew hook H has a *descent* at i. Let $\mathrm{des}(H)$ denote the set of descents of H. For each subset S of $[n-1]$, there is exactly one skew hook with n cells and descent set S. For example, the skew hook



is the only skew hook with 11 cells and descent set $\{4,5,6,7,10\}$.

The Specht modules of skew hook shape are called *Foulkes representations*. Note that for any skew hook H with n cells, the set of standard tableaux of shape H corresponds bijectively to the set of permutations in $\mathfrak{S}_n$ with descent set $\mathrm{des}(H)$. (The descent set $\mathrm{des}(\sigma)$ of a permutation $\sigma \in \mathfrak{S}_n$ is the set of all $i \in [n-1]$ such

that $\sigma(i) > \sigma(i+1)$.) Indeed, by listing the entries of cells 1 through n, one gets a permutation with descent set $\mathrm{des}(H)$. Hence by Corollary 2.2.4 for skew shapes, the dimension of the Foulkes representation S^H is the number of permutations in $\mathfrak{S}_n$ with descent set $\mathrm{des}(H)$.

A *descent* of a standard Young tableau is an entry i that is in a higher row than $i+1$. By applying the Littlewood-Richardson rule mentioned above, one gets the following decomposition of the Foulkes representation into irreducibles,

$$S^H = \bigoplus_{\lambda \vdash n} c_{H,\lambda} S^\lambda, \tag{2.2.4}$$

where $c_{H,\lambda}$ is the number of standard Young tableaux of shape λ and descent set $\mathrm{des}(H)$.

Exercise 2.2.8. Use (2.2.4) to show that the regular representation of $\mathfrak{S}_n$ decomposes into Foulkes representations as follows:

$$\mathbb{C}\mathfrak{S}_n \cong_{\mathfrak{S}_n} \bigoplus_{H \in \mathrm{SH}_n} S^H,$$

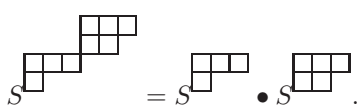
where SH_n is the set of skew hooks with n cells.

The *induction product* of an $\mathfrak{S}_j$-module U and an $\mathfrak{S}_k$-module V is the $\mathfrak{S}_{j+k}$-module

$$U \bullet V := (U \otimes V) \uparrow_{\mathfrak{S}_j \times \mathfrak{S}_k}^{\mathfrak{S}_{j+k}} .$$

(We are viewing $\mathfrak{S}_j \times \mathfrak{S}_k$ as the subgroup of $\mathfrak{S}_{j+k}$ consisting of permutations that stabilize the sets $\{1, 2 \ldots, j\}$ and $\{j+1, j+2, \ldots, j+k\}$.)

Exercise 2.2.9. If a skew shape D consists of two shapes λ and μ, where λ and μ have no rows or columns in common, we say that D is the *disjoint union* of λ and μ. Show that $S^D = S^\lambda \bullet S^\mu$ if D is the disjoint union of λ and μ. For example

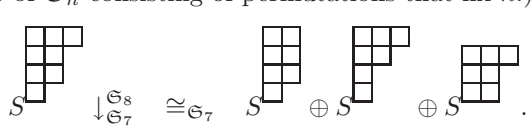


Exercise 2.2.10. Let $\lambda \vdash n$.

(a) Show that

$$S^\lambda \downarrow_{\mathfrak{S}_{n-1}}^{\mathfrak{S}_n} \cong_{\mathfrak{S}_{n-1}} \bigoplus_{\mu} S^\mu$$

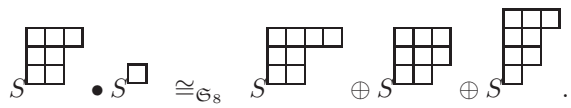
summed over all Young diagrams μ obtained from λ by removing a cell from the end of one of the rows of λ. (We are viewing $\mathfrak{S}_{n-1}$ as the subgroup of $\mathfrak{S}_n$ consisting of permutations that fix n.) For example,



(b) Show that

$$S^\lambda \uparrow_{\mathfrak{S}_n}^{\mathfrak{S}_{n+1}} = S^\lambda \bullet S^{(1)} \cong_{\mathfrak{S}_{n+1}} \bigoplus_{\mu} S^\mu$$

summed over all Young diagrams μ obtained from λ by adding a cell to the end of one of the rows of λ. For example,



There is an important generalization of Exercise 2.2.10 (b) known as Pieri's rule.

Theorem 2.2.11 (Pieri's rule)**.** *Let $m, n \in \mathbb{Z}^+$. If $\lambda \vdash n$ then*

$$S^\lambda \bullet S^{(m)} \cong_{\mathfrak{S}_{m+n}} \bigoplus_\mu S^\mu,$$

summed over all partitions μ of $m+n$ such that μ contains λ and the skew shape μ/λ has at most one cell in each column. Similarly

$$S^\lambda \bullet S^{(1^m)} \cong_{\mathfrak{S}_{m+n}} \bigoplus_\mu S^\mu,$$

summed over all partitions μ of $m+n$ such that μ contains λ and the skew shape μ/λ has at most one cell in each row.

The conjugate of a partition λ is the partition λ' whose Young diagram is the transpose of that of λ.

Theorem 2.2.12. *For all partitions $\lambda \vdash n$,*

$$S^\lambda \otimes \mathrm{sgn}_n \cong_{\mathfrak{S}_n} S^{\lambda'},$$

where the tensor product is an inner tensor product. This also holds for skew diagrams.

Exercise 2.2.13. Let V be a representation of $\mathfrak{S}_n$. Show that

$$(V \otimes \mathrm{sgn}_n) \uparrow_{\mathfrak{S}_n}^{\mathfrak{S}_{n+1}} \cong_{\mathfrak{S}_{n+1}} V \uparrow_{\mathfrak{S}_n}^{\mathfrak{S}_{n+1}} \otimes \mathrm{sgn}_{n+1},$$

and

$$(V \otimes \mathrm{sgn}_n) \downarrow_{\mathfrak{S}_{n-1}}^{\mathfrak{S}_n} \cong_{\mathfrak{S}_{n-1}} V \downarrow_{\mathfrak{S}_{n-1}}^{\mathfrak{S}_n} \otimes \mathrm{sgn}_{n-1}.$$

2.3. Group Actions on Poset (Co)homology

Let G be a finite group. A G-simplicial complex is a simplicial complex together with an action of G on its vertices that takes faces to faces. A G-poset is a poset together with a G-action on its elements that preserves the partial order; i.e., $x < y \Rightarrow gx < gy$. So if P is a G-poset then its order complex $\Delta(P)$ is a G-simplicial complex and if Δ is a G-simplicial complex then its face poset $P(\Delta)$ is a G-poset.

A *G-space* is a topological space on which G acts as a group of homeomorphisms. If Δ is a G-simplicial complex then the geometric realization $\|\Delta\|$ is a G-space under the natural induced action of G.

Example 2.3.1. The subset lattice B_n is an $\mathfrak{S}_n$-poset. The action of a permutation $\sigma \in \mathfrak{S}_n$ on a subset $\{a_1, \ldots, a_k\}$ is given by

$$\sigma\{a_1, \ldots, a_k\} = \{\sigma(a_1), \ldots, \sigma(a_k)\}. \tag{2.3.1}$$

Example 2.3.2. The partition lattice Π_n is an $\mathfrak{S}_n$-poset. The action of a permutation $\sigma \in \mathfrak{S}_n$ on a partition $\{B_1, \ldots, B_k\}$ is given by

$$\sigma\{B_1, \ldots, B_k\} = \{\sigma B_1, \ldots, \sigma B_k\}, \tag{2.3.2}$$

where σB_i is defined in (2.3.1). The symmetric group is isomorphic to the group generated by reflections about hyperplanes in the braid arrangement. The action described here is simply the action of the reflection group on intersections of hyperplanes in the braid arrangement.

Example 2.3.3. The face lattice C_n of the n-cross-polytope (or the face lattice of the n-cube) is an $\mathfrak{S}_n[\mathbb{Z}_2]$-poset. The wreath product group $\mathfrak{S}_n[\mathbb{Z}_2]$ is also known as the hyperoctahedral group or the type B Coxeter group (see Section 2.4 for the definition of wreath product). It is the group generated by reflections about the hyperplanes in the type B braid arrangement.

By viewing the n-cross-polytope as the convex hull of the points $\pm e_i$, $i = 1, \ldots, n$, one obtains the action of $\mathfrak{S}_n[\mathbb{Z}_2]$ on C_n. We describe this action in combinatorial terms. The $(k-1)$-dimensional faces of the cross-polytope are convex hulls of certain k element subsets of $\{\pm e_i : i \in [n]\}$. These are the ones that don't contain both e_i and $-e_i$ for any i. Thus the $(k-1)$-faces can be identified with k-subsets T of $[n] \cup \{\bar{i} : i \in [n]\}$ such that $\{i, \bar{i}\} \not\subseteq T$ for all i. By ordering these sets by containment, one gets the face poset of the n-cross-polytope. For example, the 4-subset $\{3, \bar{5}, 6, \bar{8}\}$ is identified with the convex hull of the points $e_3, -e_5, e_6, -e_8$. The elements of $\mathfrak{S}_n[\mathbb{Z}_2]$ are identified with permutations σ of $[n] \cup \{\bar{i} : i \in [n]\}$ for which $\overline{\sigma(i)} = \sigma(\bar{i})$ for all i (where $\bar{\bar{a}} = a$). Then the action of a permutation $\sigma \in \mathfrak{S}_n[\mathbb{Z}_2]$ on a k-subset $\{a_1, \ldots, a_k\}$ of $[n] \cup \{\bar{i} : i \in [n]\}$ is given by (2.3.1).

Example 2.3.4. The type B partition lattice Π_n^B is also an $\mathfrak{S}_n[\mathbb{Z}_2]$-poset. Recall from Example 1.3.4 that the type B partition lattice Π_n^B is the intersection lattice of the type B braid arrangement. Since this arrangement is invariant under reflection about any hyperplane in the arrangement, elements of the reflection group $\mathfrak{S}_n[\mathbb{Z}_2]$ map intersections of hyperplanes to intersections of hyperplanes. This gives the action of $\mathfrak{S}_n[\mathbb{Z}_2]$ on Π_n^B. There is a combinatorial description of the action analogous to (2.3.2) (see e.g. [**83**]).

Example 2.3.5. The lattice of subspaces of an n-dimensional vector space over a finite field F is a $GL_n(F)$-poset, where the general linear group $GL_n(F)$ acts in the obvious way.

Example 2.3.6. The lattice of subgroups of a finite group G ordered by inclusion is a G-poset, where G acts by conjugation.

Example 2.3.7. The semilattice of p-subgroups of a group G ordered by inclusion is a G-poset, where G acts by conjugation.

Let P be a G-poset. Since $g \in G$ takes j-chains to j-chains, g acts as a linear map on $C_j(P; \mathbb{C})$. It is easy to see that

$$g\partial(c) = \partial(gc) \text{ and } g\delta(c) = \delta(gc).$$

Hence g acts as a linear map on $\tilde{H}_j(P; \mathbb{C})$ and on $\tilde{H}^j(P; \mathbb{C})$. This means that whenever P is a G-poset, $\tilde{H}_j(P; \mathbb{C})$ and $\tilde{H}^j(P; \mathbb{C})$ are G-modules. The bilinear form (1.5.1), induces a pairing between $\tilde{H}_j(P; \mathbb{C})$ and $\tilde{H}^j(P; \mathbb{C})$, which allows one

to view them as dual G-modules. For $G = \mathfrak{S}_n$ or $G = \mathfrak{S}_n[\mathbb{Z}_2]$,

$$\tilde{H}_j(P, \mathbb{C}) \cong_G \tilde{H}^j(P, \mathbb{C})$$

since dual $\mathfrak{S}_n$-modules (resp., dual $\mathfrak{S}_n[\mathbb{Z}_2]$-modules) are isomorphic.

Given a G-simplicial complex Δ, the natural homeomorphism from Δ to its barycentric subdivision $\Delta(P(\Delta))$ commutes with the G-action. Consequently, for all $j \in \mathbb{Z}$

$$\tilde{H}_j(\Delta; \mathbb{C}) \cong_G \tilde{H}_j(P(\Delta); \mathbb{C}),$$

and

$$\tilde{H}^j(\Delta; \mathbb{C}) \cong_G \tilde{H}^j(P(\Delta); \mathbb{C}).$$

Exercise 2.3.8. The maximal chains of $\bar{B}_n$ correspond bijectively to tableaux of shape 1^n via the map

$$\begin{array}{|c|}\hline t_1 \\ \hline t_2 \\ \hline \vdots \\ \hline t_n \\ \hline \end{array} \quad \mapsto \quad (\{t_1\} \subset \{t_1, t_2\} \subset \cdots \subset \{t_1, \ldots, t_{n-1}\}).$$

(a) Show that the Garnir relations map to the coboundary relations. Consequently, the representation of $\mathfrak{S}_n$ on the top cohomology $\tilde{H}^{n-2}(\bar{B}_n; \mathbb{C})$ is the sign representation.
(b) Show that the polytabloids map to cycles in top homology.

There is an equivariant version of the Euler-Poincaré formula (Theorem 1.2.8), known as the Hopf-trace formula. It is convenient to state this formula in terms of virtual representations, which we define first. The *representation group* $\mathcal{G}(G)$ of a group G is the free abelian group on the set of all isomorphism classes $[V]$ of G-modules V modulo the subgroup generated by all $[V \oplus W] - [V] - [W]$. Elements of the representation group are called *virtual representations*. If V is an actual representation, we denote the virtual representation $[V]$ by V. Note that two virtual representations $A - B$ and $C - D$, where A, B, C, D are actual representations of G, are equal in the representation group if and only if $A \oplus D \cong_G B \oplus C$ in the usual sense. We will write $A - B \cong_G C - D$.

Theorem 2.3.9 (Hopf trace formula)**.** *For any G-simplicial complex Δ,*

$$\bigoplus_{i=-1}^{\dim \Delta} (-1)^i C_i(\Delta; \mathbb{C}) \cong_G \bigoplus_{i=-1}^{\dim \Delta} (-1)^i \tilde{H}_i(\Delta; \mathbb{C})$$

We will usually suppress the $\mathbb{C}$ from our notation $\tilde{H}_i(P; \mathbb{C})$ (resp., $\tilde{H}^i(P; \mathbb{C})$) and write $\tilde{H}_i(P)$ (resp., $\tilde{H}^i(P)$) instead when viewing (co)homology as a G-module.

2.4. Symmetric Functions, Plethysm, and Wreath Product Modules

Symmetric functions provide a convenient way of describing and computing representations of the symmetric group. In this section we give the basics of symmetric function theory. Then we demonstrate its use in computing homology of an interesting example known as the matching complex.

2.4.1. Symmetric functions

Let $x = (x_1, x_2, \dots)$ be an infinite sequence of indeterminates. A *homogeneous symmetric function of degree* n is a formal power series $f(x) \in \mathbb{Q}[[x]]$ in which each term has degree n and $f(x_{\sigma(1)}, x_{\sigma(2)}, \dots) = f(x_1, x_2, \dots)$ for all permutations σ of $\mathbb{Z}^+$.

Let Λ^n denote the set of homogeneous symmetric functions of degree n and let $\Lambda = \bigoplus_{n\geq 0} \Lambda^n$. Then Λ is a graded $\mathbb{Q}$-algebra, since $\alpha f(x) + \beta g(x) \in \Lambda^n$ if $f(x), g(x) \in \Lambda^n$ and $\alpha, \beta \in \mathbb{Q}$; and $f(x)g(x) \in \Lambda_{m+n}$ if $f(x) \in \Lambda^m$ and $g(x) \in \Lambda^n$.

There are several important bases for the vector space Λ^n. We mention just two of them here; the basis of power sum symmetric functions and the basis of Schur functions. These bases are indexed by partitions of n. For $n \geq 1$, let

$$p_n = \sum_{i\geq 1} x_i^n$$

and let $p_0 = 1$. The *power sum symmetric function* indexed by $\lambda = (\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k)$ is defined by

$$p_\lambda := p_{\lambda_1} p_{\lambda_2} \dots p_{\lambda_k}.$$

Given a Young diagram λ, a *semistandard tableau* of shape λ is a filling of λ with positive integers so that the rows weakly increase and the columns strictly increase. Let SS_λ be the set of semistandard tableaux of shape λ. Given a semistandard tableau T, define $w(T) := x_1^{m_1} x_2^{m_2} \cdots$, where for each i, m_i is the number of times i appears in T. Now define the Schur function indexed by λ to be

$$s_\lambda := \sum_{T\in SS_\lambda} w(T).$$

Skew shaped Schur functions s_D are defined analogously for all skew diagrams D.

While it is obvious that the power sum symmetric functions are symmetric functions, it is not obvious that the Schur functions defined this way are.

Theorem 2.4.1. *The sets $\{p_\lambda : \lambda \vdash n\}$ and $\{s_\lambda : \lambda \vdash n\}$ form bases for Λ^n. Moreover, $\{s_\lambda : \lambda \vdash n\}$ is an integral basis, i.e., a basis for the $\mathbb{Z}$-module $\Lambda^n_{\mathbb{Z}}$ of homogeneous symmetric functions of degree n with integer coefficients.*

The Schur function $s_{(n)}$ is known as the complete homogeneous symmetric function of degree n and is denoted by h_n. The Schur function $s_{(1^n)}$ is known as the elementary symmetric function of degree n and is denoted by e_n. There is an important involution $\omega : \Lambda_n \to \Lambda_n$ defined by

$$\omega(s_\lambda) = s_{\lambda'},$$

where λ' is the conjugate of λ. Clearly

$$\omega(h_n) = e_n.$$

For the power sum symmetric functions we have

$$\omega(p_\lambda) = (-1)^{|\lambda| - l(\lambda)} p_\lambda. \tag{2.4.1}$$

Exercise 2.4.2. Prove

$$\sum_{n\geq 0} h_n = \prod_{i\geq 1} (1 - x_i)^{-1},$$

and

$$\sum_{n\geq 0} e_n = \prod_{i\geq 1}(1+x_i),$$

where $h_0 = e_0 = 1$.

The *Frobenius characteristic* $\mathrm{ch}(V)$ of a representation V of $\mathfrak{S}_n$ is the symmetric function given by

$$\mathrm{ch}(V) := \sum_{\mu\vdash n} \frac{1}{z_\mu}\chi^V(\mu)\, p_\mu,$$

where $z_\mu := 1^{m_1}m_1!2^{m_2}m_2!\dots n^{m_n}m_n!$ for $\mu = 1^{m_1}2^{m_2}\dots n^{m_n}$ and $\chi^V(\mu)$ is the character $\chi^V(\sigma)$ for $\sigma \in \mathfrak{S}_n$ of conjugacy type μ. Some basic facts on Frobenius characteristic are compiled in the next result.

Theorem 2.4.3.

(a) *For all (skew or straight) shapes λ,*

$$\mathrm{ch}(S^\lambda) = s_\lambda.$$

(b) *For all representations V of $\mathfrak{S}_n$,*

$$\omega(\mathrm{ch}V) = \mathrm{ch}(V \otimes \mathrm{sgn}_n).$$

(c) *For all representations U, V of $\mathfrak{S}_n$,*

$$\mathrm{ch}(U \oplus V) = \mathrm{ch}(U) + \mathrm{ch}(V)$$

(d) *For all representations U of $\mathfrak{S}_m$ and V of $\mathfrak{S}_n$,*

$$\mathrm{ch}(U \bullet V) = \mathrm{ch}(U)\ \mathrm{ch}(V).$$

The direct sum $\bigoplus_{n\geq 0}\mathcal{G}(\mathfrak{S}_n)$ of representation groups is a ring under the induction product. It follows from Theorem 2.4.3 that the Frobenius characteristic map is an isomorphism from the ring $\bigoplus_{n\geq 0}\mathcal{G}(\mathfrak{S}_n)$ to the ring of symmetric functions over $\mathbb{Z}$.

Definition 2.4.4. Let $f \in \Lambda$ and let g be a formal power series with positive integer coefficients. Choose any ordering of the monomials of g, where a monomial appears in the ordering m_i times if its coefficient is m_i. For example, if $g = 3y_1y_2^2 + 2y_2y_3 + \dots$ then the monomials can be arranged as

$$(y_1y_2^2, y_1y_2^2, y_1y_2^2, y_2y_3, y_2y_3, \dots).$$

Pad the sequence of monomials with zero's if g has a finite number of terms. Define the *plethysm* of f and g, denoted $f[g]$, to be the formal power series obtained from f by replacing the indeterminate x_i with the ith monomial of g for each i. Since f is a symmetric function, the chosen order of the monomials doesn't matter. For example, if $f = \sum_{n\geq 0} e_n = \prod_{i\geq 1}(1+x_i)$ and g is as above then

$$f[g] = (1+y_1y_2^2)(1+y_1y_2^2)(1+y_1y_2^2)(1+y_2y_3)(1+y_2y_3)\cdots.$$

The following proposition is immediate.

Proposition 2.4.5. *Suppose $f, g \in \Lambda$ and h is a formal power series with positive integer coefficients. Then*

- *If f has positive integer coefficients then $f[p_n] = p_n[f]$ is obtained by replacing each indeterminate x_i of f by x_i^n.*

- $(af + bg)[h] = af[h] + bg[h]$, *where* $a, b \in \mathbb{Q}$
- $fg[h] = f[h]g[h]$.

Note that if $g \in \Lambda$ has positive integer coefficients then $f[g] \in \Lambda$. One can extend the definition of plethysm to all $g \in \Lambda$, by using Proposition 2.4.5 and the fact that the power sum symmetric functions form a basis for Λ. Hence plethysm is a binary operation on Λ, which is clearly associative, but not commutative. Note that the plethystic identity is $p_1 = h_1$. We say that $f, g \in \Lambda$ are *plethystic inverses* of each other, and write $g = f^{[-1]}$, if $f[g] = g[f] = h_1$.

2.4.2. Composition product and wreath product

Our purpose for introducing plethysm in these lectures is that plethysm encodes a product operation on symmetric group representations called composition product, which is described below. (This description is based on an exposition given in [**180**].)

Let G be a finite group. The *wreath product* of $\mathfrak{S}_m$ and G, denoted by $\mathfrak{S}_m[G]$, is defined to be the set of $(m+1)$-tuples $(g_1, g_2, \ldots, g_m; \tau)$ such that $g_i \in G$ and $\tau \in \mathfrak{S}_m$ with multiplication given by

$$(g_1, \ldots, g_m; \tau)(h_1, \ldots, h_m; \gamma) = (g_1 h_{\tau^{-1}(1)}, \ldots, g_m h_{\tau^{-1}(m)}; \tau\gamma).$$

The following proposition is immediate.

Proposition 2.4.6. *The map* $(\alpha_1, \alpha_2, \ldots, \alpha_m; \tau) \mapsto \tau$ *is a homomorphism from* $\mathfrak{S}_m[\mathfrak{S}_n]$ *onto* $\mathfrak{S}_m$.

Definition 2.4.7. Let V be an $\mathfrak{S}_m$-module and W be a G-module. Then the *wreath product* of V with W, denoted $V[W]$, is the inner tensor product of two $\mathfrak{S}_m[G]$-modules:

$$V[W] = \widetilde{W^{\otimes m}} \otimes \hat{V},$$

where $\widetilde{W^{\otimes m}}$ is the vector space $W^{\otimes m}$ with $\mathfrak{S}_m[G]$ action given by

$$(\alpha_1, \ldots, \alpha_m; \tau)(w_1 \otimes \cdots \otimes w_m) = \alpha_1 w_{\tau^{-1}(1)} \otimes \cdots \otimes \alpha_m w_{\tau^{-1}(m)} \tag{2.4.2}$$

and $\hat{V}$ is the pullback of the representation of $\mathfrak{S}_m$ on V to $\mathfrak{S}_m[G]$ through the homomorphism given in Proposition 2.4.6. That is, $\hat{V}$ is the representation of $\mathfrak{S}_m[G]$ on V defined by

$$(\alpha_1, \ldots, \alpha_m; \tau)v = \tau v.$$

Given a finite set $A = \{a_1 < a_2 < \cdots < a_n\}$ of positive integers, let $\mathfrak{S}_A$ be the set of permutations of the set A. We shall view a permutation in $\mathfrak{S}_A$ as a word whose letters come from A. For $\sigma \in S_n$, let σ^A denote the word $a_{\sigma(1)} a_{\sigma(2)} \cdots a_{\sigma(n)}$. We shall view an element of the Young subgroup $\mathfrak{S}_k \times \mathfrak{S}_{n-k}$ of $\mathfrak{S}_n$ as the concatenation $\alpha \star \beta$ of words $\alpha \in \mathfrak{S}_k$ and $\beta \in \mathfrak{S}_{\{k+1,\ldots,n\}}$. The wreath product $\mathfrak{S}_m[\mathfrak{S}_n]$ is isomorphic to the normalizer of the Young subgroup $\underbrace{\mathfrak{S}_n \times \cdots \times \mathfrak{S}_n}_{m \text{ times}}$ in $\mathfrak{S}_{mn}$. The isomorphism is given by

$$(\alpha_1, \ldots, \alpha_m; \tau) \quad \mapsto \quad \alpha_{\tau(1)}^{A_{\tau(1)}} \star \cdots \star \alpha_{\tau(m)}^{A_{\tau(m)}},$$

where $A_i = [in] \setminus [(i-1)n]$.

Define the *composition product* of an $\mathfrak{S}_m$-module V and an $\mathfrak{S}_n$-module W by

$$V \circ W := V[W] \uparrow_{\mathfrak{S}_m[\mathfrak{S}_n]}^{\mathfrak{S}_{mn}}$$

The following result relates plethysm to composition product.

Theorem 2.4.8. *Let V be an $\mathfrak{S}_m$-module and W be an $\mathfrak{S}_n$-module. Then*

$$\mathrm{ch}(V \circ W) = \mathrm{ch}V[\mathrm{ch}W].$$

Example 2.4.9. Given $\lambda \vdash n$, let $\Pi(\lambda)$ be the set of partitions of $[n]$ whose block sizes form the partition λ. In this example, we set $\lambda = d^b$, where d and b are positive integers.

(a) The symmetric group $\mathfrak{S}_{bd}$ acts on $\Pi(d^b)$ as in Example 2.3.2, and this action induces a representation of $\mathfrak{S}_{bd}$ on the complex vector space $\mathbb{C}\Pi(d^b)$ generated by elements of $\Pi(d^b)$. It is not difficult to see that the $\mathfrak{S}_{bd}$-module $\mathbb{C}\Pi(d^b)$ is the induction of the trivial representation of the stabilizer of the partition

$$\pi(d^b) := 1, \ldots, d \ / \ d+1, \ldots, 2d \ / \ \ldots \ / \ (b-1)d+1, \ldots, bd$$

to $\mathfrak{S}_{bd}$. The stabilizer of $\pi(d^b)$ is $\mathfrak{S}_b[\mathfrak{S}_d]$ and the trivial representation of $\mathfrak{S}_b[\mathfrak{S}_d]$ is $S^{(b)}[S^{(d)}]$. So $\mathbb{C}\Pi(d^b)$ is the composition product $S^{(b)} \circ S^{(d)}$. By Theorem 2.4.8

$$\mathrm{ch}\ \mathbb{C}\Pi(d^b) = h_b[h_d].$$

(b) The stabilizer $\mathfrak{S}_b[\mathfrak{S}_d]$ of $\pi(d^b)$ acts on the interval $(\pi(d^b), \hat{1})$ of Π_{bd}. This induces a representation of $\mathfrak{S}_b[\mathfrak{S}_d]$ on the top homology $\tilde{H}_{b-3}(\pi(d^b), \hat{1})$ (recall $\tilde{H}_i(x,y)$ denotes complex homology of the open interval (x,y)). By observing that the interval $(\pi(d^b), \hat{1})$ is isomorphic to the poset $\bar{\Pi}_b$, one can see that

$$\tilde{H}_{b-3}(\pi(d^b), \hat{1}) \cong_{\mathfrak{S}_b[\mathfrak{S}_d]} \tilde{H}_{b-3}(\bar{\Pi}_b)[S^{(d)}].$$

Next observe that $\bigoplus_{x \in \Pi(d^b)} \tilde{H}_{b-3}(x, \hat{1})$ is the induction of $\tilde{H}_{b-3}(\pi(d^b), \hat{1})$ from $\mathfrak{S}_b[\mathfrak{S}_d]$ to $\mathfrak{S}_{bd}$. So $\bigoplus_{x \in \Pi(d^b)} \tilde{H}_{b-3}(x, \hat{1})$ is the composition product $\tilde{H}_{b-3}(\bar{\Pi}_b) \circ S^{(d)}$. By Theorem 2.4.8,

$$\begin{aligned} \mathrm{ch} \bigoplus_{x \in \Pi(d^b)} \tilde{H}_{b-3}(x, \hat{1}) &= \mathrm{ch}\ \tilde{H}_{b-3}(\pi(d^b), \hat{1}) \uparrow_{\mathfrak{S}_b[\mathfrak{S}_d]}^{\mathfrak{S}_{bd}} \\ &= (\mathrm{ch}\ \tilde{H}_{b-3}(\bar{\Pi}_b))[h_d]. \end{aligned}$$

Theorem 2.4.8 is inadequate when λ is not of the form d^b. In [**177, 201**], more general results are given, which enable one to derive the following generating function,

(2.4.3)

$$\sum_{\lambda \in \mathrm{Par}(T,b)} \left(\mathrm{ch} \bigoplus_{x \in \Pi(\lambda)} \tilde{H}_{b-3}(x, \hat{1})\right) z_{\lambda_1} \cdots z_{\lambda_b} = (\mathrm{ch}\ \tilde{H}_{b-3}(\bar{\Pi}_b)) \left[\sum_{i \in T} z_i h_i\right],$$

where T is any set of positive integers, $\mathrm{Par}(T,b)$ is the set of partitions of length b, all of whose parts are in T, the λ_i are the parts of λ, and the z_i are (commuting) indeterminates.

2.4.3. The matching complex

In this subsection, we further demonstrate the power of symmetric function theory and plethysm in computing homology. Our example is a well-studied simplicial complex known as the matching complex, which is defined to be the simplicial complex M_n whose vertices are the 2-element subsets of $[n]$ and whose faces are

collections of mutually disjoint 2-element subsets of $[n]$. Alternatively, M_n is the simplicial complex of graphs on node set $[n]$ whose degree is at most 2. Its face poset is the proper part of the poset of partitions of $[n]$ whose block sizes are at most 2. It is not difficult to see that

$$\operatorname{ch} C_{k-1}(M_n) = e_k[h_2]h_{n-2k}. \tag{2.4.4}$$

It follows from this and the Hopf trace formula (Theorem 2.3.9) that

$$\sum_{k\geq -1}(-1)^{k-1}\operatorname{ch}\tilde{H}_k(M_n) = \sum_{k\geq 0}(-1)^k e_k[h_2]h_{n-2k},$$

which implies by Exercise 2.4.2 that

$$\sum_{n\geq 0}\sum_{k\geq -1}(-1)^{k-1}\operatorname{ch}\tilde{H}_k(M_n) = \prod_{i\leq j}(1-x_ix_j)\prod_{i\geq 1}(1-x_i)^{-1}.$$

The right hand side can be decomposed into Schur functions by using the following symmetric function identity of Littlewood [**123**, p.238]:

$$\prod_{i\leq j}(1-x_ix_j)\prod_{i\geq 1}(1-x_i)^{-1} = \sum_{\lambda=\lambda'}(-1)^{\frac{|\lambda|-r(\lambda)}{2}}s_\lambda,$$

where $r(\lambda)$ is the rank of λ, i.e., the size of the main diagonal (or Durfee square) of the Young diagram for λ. From this we conclude that

$$\bigoplus_{k\geq -1}(-1)^{k-1}\tilde{H}_k(M_n) \cong_{\mathfrak{S}_n} \bigoplus_{\substack{\lambda\,:\,\lambda\vdash n\\ \lambda=\lambda'}}(-1)^{\frac{|\lambda|-r(\lambda)}{2}}S^\lambda.$$

With additional work involving long exact sequences of relative homology, Bouc obtains the following beautiful refinement.

Theorem 2.4.10 (Bouc [**47**]). *For all $n \geq 1$ and $k \in \mathbb{Z}$,*

$$\tilde{H}_{k-1}(M_n) \cong_{\mathfrak{S}_n} \bigoplus_{\substack{\lambda\,:\,\lambda\vdash n\\ \lambda=\lambda'\\ r(\lambda)=|\lambda|-2k}} S^\lambda. \tag{2.4.5}$$

From Bouc's formula one obtains a formula for the Betti number in dimension $k-1$ as the number of standard Young tableaux of self-conjugate shape and rank $n-2k$ (which can be computed from the hook-length formula, Theorem 2.2.5). This result provides an excellent illustration of the use of representation theory in the computation of Betti numbers.

We sketch a proof of Theorem 2.4.10 due to Dong and Wachs [**62**], which involves a technique called discrete Hodge theory. Let Δ be a G-simplicial complex. The *combinatorial Laplacian* $\Lambda_k : C_k(\Delta) \to C_k(\Delta)$ is defined by

$$\Lambda_k = \delta_{k-1}\partial_k + \partial_{k+1}\delta_k.$$

A basic result of discrete Hodge theory is that

$$\tilde{H}_k(\Delta) \cong_G \ker\Lambda_k. \tag{2.4.6}$$

The key observation of [**62**] is that when one applies the Laplacian Λ_{k-1} to an oriented simplex $\gamma \in C_{k-1}(M_n)$, one gets

$$\Lambda_{k-1}(\gamma) = T_n \cdot \gamma,$$

where $T_n = \sum_{1\le i<j\le n}(i,j) \in \mathbb{C}\mathfrak{S}_n$ and (i,j) denotes a transposition. It is then shown that T_n acts on the Specht module S^λ as multiplication by the scalar c_λ defined by

$$c_\lambda = \sum_{i=1}^{r}\left(\binom{\alpha_i+1}{2} - \binom{\beta_i+1}{2}\right),$$

where $r = r(\lambda)$, $\alpha_i = \lambda_i - i$, and $\beta_i = \lambda'_i - i$. That is, α_i is the number of cells to the right of and in the same row as the ith cell of the diagonal of λ and β_i is the number of cells below and in the same column as the ith cell of the diagonal. The array $(\alpha_1,\dots,\alpha_r \mid \beta_1,\dots,\beta_r)$ is known as Frobenius notation for λ. Note that λ is uniquely determined by its Frobenius notation.

Next we decompose $C_{k-1}(M_n)$ into Specht modules by using (2.4.4) and another symmetric function identity of Littlewood, cf., [**126**, I 5 Ex. 9b], namely,

$$\prod_{i\le j}(1-x_ix_j) = \sum_{\nu\in\mathcal{B}}(-1)^{|\nu|/2}s_\nu,$$

where $\mathcal{B}$ is the set of all partitions of the form $(\alpha_1+1,\dots,\alpha_r+1 \mid \alpha_1,\dots,\alpha_r)$ for some r. This and Pieri's rule (Theorem 2.2.11) yield the decomposition into irreducibles:

$$C_{k-1}(M_n) \cong \bigoplus_{\lambda\in A_n} a^k_\lambda S^\lambda,$$

where

$$A_n = \{(\alpha_1,\dots,\alpha_r \mid \beta_1,\dots,\beta_r) \vdash n : r\ge 1,\ \alpha_i \ge \beta_i \quad \forall\, i\in[r]\}$$

and a^k_λ is a nonnegative integer. If λ is self-conjugate then

$$a^k_\lambda = \begin{cases} 1 & \text{if } r(\lambda) = n-2k \\ 0 & \text{otherwise} \end{cases}.$$

Clearly for $\lambda\in A_n$, we have $c_\lambda = 0$ if and only if λ is self-conjugate. It follows that

$$\ker\Lambda_{p-1} \cong \bigoplus_{\substack{\lambda:\ \lambda\vdash n\\ \lambda=\lambda'}} a^k_\lambda\, S^\lambda \cong \bigoplus_{\substack{\lambda:\ \lambda\vdash n\\ \lambda=\lambda'\\ r(\lambda)=n-2k}} S^\lambda.$$

Theorem 2.4.10 now follows by discrete Hodge theory (2.4.6).

There is a bipartite analog of the matching complex called the $m\times n$ chessboard complex. This is the simplicial complex of subgraphs of the complete bipartite graph on node sets of size m and n, whose nodes have degree at most 1. Alternatively, it is the simplicial complex of nonattacking rook placements on an $m\times n$ chessboard. Friedman and Hanlon [**75**] use discrete Hodge theory to obtain a chessboard complex analog of Bouc's formula (2.4.5). (In fact the proof of (2.4.5) described above was patterned on the Friedman-Hanlon proof.) Not only can the Betti numbers be computed from this formula, but Shareshian and Wachs [**157**] use the formula to compute torsion in the integral homology of the chessboard complex; see Theorem 5.5.5. So representation theory has proved to be a useful tool in computing torsion in integral homology as well as Betti numbers. Karaguezian, Reiner and Wachs [**107**] use formula (2.4.5) and its chessboard complex analog to derive a more general result of Reiner and Roberts [**140**], which computes the homology of general bounded degree graph and bipartite graph complexes.

The matching complex and chessboard complex have arisen in various areas of mathematics, such as group theory (see Section 5.2), commutative algebra [**140**] and discrete geometry [**223**]. These complexes are discussed further in Lecture 5. See [**202**] for a survey article on the topology of matching complexes, chessboard complexes and general bounded degree graph complexes.

LECTURE 3
Shellability and Edge Labelings

3.1. Shellable Simplicial Complexes

Shellability is a combinatorial property of simplicial and more general cell complexes, with strong topological and algebraic consequences. Shellability first appeared in the middle of the nineteenth century in Schläfli's computation of the Euler characteristic of a convex polytope [**152**]. Schläfli made the assumption, without proof, that the boundary complex of a convex polytope is shellable. This assumption was eventually proved in 1970 by Brugesser and Mani [**52**] and was used in McMullen's proof [**128**] of the famous upper bound conjecture for convex polytopes (the upper bound conjecture is stated in Section 4.1).

The original theory of shellability applied only to pure complexes. In the early 1990's, a nonpure simplicial complex arose in the complexity theory work of Björner, Lovász and Yao [**34**] with topological properties somewhat similar to those of pure shellable complexes. This led Björner and Wachs [**40, 41**] to extend the theory of shellability to nonpure complexes. The Goresky-MacPherson formula created a need for such an extension, since, unlike for hyperplane arrangements, the intersection semilattice of a subspace arrangement is not necessarily pure. As it turned out, there were many other uses for nonpure shellability.

For each face F of a simplicial complex Δ, let $\langle F\rangle$ denote the subcomplex generated by F, i.e., $\langle F\rangle = \{G : G \subseteq F\}$. A simplicial complex Δ is said to be *shellable* if its facets can be arranged in linear order $F_1, F_2, \ldots, F_t$ in such a way that the subcomplex $\left(\bigcup_{i=1}^{k-1}\langle F_i\rangle\right) \cap \langle F_k\rangle$ is pure and $(\dim F_k - 1)$-dimensional for all $k = 2, \ldots, t$. Such an ordering of facets is called a *shelling*. We emphasize that we are not assuming purity; here shellability refers to what is commonly called nonpure shellability.

A shelling of the boundary complex of the 3-simplex is given in Figure 3.1.1. A pure nonshellable complex and a nonpure shellable complex are given in Figure 3.1.2, where the shading indicates that the triangles are filled in.

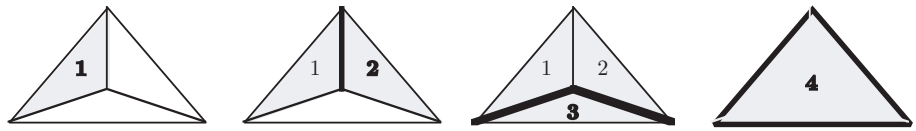


Figure 3.1.1. Shelling of the boundary of 3-simplex

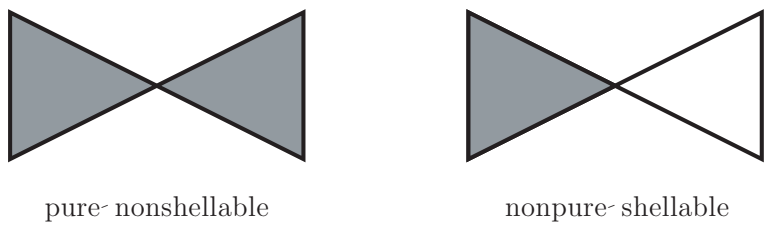
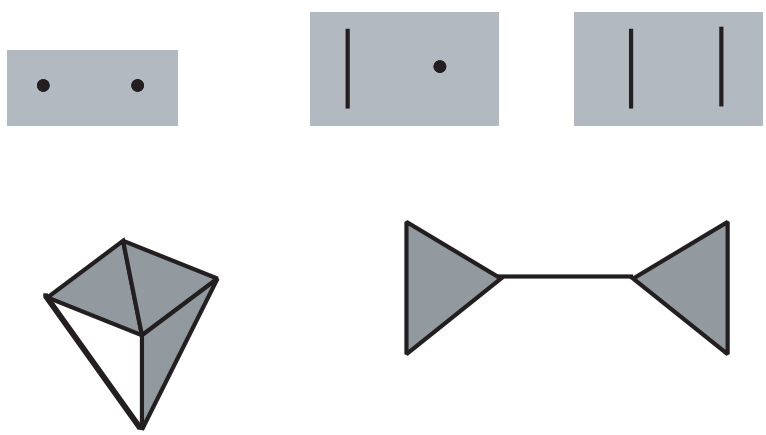


Figure 3.1.2

Exercise 3.1.1. Verify that the third and fifth simplicial complexes below are not shellable, while the others are.



Theorem 3.1.2 (Brugesser and Mani [**52**])**.** *The boundary complex of a convex polytope is shellable.*

A geometric construction called a line shelling is used to prove this result. See Ziegler's book [**220**] on polytopes for a description of this basic construction.

Shellability is not a topological property. By this we mean that shellability of a complex is not determined by the topology of its geometric realization; a shellable simplicial complex can be homeomorphic to a nonshellable simplicial complex. In fact, there exist nonshellable triangulations of the 3-ball and the 3-sphere; see [**120**],[**220**, Chapter 8], [**221**], [**125**].

Shellability does have strong topological consequences, however, as is shown by the following result. By a *wedge* $\vee_{i=1}^{n} X_i$ of n mutually disjoint connected topological

spaces X_i, we mean the space obtained by selecting a base point for each X_i and then identifying all the base points with each other.

Theorem 3.1.3 (Björner and Wachs [**40**]). *A shellable simplicial complex has the homotopy type of a wedge of spheres (in varying dimensions), where for each i, the number of i-spheres is the number of i-facets whose entire boundary is contained in the union of the earlier facets. Such facets are usually called* homology facets.

Proof idea. Let Δ be a shellable simplicial complex. We first observe that any shelling of Δ can be rearranged to produce a shelling in which the homology facets come last. So Δ has a shelling $F_1, F_2, \ldots F_k$ in which $F_1, \ldots, F_j$ are not homology facets and $F_{j+1}, \ldots, F_k$ are, where $1 \leq j \leq k$. The basic idea of the proof is that as we attach the first j facets, we construct a contractible simplicial complex at each step. The remaining homology facets attach as spheres since the entire boundary of the facet is identified with a point. □

Corollary 3.1.4. *If Δ is shellable then for all i,*

$$\tilde{H}_i(\Delta;\mathbb{Z}) \cong \tilde{H}^i(\Delta;\mathbb{Z}) \cong \mathbb{Z}^{r_i}, \tag{3.1.1}$$

where r_i is the number of homology i-facets of Δ.

The homology facets yield more than just the Betti numbers; they form a basis for cohomology. We discuss the connection with cohomology further in the context of lexicographic shellability in the next section.

There are certain operations on complexes and posets that preserve shellability. For a simplicial complex Δ and $F \in \Delta$, define the *link* to be the subcomplex given by

$$\mathrm{lk}_\Delta F := \{G \in \Delta : G \cup F \in \Delta,\, G \cap F = \emptyset\}.$$

Theorem 3.1.5 ([**40**]). *The link of every face of a shellable complex is shellable.*

Define the *suspension* of a simplicial complex Δ to be

$$\mathrm{susp}(\Delta) := \Delta * \{\{a\}, \{b\}, \emptyset\},$$

where $*$ denotes the join, and $a \neq b$ are not vertices of Δ.

Theorem 3.1.6 ([**40**]). *The join of two simplicial complexes is shellable if and only if each complex is shellable. In particular, susp(Δ) is shellable if and only if Δ is shellable.*

Define the k-skeleton of a simplicial complex Δ to be the subcomplex consisting of all faces of dimension k or less.

Theorem 3.1.7 ([**41**]). *The k-skeleton of a shellable simplicial complex is shellable for all $k \geq 0$.*

A simplicial complex Δ is said to be *r-connected* (for $r \geq 0$) if it is nonempty and connected and its jth homotopy group $\pi_j(\Delta)$ is trivial for all $j = 1, \ldots, r$. So 0-connected is the same as connected and 1-connected is the same as simply connected. A nonempty simplicial complex Δ is said to be *r-acyclic* if its jth reduced integral homology group $\tilde{H}_j(\Delta)$ is trivial for all $j = 0, 1, \ldots, r$. We say that X is (-1)-connected and (-1)-acyclic when Δ is nonempty. It is also convenient to say that every simplicial complex is r-connected and r-acyclic for all $r \leq -2$.

It is a basic fact of homotopy theory that r-connected implies r-acyclic and that the converse holds only for simply connected complexes. Another basic fact

is that a a simplicial complex is r-connected (r-acyclic) if and only if its $(r+1)$-skeleton has the homotopy type (homology) of a wedge of $(r+1)$-spheres. This makes shellability a useful tool in establishing r-connectivity and r-acyclicity.

Theorem 3.1.8. *A simplicial complex is r-connected (and therefore r-acyclic) if its $(r+1)$-skeleton is pure shellable.*

A poset P is said to be shellable if its order complex $\Delta(P)$ is shellable. The following is an immediate consequence of Theorems 3.1.5 and 3.1.6.

Corollary 3.1.9.
(a) *A bounded poset P is shellable if and only if its proper part $\bar{P}$ is shellable.*
(b) *Every (open or closed) interval of a shellable poset is shellable.*
(c) *The join of two posets is shellable if and only if each of the posets is shellable.*

Theorem 3.1.10 (Björner and Wachs [**41**])**.** *The product of bounded posets is shellable if and only if each of the posets is shellable.*

Sometimes shellability has stronger topological consequences than homotopy type.

Theorem 3.1.11 (Danaraj and Klee [**56**])**.** *Let Δ be a pure shellable d-dimensional simplicial complex in which every codimension* 1 *face is contained in at most* 2 *facets. Then Δ is homeomorphic to a d-sphere or a d-ball. Moreover, Δ is homeomorphic to a d-sphere if and only if every codimension* 1 *face is contained in exactly* 2 *facets.*

Note that the condition on codimension 1 faces in the Danaraj and Klee theorem can be expressed as follows: Closed length 2 intervals of $L(\Delta)$ have at most 4 elements. Pure posets in which every length 2 interval has exactly 4 elements are said to be *thin*. We have the following poset version of the Danaraj and Klee result.

Theorem 3.1.12 (Björner [**23**])**.** *If $\hat{P}$ is pure, thin and shellable then P is isomorphic to the face poset of a regular cell decomposition (a generalization of simplicial decomposition; see* [**92**]*) of an $l(P)$-sphere.*

Many natural classes of simplicial complexes and posets, which have arisen in various fields of mathematics, have turned out to be shellable (pure or nonpure). A striking illustration of the ubiquity of shellability is given by the following result.

Theorem 3.1.13 (Shareshian [**153**])**.** *The lattice of subgroups of a finite group G is shellable if and only if G is solvable.*

The pure version of this result, which states that *the lattice of subgroups of a finite group G is pure shellable if and only if G is supersolvable*, was proved by Björner [**21**] in the late 70's by introducing a technique called lexicographic shellability. Shareshian uses a general (nonpure) version of lexicographic shellability more recently introduced by Björner and Wachs [**40**]. We discuss lexicographic shellability in the remaining sections of this lecture. Further discussion of connections between group theory and poset topology can be found in Section 5.2.

We end this section with two interesting conjectures. A pure simplicial complex Δ is said to be *extendably shellable* if every partial shelling of Δ extends to a shelling of Δ. Tverberg (see [**56**]) conjectured that the boundary complex of every polytope

is extendably shellable. The conjecture was settled in the negative by Ziegler [**221**] who showed that there are simple and simplicial polytopes whose boundary complex is not extendably shellable. The conjecture is, however, true for the simplex since every ordering of the facets of the boundary complex is a shelling. Hence the $k = d - 1$ case of the following conjecture holds.

Conjecture 3.1.14 (Simon [**162**]). *For all $k \leq d$, the k-skeleton of the d-simplex is extendably shellable.*

Simon's conjecture was shown to be true for $k \leq 2$ by Björner and Eriksson [**31**], who extended the conjecture to all matroid complexes. Recently Hall [**84**] showed that the boundary complex of the 12-dimensional cross-polytope is a counterexample to the extended conjecture of Björner and Eriksson.

A simplicial complex Δ is said to be *minimally nonshellable* (called an obstruction to shellability in [**200**]) if Δ is not shellable, but every proper induced subcomplex of Δ is shellable. For example, the complex consisting of two disjoint 1-simplexes is a minimally nonshellable complex (see Exercise 3.1.1). Billera and Myers [**20**] showed that this is the only minimally nonshellable order complex. It is also the only 1-dimensional minimally nonshellable simplicial complex. It was shown by Wachs [**200**] that the number of vertices in any 2-dimensional minimally nonshellable simplicial complex is either $5, 6$ or 7; hence there are only finitely many such complexes.

Conjecture 3.1.15 (Wachs). *There are only finitely many d-dimensional minimally nonshellable simplicial complexes (obstructions to shellability) for each d.*

3.2. Lexicographic Shellability

In the early 1970's, Stanley [**165, 166**] introduced a technique for showing that the Möbius function of rank-selected subposets of certain posets alternates in sign. This technique involved labeling the edges of the Hasse diagram of the poset in a certain way. Stanley conjectured that the posets that he was considering were Cohen-Macaulay, a topological (and algebraic) property of simplicial complexes implied by shellability (cf., Section 4.1). Björner [**21**] proved this conjecture by finding a condition on edge labelings which implies shellability of the poset. From this emerged the theory of lexicographic shellability, which was further developed in a series of papers by Björner and Wachs, first in the pure case [**37, 38**] and later in the general (nonpure) case [**40, 41**].

There are two basic versions of lexicographic shellability, EL-shellability and CL-shellability. In this section we begin with the simpler but less powerful version, EL-shellability, and discuss some of its consequences.

An *edge labeling* of a bounded poset P is a map $\lambda : \mathcal{E}(P) \to \Lambda$, where $\mathcal{E}(P)$ is the set of edges of the Hasse diagram of P, i.e., the covering relations $x \lessdot y$ of P, and Λ is some poset (usually the integers $\mathbb{Z}$ with its natural total order relation). Given an edge labeling $\lambda : \mathcal{E}(P) \to \Lambda$, one can associate a word

$$\lambda(c) = \lambda(\hat{0}, x_1)\lambda(x_1, x_2) \cdots \lambda(x_t, \hat{1})$$

with each maximal chain $c = (\hat{0} \lessdot x_1 \lessdot \cdots \lessdot x_t \lessdot \hat{1})$. We say that c is *increasing* if the associated word $\lambda(c)$ is *strictly* increasing. That is, c is increasing if

$$\lambda(\hat{0}, x_1) < \lambda(x_1, x_2) < \cdots < \lambda(x_t, \hat{1}).$$

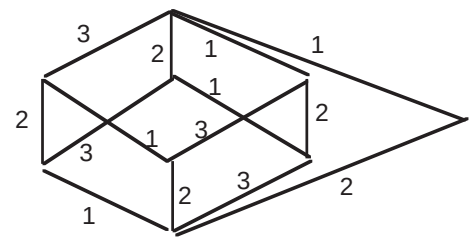


Figure 3.2.1. EL-labeling

We say that c is *decreasing* if the associated word $\lambda(c)$ is *weakly* increasing. We can order the maximal chains lexicographically by using the lexicographic order on the corresponding words. Any edge labeling λ of P restricts to an edge labeling of any closed interval $[x, y]$ of P. So we may refer to increasing and decreasing maximal chains of $[x, y]$, and lexicographic order of maximal chains of $[x, y]$.

Definition 3.2.1. Let P be a bounded poset. *An edge-lexicographical labeling* (EL-labeling, for short) of P is an edge labeling such that in each closed interval $[x, y]$ of P, there is a unique increasing maximal chain, which lexicographically precedes all other maximal chains of $[x, y]$.

An example of an EL-labeling of a poset is given in Figure 3.2.1. The leftmost chain, which has associated word 123, is the only increasing maximal chain of the interval $[\hat{0}, \hat{1}]$. It is also lexicographically less than all other maximal chains. One needs to check each interval to verify that the labeling is indeed an EL-labeling.

A bounded poset that admits an EL-labeling is said to be *edge-lexicographic shellable* (EL-shellable, for short). The following theorem justifies the name.

Theorem 3.2.2 (Björner [**21**], Björner and Wachs [**40**])**.** *Suppose P is a bounded poset with an EL-labeling. Then the lexicographic order of the maximal chains of P is a shelling of $\Delta(P)$. Moreover, the corresponding order of the maximal chains of $\bar{P}$ is a shelling of $\Delta(\bar{P})$.*

Exercise 3.2.3. Prove Theorem 3.2.2.

It is for nonbounded posets that shellability has interesting topological consequences since the order complex of a bounded poset is a just a cone.

Theorem 3.2.4 (Björner and Wachs [**40**])**.** *Suppose P is a poset for which $\hat{P}$ admits an EL-labeling. Then P has the homotopy type of a wedge of spheres, where the number of i-spheres is the number of decreasing maximal $(i+2)$-chains of $\hat{P}$. The decreasing maximal $(i+2)$-chains, with $\hat{0}$ and $\hat{1}$ removed, form a basis for cohomology $\tilde{H}^i(P; \mathbb{Z})$.*

The first part of Theorem 3.2.4 is a consequence of Theorems 3.1.3 and 3.2.2. Indeed, the proper parts of the decreasing chains are the homology facets of the shelling of $\Delta(P)$ induced by the lexicographic order of maximal chains of $\hat{P}$. To establish the second part, one needs only show that the proper parts of the decreasing chains span cohomology. This is proved by showing that the cohomology relations enable one to express a maximal chain with an ascent as the negative of a sum

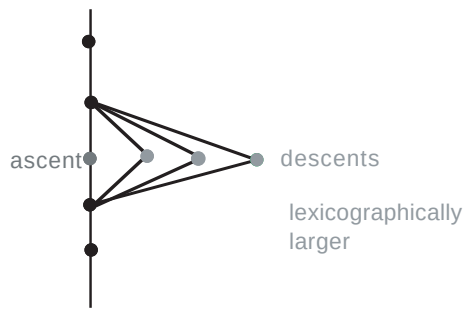


Figure 3.2.2. Straightening step

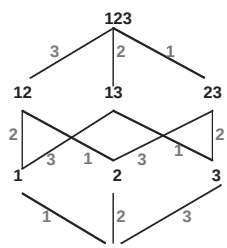


Figure 3.2.3. EL-labeling of B_3

of lexicographically larger maximal chains. This provides a step in a "straightening" algorithm for expressing maximum chains as linear combinations of decreasing maximum chains. See Figure 3.2.2.

Remark 3.2.5. In Björner's original version of EL-shellability, the unique increasing maximal chain was required to be weakly increasing and the decreasing chains were required to be strictly decreasing. The two versions have the same topological and algebraic consequences, but it is unknown whether they are equivalent.

Example 3.2.6. In the EL-labeling given in Figure 3.2.1, the two rightmost maximal chains are the only decreasing maximal chains. One has length 3 and the other has length 2. It follows from Theorem 3.2.4 that the order complex of the proper part of the poset has the homotopy type of a wedge of a 1-sphere and a 0-sphere. This is consistent with the fact that the order complex of the proper part of the poset consists of the barycentric subdivision of the boundary of a 2-simplex and an isolated point.

3.2.1. The Boolean algebra B_n.

There is a very natural EL-labeling of the Boolean algebra B_n; simply label the covering relation $A_1 \subset A_2$ with the unique element in the singleton set $A_2 - A_1$. The maximal chains correspond to the permutations in $\mathfrak{S}_n$. It is easy to see that each interval has a unique increasing chain that is lexicographically first. There is only one decreasing chain which is consistent with the fact that $\Delta(\bar{B}_n)$ is a sphere. See Figure 3.2.3.

For each $k \leq n$, define the truncated Boolean algebra B_n^k to be the subposet of B_n given by

$$B_n^k = \{A \subseteq [n] : |A| \geq k\}.$$

Define an edge labeling λ of $B_n^k \cup \{\hat{0}\}$ as follows:

$$\lambda(A_1, A_2) = \begin{cases} \max A_2 & \text{if } A_1 = \hat{0} \text{ and } |A_2| = k \\ a & \text{if } A_2 - A_1 = \{a\}. \end{cases}$$

It is easy to check that this is an EL-labeling. The decreasing chains correspond to permutations with descent set $\{k, k+1, \ldots, n-1\}$. Hence $\dim \tilde{H}^{n-k-1}(\bar{B}_n^k)$ equals the number of permutations in $\mathfrak{S}_n$ with descent set $\{k, \ldots, n-1\}$, which equals the number of standard Young tableaux of hook shape $k1^{n-k}$. An equivariant version of this result is given by the following special case of a result of Solomon (the general result is given in Section 3.4).

Theorem 3.2.7 (Solomon [**163**]). *For all $k \leq n$,*

$$\tilde{H}_{n-k-1}(\bar{B}_n^k) \cong_{\mathfrak{S}_n} S^{(k1^{n-k})}.$$

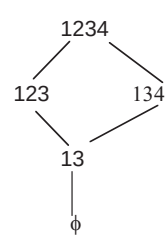
Combinatorial proof idea. We use a surjection from the set of tableaux of shape $k1^{n-k}$ to the set of maximal chains of $\bar{B}_n^k$, illustrated by the following example in which $n = 5$ and $k = 2$.

$$\begin{array}{|c|c|} \hline 3 & 1 \\ \hline 2 \\ \cline{1-1} 4 \\ \cline{1-1} 5 \\ \cline{1-1} \end{array} \mapsto (\{3,1\} \lessdot \{3,1,2\} \lessdot \{3,1,2,4\}).$$

The surjection determines a surjective $\mathfrak{S}_n$-homomorphism from the tableaux module M^λ to the chain space $C_{n-k-2}(\bar{B}_n^k; \mathbb{C})$. To show that this homomorphism induces a surjective homomorphism from the quotient space S^λ to the quotient space $\tilde{H}^{n-k-2}(\bar{B}_n^k) = C_{n-k-2}(\bar{B}_n^k; \mathbb{C})/B^{n-k-2}(\bar{B}_n^k; \mathbb{C})$, we observe that the row relations map to 0 in $C_{n-k-2}(\bar{B}_n^k; \mathbb{C})$ and show that the Garnir relations map to coboundary relations. Since the dimensions of the two vector spaces are equal, the homomorphism is an isomorphism. We demonstrate the fact that Garnir relations map to cohomology relations on two examples and leave the general proof as an exercise. The Garnir relation

$$\begin{array}{|c|c|} \hline 3 & 1 \\ \hline 2 \\ \cline{1-1} 4 \\ \cline{1-1} \end{array} + \begin{array}{|c|c|} \hline 3 & 1 \\ \hline 4 \\ \cline{1-1} 2 \\ \cline{1-1} \end{array}$$

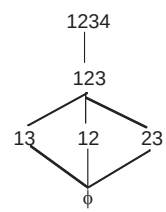
maps to the sum of maximal chains of the proper part of the poset



The Garnir relation

$$\begin{array}{|c|c|}\hline 3 & 1 \\ \hline 2 \\ \cline{1-1} 4 \\ \cline{1-1}\end{array} + \begin{array}{|c|c|}\hline 2 & 1 \\ \hline 3 \\ \cline{1-1} 4 \\ \cline{1-1}\end{array} + \begin{array}{|c|c|}\hline 2 & 3 \\ \hline 1 \\ \cline{1-1} 4 \\ \cline{1-1}\end{array}$$

maps to the sum of maximal chains of the proper part of the poset



□

3.2.2. The partition lattice Π_n.

We give two different EL-labelings of the partition lattice. The first one, due to Gessel, appears in [**21**], and the second one is due to Wachs [**198**]. If $x \lessdot y$ in Π_n then y is obtained from x by merging two blocks, say B_1 and B_2. For the first edge labeling λ_1, let

$$\lambda_1(x, y) = \max\{\min B_1, \min B_2\}$$

and for the second edge labeling λ_2, let

$$\lambda_2(x, y) = \max B_1 \cup B_2.$$

The increasing chain from $\hat{0}$ to $\hat{1}$ is the same for both labelings; it consists of partitions with only one nonsingleton block. More precisely, the chain is given by

$$\hat{0} \lessdot \{1, 2\} \lessdot \{1, 2, 3\} \lessdot \cdots \lessdot \hat{1},$$

where we have written only the nonsingleton block of each partition in the chain. We leave it as an exercise to show that these labelings are EL-labelings of the partition lattice.

The decreasing maximal chains for λ_1 and λ_2 are not the same. We describe those for λ_2 first. They also consist of partitions with only one nonsingleton block and are of the form

$$\begin{aligned} c_\sigma := (\hat{0} &\lessdot \{\sigma(n), \sigma(n-1)\} \lessdot \{\sigma(n), \sigma(n-1), \sigma(n-2)\} \lessdot \cdots \\ &\lessdot \{\sigma(n), \sigma(n-1), \sigma(n-2), \ldots, \sigma(1)\} = \hat{1}), \end{aligned}$$

where $\sigma \in \mathfrak{S}_n$ and $\sigma(n) = n$. We conclude that the homotopy type of $\bar{\Pi}_n$ is given by

$$\bar{\Pi}_n \simeq \bigvee_{(n-1)!} \mathbb{S}^{n-3} \tag{3.2.1}$$

and that the chains $\bar{c}_\sigma$, where $\sigma \in \mathfrak{S}_n$ and $\sigma(n) = n$, form a basis for $\tilde{H}^{n-3}(\bar{\Pi}_n, \mathbb{Z})$. From this basis one can immediately see that

$$\tilde{H}_{n-3}(\Pi_n) \downarrow_{\mathfrak{S}_{n-1}}^{\mathfrak{S}_n} \cong_{\mathfrak{S}_{n-1}} \tilde{H}^{n-3}(\Pi_n) \downarrow_{\mathfrak{S}_{n-1}}^{\mathfrak{S}_n} \cong_{\mathfrak{S}_{n-1}} \mathbb{C}\mathfrak{S}_{n-1}, \tag{3.2.2}$$

a result obtained by Stanley [**169**] as a consequence of his formula for the full representation of $\mathfrak{S}_n$ on $\tilde{H}_{n-3}(\bar{\Pi}_n)$ (Theorem 4.4.7).

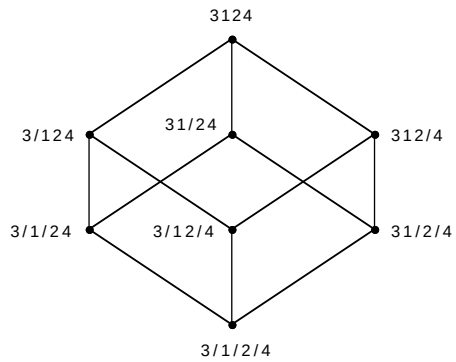


Figure 3.2.4. Π_{3124}

Next we describe a nice basis for homology of $\bar{\Pi}_n$ that is dual to the decreasing chain basis for λ_2. To *split* a permutation $\sigma \in \mathfrak{S}_n$ at positions $j_1 < j_2 < \cdots < j_k$ in $[n-1]$ is to form the partition

$$\sigma(1), \sigma(2), \ldots, \sigma(j_1) \,/\, \sigma(j_1{+}1), \sigma(j_1{+}2), \ldots, \sigma(j_2) \,/\, \ldots \,/\, \sigma(j_k{+}1), \sigma(j_k{+}2), \ldots, \sigma(n)$$

of $[n]$. For each $\sigma \in \mathfrak{S}_n$, let Π_σ be the induced subposet of the partition lattice Π_n consisting of partitions obtained by splitting the permutation σ at any set of positions in $[n-1]$. The subposet Π_{3124} of Π_4 is shown in Figure 3.2.4. Each poset Π_σ is isomorphic to the subset lattice B_{n-1}. Therefore $\Delta(\bar{\Pi}_\sigma)$ is an $(n-3)$-sphere embedded in $\Delta(\bar{\Pi}_n)$, and hence it determines a fundamental cycle $\rho_\sigma \in \tilde{H}_{n-3}(\bar{\Pi}_n; \mathbb{Z})$.

Theorem 3.2.8 (Wachs [**198**]). *The set $\{\rho_\sigma : \sigma \in \mathfrak{S}_n, \sigma(n) = n\}$ forms a basis for $\tilde{H}_{n-3}(\bar{\Pi}_n; \mathbb{Z})$ dual to the decreasing chain basis $\{\bar{c}_\sigma : \sigma \in \mathfrak{S}_n, \sigma(n) = n\}$ for cohomology. Call the homology basis, the* splitting basis.

Now we describe the decreasing chain basis for cohomology for the EL-labeling λ_1 and its dual basis for homology. Given any rooted nonplanar (i.e. children of a node are unordered) tree T on node set $[n]$, by removing any set of edges of T, one forms a partition of $[n]$ whose blocks are the node sets of the connected components of the resulting graph. Let Π_T be the induced subposet of the partition lattice Π_n consisting of partitions obtained by removing edges of T; see Figure 3.2.5. (If T is a linear tree then Π_T is the same as Π_σ, where σ is the permutation obtained by reading the nodes of the tree from the root down.) Each poset Π_T is isomorphic to the subset lattice B_{n-1}. We let ρ_T be the fundamental cycle of the spherical complex $\Delta(\bar{\Pi}_T)$. Let T be an increasing tree on node set $[n]$, i.e., a rooted nonplanar tree on node set $[n]$ in which each node i is greater than its parent $p(i)$. We form the chain c_T in Π_T, from top down, by removing the edges $\{i, p(i)\}$, one at a time, in increasing order of i. For the increasing tree T in Figure 3.2.5,

$$c_T = (1/2/3/4 \;\lessdot\; 3/24/1 \;\lessdot\; 324/1 \;\lessdot\; 1234).$$

We claim that the c_T, where T is an increasing tree on node set $[n]$, are the decreasing chains of λ_1.

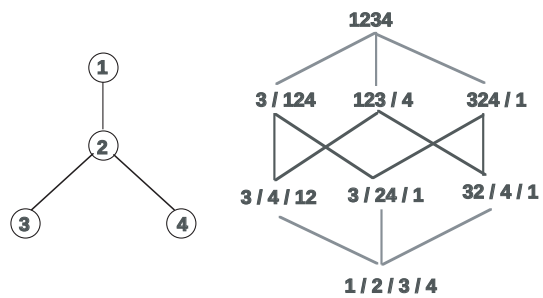


Figure 3.2.5. T and Π_T

Theorem 3.2.9. *Let $\mathcal{T}_n$ be the set of increasing trees on node set $[n]$. The set $\{\rho_T : T \in \mathcal{T}_n\}$ forms a basis for $\tilde{H}_{n-3}(\bar{\Pi}_n;\mathbb{Z})$ dual to the decreasing chain basis $\{\bar{c}_T : T \in \mathcal{T}_n\}$ for cohomology. Call the homology basis, the* tree splitting basis.

Exercise 3.2.10.

(a) Show λ_1 is an EL-labeling whose decreasing maximal chains are of the form c_T, where T is an increasing tree.
(b) Prove Theorem 3.2.9
(c) Show λ_2 is an EL-labeling whose decreasing maximal chains are of the form c_σ, where $\sigma \in \mathfrak{S}_n$ is such that $\sigma(n) = n$.
(d) Prove Theorem 3.2.8.

Both of the EL-labelings and their corresponding bases are special cases of more general constructions. The first EL-labeling λ_1 and its bases are specializations of constructions for geometric lattices due to Björner [**21, 22**]; this is discussed in the next section. The second EL-labeling λ_2 generalizes to all Dowling lattices; see the next exercise. A geometric interpretation of the splitting basis, in which the fundamental cycles correspond to bounded regions in an affine slice of the real braid arrangement, is given by Björner and Wachs [**42**]. This leads to analogs of the splitting basis for intersection lattices of other Coxeter arrangements, in particular the type B partition lattice Π_n^B. An analog of the splitting basis for all Dowling lattices is given by Gottlieb and Wachs [**83**].

Exercise 3.2.11 (Gottlieb [**81**, Section 7.3])**.**

(a) Find an EL-labeling of the type B partition lattice Π_n^B analogous to the EL-labeling λ_2 of Π_n.
(b) Describe the decreasing chains in terms of signed permutations in $\mathfrak{S}_n[\mathbb{Z}_2]$ whose right-to-left maxima are positive, where for $i = 1, 2, \ldots n$, we say that $\sigma(i)$ is a right-to-left maxima of σ if $|\sigma(i)| > |\sigma(j)|$ for all $j = i \ldots n$.
(c) Show that the number of decreasing chains is $(2n-1)!! := 1 \cdot 3 \cdots (2n-1)$, thereby recovering the well-known fact that $\bar{\Pi}_n^B$ has the homotopy type of a wedge of $(2n-1)!!$-spheres of dimension $n-2$.

A partition $\pi \in \Pi_n$ is said to be *noncrossing* if for all $a < b < c < d$, whenever a, c are in a block B of π and b, d are in a block B' of π then $B = B'$. Let NC_n be the induced subposet of Π_n consisting of noncrossing partitions. This poset, known as the *noncrossing partition lattice*, was first introduced by Kreweras [**115**] who

showed that it is a graded lattice with Möbius invariant equal to the signed Catalan number $(-1)^{n-1}\frac{1}{n}\binom{2n-2}{n-1}$. It has since undergone extensive study due to its many fascinating properties and its connections to various diverse fields of mathematics such as combinatorics, discrete geometry, mathematical biology, geometric group theory, low dimensional topology, and free probability; see the survey papers of Simion [**161**] and McCammond [**127**].

Björner and Edelman showed that NC_n is EL-shellable (see Exercise 3.2.12). Reiner [**139**] introduced an analog of the noncrossing partition lattice for the type B Coxeter group, and established many of the same properties held by NC_n, including EL-shellability. A generalization of the types A and B noncrossing partition lattices to all finite Coxeter groups was recently introduced in the geometric group theory /low dimensional topology work of Bessis [**17**] and Brady and Watt [**49**]. This generalization is discussed in Section 3.3. The Möbius function for the type D noncrossing partition lattice was computed by Athanasiadis and Reiner and the question of EL-shellability was raised [**7**]. The general Coxeter group noncrossing partition lattices were subsequently shown to be EL-shellable by Athanasiadis, Brady and Watt [**6**].

Exercise 3.2.12 (Björner and Edelman, cf. [**21**])**.**

(a) Show that the restriction to NC_n of one of the EL-labelings of Π_n described above is an EL-labeling.
(b) Describe the decreasing chains for the EL-labeling and show that their number is the Catalan number $\frac{1}{n}\binom{2n-2}{n-1}$.

Exercise 3.2.13 (Stanley [**173**])**.** Define an edge labeling for NC_n by

$$\lambda(x,y) = \max\{i \in B_1 : i < \min B_2\},$$

where B_1 and B_2 are the blocks of x that are merged to obtain y and $\min B_1 < \min B_2$.

(a) Show that λ is an EL-labeling of NC_n.
(b) Show that the decreasing chains correspond bijectively, via the labeling, to the sequences of the form $(a_1 \geq a_2 \geq \cdots \geq a_{n-1})$ where $a_{n-i} \in [i]$ for all $i \in [n-1]$. Show that the number of such sequences is a Catalan number.

This labeling is particularly interesting because the maximal chains correspond bijectively, via this labeling, to a well-studied class of sequences called parking functions; see [**173**].

3.2.3. Geometric lattices

A *geometric lattice* is a lattice L that is semimodular (which means that for all $x, y \in L$, the join $x \vee y$ covers y whenever x covers the meet $x \wedge y$) and atomic (which means every element of L is the join of atoms). Some fundamental examples are the subset lattice B_n, the subspace lattice $B_n(q)$, the partition lattice Π_n, the type B partition lattice Π_n^B, and more generally, the intersection lattice $L(\mathcal{A})$ of any central hyperplane arrangement. Geometric lattices are fundamental structures of matroid theory. It is easy to show that geometric lattices are pure.

We describe an edge labeling for geometric lattices, which comes from Stanley's early work [**166**] on rank-selected Möbius invariants and was one of the main motivating examples for Björner's original work [**21**] on EL-shellability. Fix an ordering

$a_1, a_2, \ldots, a_k$ of the atoms of the geometric lattice L. Then label each edge $x \lessdot y$ of the Hasse diagram with the smallest i for which $x \vee a_i = y$. Note that if the atoms of the subset lattice B_n are ordered $\{1\}, \{2\}, \ldots, \{n\}$, then the geometric lattice EL-labeling is precisely the labeling given in Section 3.2.1.

Exercise 3.2.14.

(a) Show that the edge labeling for geometric lattices described above is an EL-labeling.
(b) Find an ordering of the atoms of Π_n such that the induced geometric lattice EL-labeling is equivalent to the EL-labeling λ_1 of Section 3.2.2.
(c) Is λ_2 of Section 3.2.2 equivalent to a geometric lattice EL-labeling for some ordering of the atoms?
(d) Show that every semimodular lattice has an EL-labeling.

The decreasing chains of the geometric lattice EL-labeling have a very nice characterization, due to Björner [**22**], which is described in the language of matroid theory. A set A of atoms in a geometric lattice L is said to be independent if $r(\vee A) = |A|$. A set of atoms that is minimally dependent (i.e., every proper subset is independent) is called a *circuit*. An independent set of atoms that can be obtained from a circuit by removing its smallest element (with respect to the fixed ordering $a_1, a_2, \ldots, a_k$ of the atoms of L) is called a *broken circuit*. A maximal independent set of atoms is said to be an NBC base if contains no broken circuits. There is a natural bijection from the NBC bases of L to the decreasing chains of L. Indeed, given any NBC base $A = \{a_{i_1}, \ldots, a_{i_r}\}$, where $1 \leq i_1 < i_2 < \cdots < i_r \leq k$, construct the maximal chain

$$c_A := (\hat{0} < a_{i_r} < a_{i_r} \vee a_{i_{r-1}} < \cdots < a_{i_r} \vee a_{i_{r-1}} \vee \cdots \vee a_{i_1} = \hat{1}).$$

It is not difficult to check that the label sequence of c_A is $(i_r, i_{r-1}, \ldots, i_1)$, which is decreasing, and that the map $A \mapsto c_A$ is a bijection from the NBC bases of L to the decreasing chains of L. We conclude that the set $\{\bar{c}_A : A \text{ is an NBC base of } L\}$ is a basis for top cohomology of L.

Björner [**22**] also constructs a basis for homology of L indexed by the NBC bases, which is dual to the cohomology basis described above. Any independent set of atoms in a geometric lattice generates (by taking joins) a Boolean algebra embedded in the geometric lattice. Given any independent set of atoms A, let L_A be the Boolean algebra generated by A and let ρ_A be the fundamental cycle of $\bar{L}_A$.

Theorem 3.2.15 (Björner [**22**])**.** *Fix an ordering of the set of atoms of a geometric lattice L. The set*

$$\{\rho_A : A \textit{ is an NBC base of } L\}$$

is a basis for top homology of $\bar{L}$, which is dual to the decreasing chain basis

$$\{\bar{c}_A : A \textit{ is an NBC base of } L\}$$

for top cohomology.

Exercise 3.2.16. Prove Theorem 3.2.15.

Exercise 3.2.17. Show that the tree splitting basis for homology of the partition lattice is an example of a Björner NBC basis.

For further reading on the homology of geometric lattices see Björner's book chapter [**26**].

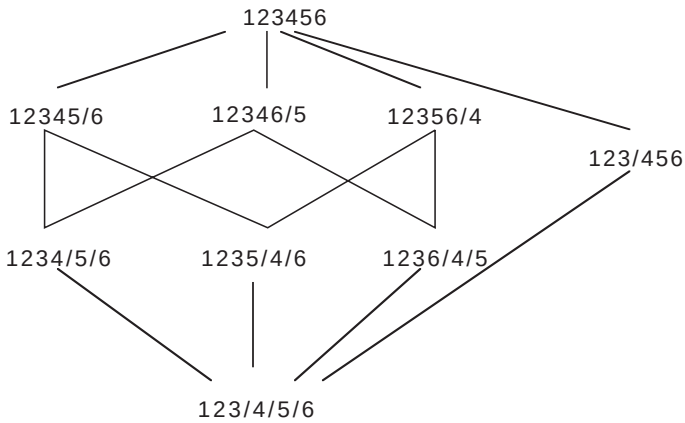


Figure 3.2.6. Interval of $\Pi_{6,3}$

3.2.4. The k-equal partition lattice

In this section we present the original example that motivated the Björner-Wachs extension of the theory of shellability from pure to nonpure complexes. This interesting example played a pivotal role in two other important developments, namely the use of poset topology in complexity theory (Björner, Lovász and Yao [**34, 33**]) and the derivation of the equivariant version of the Goresky-MacPherson formula (Sundaram and Welker [**186**]).

The k-equal partition lattice $\Pi_{n,k}$ is the subposet of Π_n consisting of partitions that have no blocks of size $\{2, 3, \ldots, k-1\}$. This lattice is not a geometric lattice when $k > 2$. It's not even pure; consider the interval of $\Pi_{6,3}$ shown in Figure 3.2.6.

The k-equal partition lattice arose in the early 1990's from the "k-equal problem" in complexity theory: Given a sequence of n real numbers $x_1, \ldots, x_n$, determine whether or not some k of them are equal. Björner, Lovász and Yao [**34, 33**] obtained the following lower bound on the complexity of the k-equal problem by studying the topology of $\bar{\Pi}_{n,k}$,

$$\max\{n-1, n\log_3 \frac{n}{3k}\}.$$

This lower bound differs by a factor of only 16 from the best upper bound. The k-equal problem translates into a subspace arrangement problem; namely determine whether or not the point $(x_1, \ldots, x_n)$ is in the complement of the real subspace arrangement $\mathcal{A}_{n,k}$, called the k-equal arrangement, consisting of subspaces of the form

$$\{(x_1, \ldots, x_n) \in \mathbb{R}^n : x_{i_1} = x_{i_2} = \cdots = x_{i_k}\},$$

where $1 \leq i_1 < \cdots < i_k \leq n$. Björner and Lovász [**33**] show that the sum of the Betti numbers of the complement of the k-equal arrangement gives a lower bound for the depth of the best decision tree for the k-equal problem. They then use the Goresky-MacPherson formula (Theorem 1.3.8) to reduce the complexity problem to that of studying the topology of lower intervals in the intersection lattice of $\mathcal{A}_{n,k}$, which is isomorphic to $\Pi_{n,k}$.

Björner and Welker [**45**] prove that all open intervals of $\Pi_{n,k}$ have the homotopy type of a wedge of spheres of varying dimensions. Since homotopy type of a wedge of spheres (of top dimension) is the main topological consequence of pure shellability, this result led Björner and Wachs to consider shellability for nonpure complexes.

We now present the nonpure EL-labeling of $\Pi_{n,k}$ obtained in [**40**] . First linearly order the label set $\{\bar{1} < \bar{2} < \cdots < \bar{n} < 1 < 2 < \cdots < n\}$. Now label the edge $\pi \lessdot \tau$ as follows:

$$\lambda(\pi,\tau) = \begin{cases} \max B & \text{if a new block } B \text{ is formed from singleton blocks} \\ a & \text{if a nonsingleton block is merged with a singleton } \{a\} \\ \overline{\max B_1 \cup B_2} & \text{if two nonsingletons } B_1 \text{ and } B_2 \text{ are merged.} \end{cases}$$

Exercise 3.2.18. Show that λ is an EL-labeling of $\Pi_{n,k}$.

We next describe the decreasing maximal chains. Given a hook shape Young diagram $k1^m$, by the *corner* of the hook we mean the cell in the first row and first column, and by the leg of T we mean the first column of T minus the corner. We refer to a skew diagram as a k-broken skew hook diagram if it is the disjoint union of hooks $H_1, H_2, \ldots, H_t$ of the form $k1^m$, where m is arbitrary. The tableaux in Figure 3.2.7 are reverse standard tableaux (i.e. column and row entries are decreasing rather than increasing) of 3-broken skew hook shape. For each reverse standard tableau T of k-broken skew hook shape, with entries in $[n]$, let π_T be the partition in $\Pi_{n,k}$ whose nonsingleton blocks are $B_1, \ldots, B_t$, where B_i is the set of entries of T in the hook H_i. For the tableaux in Figure 3.2.7,

$$\pi_T = 15, 14, 3, 9, 8, 5/10, 7, 1, 2/13, 11, 4, 12, 6$$

and

$$\pi_{T'} = 15, 14, 3, 9, 8, 5/10, 7, 1/13, 11, 4, 12, 6/2.$$

Given a reverse standard tableau T of k-broken skew hook shape, we let a be the smallest entry among all entries of T that are either in the leg of a hook or in the corner of a hook that has no leg. In the former case we let T' be the standard tableau of k-broken skew hook shape obtained by removing a from T. In the latter case we let T' be the standard tableau of k-broken skew hook shape obtained by removing the entire hook (which is a row) containing a. In either case, we refer to the tableau T' as the *predecessor* of T. In Figure 3.2.7, T' is the predecessor of T. For each reverse standard tableau T of k-broken skew hook shape, with n cells and entries in $[n]$, let c_T be the maximal chain of $\Pi_{n,k}$ such that

- $\pi_T \in c_T$
- the upper segment $c_T \cap [\pi_T, \hat{1}]$ is the chain

$$\pi_T = B_1/B_2/\ldots/B_t < B_1 \cup B_2/B_3/\ldots/B_t < \cdots < B_1 \cup \cdots \cup B_t = \hat{1}$$

- the lower segment $c_T \cap [\hat{0}, \pi_T]$ is the chain

$$\hat{0} = \pi_{T_0} < \pi_{T_1} < \cdots < \pi_{T_{n-t(k-1)}} = \pi_T,$$

  where T_i is the predecessor of T_{i+1}.

For the tableau T of Figure 3.2.7, the maximal chain c_T is given by

$$T = \begin{array}{cccccccccccc} & & & & & & & 13 & 11 & 4 \\ & & & & & & & 12 \\ & & & & & & & 6 \\ & & & & 10 & 7 & 1 \\ & & & & 2 \\ 15 & 14 & 3 \\ 9 \\ 8 \\ 5 \end{array} \qquad T' = \begin{array}{cccccccccccc} & & & & & & & 13 & 11 & 4 \\ & & & & & & & 12 \\ & & & & & & & 6 \\ & & & & 10 & 7 & 1 \\ 15 & 14 & 3 \\ 9 \\ 8 \\ 5 \end{array}$$

Figure 3.2.7.

$$\begin{aligned}
\hat{1} &> 15, 14, 3, 9, 8, 5, 10, 7, 1, 2/13, 11, 4, 12, 6 \\
&> 15, 14, 3, 9, 8, 5/10, 7, 1, 2/13, 11, 4, 12, 6 \\
&> 15, 14, 3, 9, 8, 5/10, 7, 1/13, 11, 4, 12, 6 \\
&> 15, 14, 3, 9, 8/10, 7, 1/13, 11, 4, 12, 6 \\
&> 15, 14, 3, 9, 8/10, 7, 1/13, 11, 4, 12 \\
&> 15, 14, 3, 9/10, 7, 1/13, 11, 4, 12 \\
&> 15, 14, 3/10, 7, 1/13, 11, 4, 12 \\
&> 15, 14, 3/13, 11, 4, 12 \\
&> 15, 14, 3/13, 11, 4 \\
&> 15, 14, 3 \\
&> \hat{0}
\end{aligned}$$

where only the nonsingleton blocks are shown.

Exercise 3.2.19.

(a) [**40**] Show that the set of decreasing chains with respect to λ is the set

$$\{c_T : T \text{ is a reverse standard tableau of } k\text{-broken skew hook shape, with } n \text{ cells and } n \text{ in the corner of the leftmost hook}\}.$$

(b) [**40**] Show that the Betti numbers are given by

$$\tilde{\beta}_{n-3-t(k-2)}(\bar{\Pi}_{n,k}) = \sum_{\substack{j_1 + j_2 + \ldots + j_t = n \\ j_i \geq k}} \binom{n-1}{j_1 - 1, j_2, \ldots, j_t} \prod_{i=1}^{t} \binom{j_i - 1}{k-1}$$

where $t \geq 1$, and

$$\tilde{\beta}_i(\bar{\Pi}_{n,k}) = 0$$

if i is not of the form $n - 3 - t(k-2)$ for any $t \geq 1$.

(c) [**184**] Prove that

$$\tilde{H}_{n-3-t(k-2)}(\bar{\Pi}_{n,k}) \downarrow_{\mathfrak{S}_{n-1}}^{\mathfrak{S}_n} \cong_{\mathfrak{S}_{n-1}} \bigoplus_D S^D$$

summed over all k-broken skew hook diagrams D of size n in which the corner of the leftmost hook is removed. This is equivalent to

$$\tilde{H}_{n-3-t(k-2)}(\bar{\Pi}_{n,k}) \downarrow_{\mathfrak{S}_{n-1}}^{\mathfrak{S}_n} \cong \bigoplus_{\substack{j_1+j_2+\ldots+j_t=n \\ j_i \geq k}} S^{1^{j_t-k}} \bullet S^{(k-1)} \bullet \prod_{i=1}^{t-1} S^{(k1^{j_i-k})},$$

where $\bullet$ and $\prod$ denote induction product.

The decreasing chains given in Exercise 3.2.19(a) were also used by Sundaram and Wachs [**184**] to obtain a formula for the unrestricted representation of $\mathfrak{S}_n$ on $\tilde{H}_{n-3-t(k-2)}(\bar{\Pi}_{n,k})$, involving composition product of representations. In order to transfer the representation of the symmetric group on homology of the k-equal partition poset to representations of the symmetric group on the homology of the complement $M_{\mathcal{A}_{n,k}} := \mathbb{R}^n - \cup\mathcal{A}_{n,k}$ of the real k-equal arrangement $\mathcal{A}_{n,k}$ and the complement $M_{\mathcal{A}^{\mathbb{C}}_{n,k}} := \mathbb{C}^n - \cup\mathcal{A}^{\mathbb{C}}_{n,k}$ of the complex k-equal arrangement $\mathcal{A}^{\mathbb{C}}_{n,k}$, Sundaram and Welker derived an equivariant version of the Goresky-MacPherson formula; see Theorem 5.4.2. By computing the multiplicity of the trivial representation in the homology of the complement, they obtain Betti numbers for the orbit spaces $M_{\mathcal{A}_{n,k}}/\mathfrak{S}_n$ and $M^{\mathbb{C}}_{\mathcal{A}_{n,k}}/\mathfrak{S}_n$. For instance, they recover the following result of Arnol'd [**3**]:

$$\tilde{\beta}_i(M^{\mathbb{C}}_{\mathcal{A}_{n,k}}/\mathfrak{S}_n) = \begin{cases} 1 & \text{if } i = 2k-3 \\ 0 & \text{otherwise.} \end{cases} \tag{3.2.3}$$

The orbit space $M^{\mathbb{C}}_{\mathcal{A}_{n,k}}/\mathfrak{S}_n$ is homeomorphic to the space of monic polynomials of degree n whose roots have multiplicity at most $k-1$.

Type B and D analogs of the k-equal arrangement and the k-equal partition lattice were studied by Björner and Sagan [**36**]. Gottlieb [**82**] extended this study to Dowling lattices.

3.3. CL-shellability and Coxeter Groups

We now consider a more powerful version of lexicographic shellability called chain-lexicographic shellability (CL-shellability for short). This tool was introduced by Björner and Wachs in the early 1980's in order to establish shellability of Bruhat order on a Coxeter group [**37**].

For a bounded poset P, let $\mathcal{ME}(P)$ be the set of pairs $(c, x \lessdot y)$ consisting of a maximal chain c and an edge $x \lessdot y$ along that chain. A *chain-edge labeling* of P is a map $\lambda : \mathcal{ME}(P) \to \Lambda$, where Λ is some poset, satisfying: If two maximal chains coincide along their bottom d edges, then their labels also coincide along these edges.

An example of a chain-edge labeling is given in Figure 3.3.1. Note that one of the edges has two labels. This edge receives label 3 if it is paired with the leftmost maximal chain and receives label 1 if it is paired with the other maximal chain that contains the edge.

Just as for edge labelings, we need to restrict chain-edge labelings $\lambda : \mathcal{ME}(P) \to \Lambda$ to intervals $[x, y]$. The complication is that since there can be more than one way to label an edge depending upon which maximal chain it is paired with, there can be more than one way to restrict λ to $\mathcal{ME}([x, y])$. It follows from the definition of

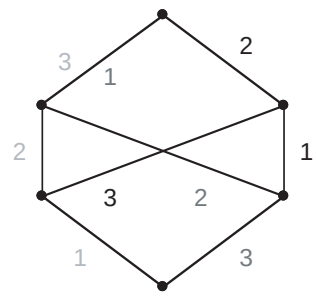


Figure 3.3.1. Chain edge labeling

chain-edge labeling that each maximal chain r of $[\hat{0}, x]$ determines a unique restriction of λ to $\mathcal{ME}([x, y])$. This enables one to talk about increasing and decreasing maximal chains and lexicographic order of maximal chains in the rooted interval $[x, y]_r$.

Definition 3.3.1. Let P be a bounded poset. *A chain-lexicographic labeling* (CL-labeling, for short) of P is a chain-edge labeling such that in each closed rooted interval $[x, y]_r$ of P, there is a unique strictly increasing maximal chain, which lexicographically precedes all other maximal chains of $[x, y]_r$. A poset that admits a CL-labeling is said to be CL-shellable.

The chain-edge labeling of Figure 3.3.1 is a CL-labeling. The unique increasing maximal chain of the poset is the leftmost maximal chain.

It is easy to see that EL-shellability implies CL-shellability. All the consequences of EL-shellability discussed in Section 3.2 are also consequences of the more general CL-shellability. It is unknown whether CL-shellability and EL-shellability are equivalent notions.

We now present the original example that motivated this more technical version of lexicographic shellability.

Definition 3.3.2. A Coxeter system (W, S) consists of a a group W together with a set of generators S such that the following relations form a presentation of W:

- $s^2 = e$, for all $s \in S$
- $(st)^{m_{s,t}} = e$, where $m_{s,t} \geq 2$, for certain $s \neq t \in S$.

The group W is said to be a Coxeter group.

Finite Coxeter groups can be characterized as finite reflection groups, i.e, finite groups generated by linear reflections in Euclidean space. Coxeter groups are an important class of groups, which have fascinating connections to many areas of mathematics, including combinatorics; see the chapter by Fomin and Reading in this volume [**71**] and the recent book of Björner and Brenti [**30**]. The finite irreducible Coxeter groups have been completely classified. There are four infinite families and six exceptional irreducible Coxeter groups. The most basic family consists of the type A Coxeter groups, which are the symmetric groups $\mathfrak{S}_n$ with the adjacent transpositions $(i, i+1)$, $i = 1, \ldots, n-1$, forming the generating set. This is the reflection group of the (type A) braid arrangement. It is also the group of symmetries of the n-simplex. The hyperoctahedral groups $\mathfrak{S}_n[\mathbb{Z}_2]$ with generators given by signed adjacent transpositions $(1, -1)$ and $(i, i+1)$, $i = 1, \ldots, n-1$, form

the type B family. This is the reflection group of the type B braid arrangement. It is also the symmetry group of the n-cube and the n-cross-polytope.

Let (W, S) be a Coxeter system. Every $\sigma \in W$ can be expressed as a product,

$$\sigma = s_1 \dots s_k,$$

where $s_i \in S$. The word $s_1 \dots s_k$ is said to be a *reduced expression* for σ if its length k is minimum among all words whose product is σ. The *length* $l(\sigma)$ of σ is defined to be the length of a reduced expression for σ.

Bruhat order on W is a partial order relation on W that is defined via the covering relation, $\sigma < \tau$ if

- $\tau = t\sigma$, for some $t \in T := \{\alpha s \alpha^{-1} : s \in S, \alpha \in W\}$
- $l(\tau) = l(\sigma) + 1$.

For any subset J of the generating set S, there is an induced subposet of Bruhat order on W, called the *quotient* by J, defined as follows

$$W^J := \{\sigma \in W : s\sigma > \sigma \text{ for all } s \in J\}.$$

Note $W = W^{\emptyset}$.

Bruhat order describes the inclusion relationships of the Schubert subvarieties of a flag manifold. It was conjectured by de Concini and Stanley in the late 1970's that any open interval of of Bruhat order on a quotient of a Coxeter group is homeomorphic to a d-sphere or a d-ball, where d is the length of the interval. This result was needed by de Concini and Lakshmibai [**57**] in their work on Cohen-Macaulayness of homogeneous coordinate rings of certain generalized Schubert varieties. In an attempt to prove the conjecture by establishing EL-shellability, Björner and Wachs instead came up with the notion of CL-shellability and constructed a CL-labeling of the dual Bruhat poset (call this a dual CL-labeling). This labeling relied on the following well-known characterization of Bruhat order.

Theorem 3.3.3 (Subword characterization of Bruhat order)**.** *Let (W, S) be a Coxeter system. Then $\sigma < \tau$ in Bruhat order on W if and only if for any reduced expression w for τ there is a reduced expression for σ that is a subword of w.*

We describe the dual CL-labeling of intervals $[\sigma, \tau]$ of W^J (this works for infinite W as well as finite W). Fix a reduced expression w for τ. It follows easily from the subword characterization that as we travel down a maximal chain, we delete (unique) letters of w one at a time until we reach a reduced word for σ. Label the edges of each maximal chain from top down with the position of the letter in w that is crossed out. This is illustrated in Figure 3.3.2 on the full interval $[e, 321]$ of Bruhat order on $\mathfrak{S}_3$, where s_i denotes the adjacent transposition $(i, i+1)$.

Theorem 3.3.4 (Björner and Wachs [**37**])**.** *The dual chain-edge labeling described above is a dual CL-labeling of closed intervals of Bruhat order on W^J. The number of decreasing chains (from top to bottom) is 1 if $J = \emptyset$ and is at most 1 otherwise.*

It follows from Theorem 3.3.4 that every open interval (σ, τ) of W^J has the homotopy type of a $(l(\tau) - l(\sigma) - 2)$-sphere or is contractible. A stronger topological consequence can be obtained by using the Danaraj and Klee result (Theorem 3.1.11). Indeed, it can be shown that every closed interval of length 2 in W has exactly 4 elements and that every closed interval of length 2 in W^J has at most 4 elements.

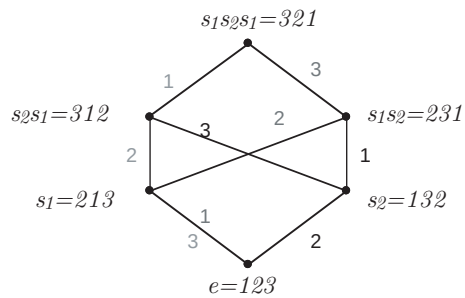


Figure 3.3.2. Dual CL-labeling of $\mathfrak{S}_3$

Corollary 3.3.5. *Every open interval (σ, τ) of Bruhat order on a quotient W^J of a Coxeter group W is homeomorphic to a $(l(\tau)-l(\sigma)-2)$-sphere or a $(l(\tau)-l(\sigma)-2)$-ball. If $J = \emptyset$ then (σ, τ) is homeomorphic to a $(l(\tau) - l(\sigma) - 2)$-sphere.*

Besides for quotients, there are other interesting classes of induced subposets of Bruhat order such as descent classes, studied by Björner and Wachs [**39**], and the subposet of involutions, which arose in algebraic geometry work of Richardson and Springer [**145**]. There is a natural notion of descent set for general Coxeter groups which generalizes the descent set of a permutation. A descent class of a Coxeter system (W, S) is the set of all elements whose descent set is some fixed subset of S. Björner and Wachs used the chain-edge labeling described above to show that every finite interval of any descent class of any Coxeter group is dual CL-shellable. This chain-edge labeling does not work for the subposet of involutions because the maximal chains of the subposet are not maximal in the full poset. Recently Incitti found an EL-labeling that does work for the classical Weyl groups (i.e., the type A, B or D Coxeter groups) and conjectured that his result extends to all intervals of all (finite or infinite) Coxeter groups. More recently, Hultman [**99**] proved the main consequence of this conjecture, namely that all intervals of Bruhat order on the set of involutions of a Coxeter group are homeomorphic to spheres.

An alternative way to label the edges of Hasse diagram of W^J is by labeling the edge $\sigma \lessdot \tau$ with the element $t \in T$ such that $\tau = t\sigma$. By imposing a certain linear order on T, Edelman [**67**] showed that this edge labeling is an EL-labeling for the symmetric group. Proctor [**134**] did the same for the classical Weyl groups. Several years after the introduction of CL-shellability, Dyer [**65**] found a way to linearly order T so that the edge labeling by elements of T is an EL-labeling for all Coxeter groups and all quotients. Dyer's linear order (called reflection order) was recently used by Williams [**212**] to obtain an EL-labeling of the poset of cells in a certain cell decomposition of Rietsch [**146**] of the totally nonnegative part of an arbitrary flag variety (for any reductive linear algebraic group). This and the fact that the poset is thin (cf., Theorem 3.1.12) led Williams to conjecture that Rietsch's cell decomposition is a regular cell decomposition of a ball (if true this would improve a result of Lusztig [**124**] asserting that the totally nonnegative part of the flag variety is contractible).

There are two other natural partial order relations on a Coxeter group (W, S) that are important and interesting. By replacing T by S in the definition of Bruhat order, one gets the definition of *weak order*. By replacing ordinary length l by absolute length al in the definition of Bruhat order, where $al(\sigma)$ is the length of the shortest factorization of σ in elements of T, one gets the definition of *absolute length order*.

The Hasse diagram of weak order is the same as the Cayley graph of the group W with respect to S, directed away from the identity. It is pure and bounded with the same rank function as Bruhat order, ordinary length l. Its topology was first studied by Björner [**24**] who showed that although it is not lexicographically shellable, it has the homotopy type of a sphere.

The Hasse diagram of absolute length order W^{al} is the same as the Cayley graph of the group W with respect to T, directed away from the identity. It is pure with rank function, absolute length al; its bottom element is e, but it has no unique top element. In the symmetric group $\mathfrak{S}_n$, the maximal elements are the n-cycles. Interest in the absolute length order is a fairly recent development, which arose in work on the braid group [**48**], [**49**], [**17**]. It was shown by Brady [**48**] that if W is the symmetric group $\mathfrak{S}_n$, then every interval $[e, c]$ of W^{al}, where c in an n-cycle, is isomorphic to the noncrossing partition lattice NC_n discussed in Section 3.2.2. This observation (and connections to finite type Artin groups) led Bessis [**17**] and Brady and Watt [**49**] to define the noncrossing partition poset for any finite Coxeter group to be the interval $[e, c]$ of W^{al}, where c is a Coxeter element of (W, S). A Coxeter element is a product of all the elements of S in some order. In $\mathfrak{S}_n$ the Coxeter elements are the n-cycles. Reiner's type B noncrossing partition lattice as well as the classical (type A) noncrossing partition lattice are recovered from this definition.

Exercise 3.3.6 (Brady [**48**])**.** Show that the interval $[e, c]$, where c is an n-cycle, in absolute length order of the symmetric group is isomorphic to the noncrossing partition lattice NC_n.

Although the topology of the interval (e, c), where c is a Coxeter element, is known to have the homotopy type of a wedge of spheres via the Athanasiadis-Brady-Watt proof of EL-shellability [**6**], little is known about the topology of the full absolute length poset, W^{al}, even for the symmetric group.

Problem 3.3.7 (Reiner [**2**, Problem 3.1])**.** What can be said about the topology of $W^{al} - \{e\}$, where W is an arbitrary finite Coxeter group? Is W^{al} lexicographically shellable, for types A and B? It is known that for type D, the poset W^{al} is not shellable.

In the next exercise we see other examples of posets that admit CL-labelings.

Exercise 3.3.8. Let $\mathcal{W}_n$ be the poset of finite words over alphabet $[n]$, ordered by the subword relation. We have $34 < 23244$ in $\mathcal{W}_n$. Let $\mathcal{N}_n$ be the induced subposet of normal words, where a word is said to be *normal* if no two adjacent letters are equal. For example, 2324 is normal while 23244 is not.

(a) (Björner and Wachs [**38**]) Find a dual CL-labeling of each interval $[u, v]$ in $\mathcal{N}_n$, and show that (u, v) is homeomorphic to an $(l(v) - l(u) - 2)$-sphere.
(b) (Björner [**25**]) Find a dual CL-labeling of each interval $[u, v]$ in $\mathcal{W}_n$.

(c) (Björner [**25, 27**]) Given a word $v = a_1 \dots a_n$ in $\mathcal{W}_n$, define its repetition set $R(v) = \{i : a_i = a_{i-1}\}$. A *normal embedding* of a word u in $v = a_1 \dots a_n$ is a sequence $1 \le i_i < \dots < i_k \le n$ such that $u = a_{i_1} \dots a_{i_k}$ and $R(v) \subseteq \{i_1, \dots, i_k\}$. Show that the interval (u, v) in $\mathcal{W}_n$ is homotopy equivalent to a wedge of $n_{u,v}$ spheres of dimension $l(v) - l(u) - 2$, where $n_{u,v}$ is the number of normal embeddings of u in v.

In the next lecture, a formulation of CL-shellability, called recursive atom ordering, which does not involve edge labelings, will be presented. This formulation has proved to be quite useful in many applications, indeed, even more useful than the original chain labeling version. A more general form of lexicographic shellability which does involve edge labelings was introduced by Kozlov [**112**]. A generalization of lexicographic shellability to balanced complexes was formulated by Hersh [**94**] and was further studied by Hultman [**98**]. A relative version of lexicographic shellability was introduced by Stanley [**172**]. A Morse theory version of lexicographical shellability, which is even more general than Kozlov's version was formulated by Babson and Hersh [**9**].

3.4. Rank Selection

Edge labelings of posets were first introduced by Stanley [**164, 165, 166**] for the purpose of studying the Möbius function of rank-selected subposets of certain pure lattices. If one drops the requirement that the unique increasing maximal chain be lexicographically first in the definition of EL-shellability then one has the kind of labeling that Stanley considered. Such a labeling was called an R-labeling (we now call it an ER-labling and we call the chain-edge version a CR-labeling). Here we discuss how CL-labelings are used to determine homotopy type and homology of rank-selected subposets of pure CL-shellable posets.

Let P be a bounded poset. For $x \in P$, define the *rank*,

$$r(x) := l([\hat{0}, x]).$$

For $R \subseteq [l(P) - 1]$, define the rank-selected subposet

$$P_R := \{x \in P : r(x) \in R\}.$$

Given a CL-labeling of a bounded poset P, the *descent set* of a maximal chain $c := (\hat{0} = x_0 \lessdot x_1 \lessdot x_2 \lessdot \dots \lessdot x_t \lessdot x_{t+1} = \hat{1})$ of P is defined to be

$$\mathrm{des}(c) := \{i \in [t] : \lambda(c, x_{i-1} \lessdot x_i) \ge \lambda(c, x_i \lessdot x_{i+1})\}.$$

Theorem 3.4.1 (Björner and Wachs [**38**])**.** *Suppose P is pure and CL-shellable. Let $R \subseteq [l(P) - 1]$. Then $\hat{P}_R$ is also CL-shellable, and hence P_R has the homotopy type of a wedge of $(|R|-1)$-spheres. The number of spheres is the number of maximal chains of P with descent set R. Moreover, the set*

$$\{c_R : c \in \mathcal{M}(P) \ \&\ \mathrm{des}(c) = R\}$$

forms a basis for $\tilde{H}^{|R|-1}(P_R; \mathbb{Z})$.

Remark 3.4.2. Theorem 3.4.1 holds for nonpure posets only when the rank set R is rather special, see [**41**]. It was shown by Björner [**21**] that pure shellability is preserved by rank selection. It is not known, however, whether pure EL-shellability is preserved by rank-selection. Stanley [**165, 166**] proved the precursor to Theorem 3.4.1 that for pure posets P that admit ER-labelings, the rank selected Möbius

invariant $\mu(\hat{P}_R)$ is equal to $(-1)^{|R|-1}$ times the number of maximal chains of P with descent set R.

Example 3.4.3. *The rank-selected Boolean algebra* $(B_n)_R$. Consider the EL-labeling of B_n given in Example 3.2.1. Since the label sequences of the maximal chains are the permutations in $\mathfrak{S}_n$, the number of maximal chains with descent set $R \subseteq [n-1]$ is the number of permutations with descent set R. So $(B_n)_R$ has the homotopy type of a wedge of $d_{n,R}$ spheres of dimension $|R|-1$, where $d_{n,R}$ is the number of permutations in $\mathfrak{S}_n$ with descent set R. (Note that the truncated Boolean algebra given in Section 3.2.1 is an example of a rank-selected Boolean algebra.) We have the following equivariant homology version of the homotopy result.

Theorem 3.4.4 (Solomon [**163**]). *Let $R \subseteq [n-1]$ and let H be the skew hook with n cells and descent set R (the definition of descent set of a skew hook was given in Section 2.2). Then*

$$\tilde{H}_{|R|-1}((B_n)_R) \cong_{\mathfrak{S}_n} S^H.$$

Exercise 3.4.5. Prove Theorem 3.4.4 by using a map from the set of tableaux of shape H to the set of maximal chains of $(B_n)_R$ thereby generalizing the proof of Theorem 3.2.7.

We consider a type B-analog and a q-analog of Example 3.4.3 in the next two exercises.

Exercise 3.4.6. *The rank-selected lattice of faces of the n-cross-polytope* $(C_n)_R$. Recall form Example 2.3.3 that we can identify the $(k-1)$-faces of C_n with k-subsets T of $[n] \cup \{-i : i \in [n]\}$ such that $\{i, -i\} \nsubseteq T$ for all i. Note that the maximal chains of C_n correspond bijectively to elements of the hyperoctahedral group $\mathfrak{S}_n[\mathbb{Z}_2]$, i.e. the signed permutations. As was mentioned earlier, there is a natural notion of descent set for general Coxeter groups which generalizes the descent set of a permutation. For the hyperoctahedral group, descent can be characterized as follows: $i \in \{0, 1, \ldots, n-1\}$ is a *descent* of a signed permutation $\sigma \in \mathfrak{S}_n[\mathbb{Z}_n]$ if either

(a) $\sigma(i) > \sigma(i+1)$ (viewed as elements of $\mathbb{Z}$) or
(b) $i = 0$ and $\sigma(1) < 0$.

Find an EL-labeling of C_n such that for all $R \subseteq [n]$, the number of maximal chains with descent set R equals $\bar{d}_{n,R}$, the number of signed permutations with descent set R. Consequently, $(C_n)_R$ has the homotopy type of a wedge of $\bar{d}_{n,R}$ spheres of dimension $|R|-1$.

Exercise 3.4.7 (Simion [**160**]). *The rank-selected subspace lattice* $(B_n(q))_R$. Find an EL-labeling λ of $B_n(q)$ with label set $[n]$ such that

- for each maximal chain c, the label sequence $\lambda(c)$ is a permutation in $\mathfrak{S}_n$,
- for each permutation $\sigma \in \mathfrak{S}_n$, the number of maximal chains c such that $\lambda(c) = \sigma$ is $q^{\mathrm{inv}(\sigma)}$, where $\mathrm{inv}(\sigma)$ is the number of inversions of σ.

Consequently, for all $R \subseteq [n-1]$, the rank-selected subspace lattice $(B_n(q))_R$ has the homotopy type of a wedge of $d_{n,R}(q)$ spheres of dimension $|R|-1$, where

$$d_{n,R}(q) = \sum_{\substack{\sigma \in \mathfrak{S}_n \\ \mathrm{des}(\sigma) = R}} q^{\mathrm{inv}(\sigma)}.$$

There are also interesting results on $\mathfrak{S}_n$-equivariant rank-selected homology of the partition lattice with ties to permutation enumeration due to Sundaram [**177, 179**]. See also the related recent work of Hanlon and Hersh [**88**] on rank-selected partition lattices [**88**]

Exercise 3.4.8. Let L be a geometric lattice with the EL-labeling given in Section 3.2.3. Also let $R \subseteq [l(L)-1]$. Show that there is a basis for top homology of L_R consisting of fundamental cycles of subposets of L_R that are all isomorphic to the join $\bar{B}_{s_1} * \bar{B}_{s_2} * \cdots * \bar{B}_{s_k}$, where $\{s_1, s_1+s_2, \ldots, s_1+s_2+\cdots+s_k\} = [l(L)] - R$.

Stanley [**169**] gives a type B analog of Theorem 3.4.4 for the lattice of Exercise 3.4.6 and a q-analog of Theorem 3.4.4 for the lattice of Exercise 3.4.7. These results involve the representation theories of the hyperoctahedral group and the general linear group, which are analogous to that of the symmetric group. He proves Theorem 3.4.4 and the analogous results by invoking the following general theorem.

Theorem 3.4.9 ([**169**])**.** *Let P be a bounded pure shellable G-poset of length l. If $R \subseteq [l-1]$ then*

$$\tilde{H}_{|R|-1}(P_R) \cong_G \bigoplus_{T \subseteq R} (-1)^{|R-T|} C_{|T|-1}(P_T). \tag{3.4.1}$$

Equivalently,

$$C_{|R|-1}(P_R) \cong_G \bigoplus_{T \subseteq R} \tilde{H}_{|T|-1}(P_T). \tag{3.4.2}$$

In [**169**] Theorem 3.4.9 is proved for a more general class of posets called Cohen-Macaulay posets. These posets are discussed in the next section. Equation (3.4.1) is a consequence of the Hopf trace formula and the fact that pure shellability, or more generally Cohen-Macaulayness, is preserved by rank-selection. Equation (3.4.2) follows by the principle of inclusion-exclusion or Möbius inversion on the subset lattice.

Note that Theorems 3.4.4 and 3.4.9 imply the fact that the regular representation of $\mathfrak{S}_n$ decomposes into a direct sum of Foulkes modules, cf., Exercise 2.2.8.

Exercise 3.4.10. Prove Theorem 3.4.4 by using Theorem 3.4.9.

As was mentioned in the introduction to this lecture, shellability theory is intimately connected with the enumeration of faces of polyhedral complexes, a central topic in geometric combinatorics; see the books of Stanley [**170**] and Ziegler [**220**]. The descent set of a maximal chain is a specialization of a more general concept in shellability theory known as the restriction of a facet, which is just the smallest new face that is added to the complex when the facet is added. When facets of pure shellable simplicial complexes are enumerated according to the size of their restriction, an important combinatorial invariant of the simplicial complex, known as the h-vector, is computed. For any $(d-1)$-dimensional simplicial complex Δ (shellable or not), the h-vector $(h_0(\Delta), h_1(\Delta), \ldots, h_d(\Delta))$ and the f-vector $(f_{-1}(\Delta), f_0(\Delta), \ldots, f_{d-1}(\Delta))$ determine each other,

$$\sum_{i=0}^{d} h_i x^{d-i} = \sum_{i=0}^{d} f_{i-1}(x-1)^{d-i},$$

but the h-vector is usually more convenient for expressing relations such as the upper bound conjecture, a symmetry relation known as the Dehn-Sommerville equations, the celebrated Billera-Lee [**19**] and Stanley [**168**] characterization of the f-vector of a simplicial polytope, and the conjectured extension to convex polytopes and homology spheres. Although shellability has played an important role in the study of f-vectors and h-vectors, much fancier tools from commutative algebra and algebraic geometry have come into play. See the books of Stanley [**170**] and Ziegler [**220**] for basic treatments of this material, the survey article of Stanley [**175**] for important recent developments, and the paper of Swartz [**187**] for even more recent developments.

Refinements of the f-vector and h-vector called the flag f-vector and the flag h-vector, respectively, (defined for all pure posets and the more general balanced complexes) have been extensively studied, beginning with the Bayer-Billera [**15**] analog of the Dehn-Sommerville equations; see [**170**] for further information. For pure lexicographically shellable posets the entries of the flag h-vector have a simple combinatorial interpretation as the number of maximal chains with fixed descent set.

For nonpure complexes, two-parameter generalizations of the f-vector and h-vector are defined in [**40**]. The f-triangle and h-triangle also determine each other. The f-triangle counts faces of a simplicial complex according to the maximum size of a facet containing the face and the size of the face. For shellable complexes, it is shown in [**40**] that the h-triangle counts facets according to the size of the restriction set and the size of the facet. Duval [**64**] shows that the entries of the h-triangle are nonnegative for a more general class of complexes than the shellable complexes; namely the sequentially Cohen-Macaulay complexes, which are discussed in the next lecture. It is pointed out by Herzog, Reiner, and Welker [**96**] that the h-triangle of a sequentially Cohen-Macaulay complex Δ encodes the multigraded Betti numbers appearing in the minimal free resolution of the Stanley-Reisner ideal of the Alexander dual of Δ. Stanley-Reisner rings and ideals are discussed briefly in the next lecture and Alexander duality is discussed in the last lecture.

Descent sets and the more general restriction sets also play an important role in direct sum decompositions of Stanley-Reisner rings of shellable complexes; due to Kind and Kleinschmidt [**109**], Garsia [**77**], and Björner and Wachs [**41**, Section 12].

LECTURE 4
Recursive Techniques

The recursive definition of the Möbius function of a poset provides a recursive technique for computing the reduced Euler characteristic of the order complex of a poset. More refined recursive techniques for computing the homology of a poset are discussed in this lecture. A general class of posets to which these techniques can be applied, the Cohen-Macaulay posets or more generally the sequentially Cohen-Macaulay posets, are discussed in Section 4.1. A recursive formulation of CL-shellability is presented in Section 4.2. In Section 4.3 the recursive techniques for computing Betti numbers are demonstrated on various examples, and in Section 4.4 equivariant versions of these techniques are also demonstrated. In Section 4.5 bases and generating sets for homology and cohomology of some of these examples are presented together with their use in computing $\mathfrak{S}_n$-equivariant homology.

4.1. Cohen-Macaulay Complexes

Recall that the link of a face F of a simplicial complex Δ is the subcomplex

$$\mathrm{lk}_\Delta F := \{G \in \Delta : G \cup F \in \Delta,\ G \cap F = \emptyset\}.$$

Definition 4.1.1. A simplicial complex Δ is said to be *Cohen-Macaulay* over $\mathbf{k}$ if

$$\tilde{H}_i(\mathrm{lk}_\Delta F; \mathbf{k}) = 0,$$

for all $F \in \Delta$ and $i < \dim \mathrm{lk}_\Delta F$.

The following result shows that Cohen-Macaulayness is a topological property.

Theorem 4.1.2 (Munkres [**130**])**.** *The simplicial complex Δ is Cohen-Macaulay over $\mathbf{k}$ if and only if for all $p \in \|\Delta\|$ and all $i < \dim \Delta$,*

$$\tilde{H}_i(\|\Delta\|; \mathbf{k}) = H_i(\|\Delta\|, \|\Delta\| - p; \mathbf{k}) = 0,$$

where the homology is reduced singular homology and relative singular homology, respectively.

The Cohen-Macaulay property has its origins in commutative algebra, in the theory of Cohen-Macaulay rings. Associated with every simplicial complex Δ on vertex set $[n]$ is a ring $\mathbf{k}[\Delta]$ called the *Stanley-Reisner ring* of the simplicial complex, which is defined to be the quotient of the polynomial ring $\mathbf{k}[x_1, \ldots, x_n]$ by the *Stanley-Reisner ideal* I_Δ generated by monomials of the form $x_{i_1} x_{i_2} \ldots x_{i_j}$ where

$\{i_1, i_2, \ldots, i_j\} \notin \Delta$. The Stanley-Reisner construction is a two-way bridge used to obtain topological and enumerative properties of the simplicial complex from properties of the ring and vice versa. A simplicial complex is Cohen-Macaulay if and only if its Stanley-Reisner ring is a Cohen-Macaulay ring (see [**170**] for the definition of Cohen-Macaulay ring). The equivalence of the characterization given in Definition 4.1.1 and the ring theoretic characterization is due to Reisner [**144**].

It follows from Theorem 4.1.2 that any triangulation of a d-sphere is Cohen-Macaulay. The ring theoretic consequence of this fact played an essential role in Stanley's celebrated proof of the Upper Bound Conjecture for spheres. For $n > d$, define the cyclic polytope $C(n, d)$ to be the convex hull of any n distinct points on the moment curve $\{(t, t^2, \ldots, t^d) \in \mathbb{R}^d : t \in \mathbb{R}\}$ (the face poset of the polytope is independent of the choice of points). The boundary complex of the cyclic polytope is a simplicial complex. The upper bound conjecture for spheres asserts that the boundary complex of the cyclic polytope achieves the maximum number of faces of each dimension, over all simplicial complexes on n vertices that triangulate a d-sphere. See [**170**] for Stanley's proof of this conjecture and other very important uses of the Stanley-Reisner ring and commutative algebra in the enumeration of faces of simplicial complexes.

The Stanley-Reisner bridge can also be crossed in the opposite direction obtaining ring theoretic information from the topology and combinatorics of the simplicial complex, as exemplified by the commutative algebra results mentioned at the end of the last lecture. Another example is a fundamental result of Eagon and Reiner [**66**] which states that a square free monomial ideal has a linear resolution if and only if it is the Stanley-Reisner ideal I_Δ of a simplicial complex Δ whose Alexander dual is Cohen-Macaulay. We will not define linear resolution, but Alexander duality is discussed in Section 5.1. A nonpure generalization of this result involving sequential Cohen-Macaulayness (discussed below) appears in papers of Herzog and Hibi [**95**] and Herzog, Reiner and Welker [**96**]. For further reading on the extensive connections between simplicial topology and commutative algebra, see Stanley's classic book [**170**] and the recent book of Miller and Sturmfels [**129**].

Exercise 4.1.3 (Walker [**206**])**.** Show that if $\mathrm{lk}_\Delta F$ is empty, 0-dimensional, or connected for all $F \in \Delta$, then Δ is pure. Consequently, Cohen-Macaulay simplicial complexes are pure.

The main tool for establishing Cohen-Macaulayness is shellability. Indeed, it follows from Theorem 3.1.5 and Corollary 3.1.4 that pure shellability implies the Cohen-Macaulay property over any $\mathbf{k}$. From Exercise 4.1.3 we see that this implication does not hold for nonpure shellability. In order to extend the implication to the nonpure setting, Stanley [**170**] introduced nonpure versions of Cohen-Macaulay for complexes and rings, called sequentially Cohen-Macaulay, and showed that all shellable complexes are sequentially Cohen-Macaulay. Duval [**64**] and Wachs [**201**] found similar simpler characterizations of Stanley's sequential Cohen-Macaulayness for simplicial complexes. Here we take Wachs' characterization as the definition.

Definition 4.1.4. Let Δ be a simplicial complex. For each $m = 1, 2, \ldots, \dim(\Delta)$, let $\Delta^{\langle m \rangle}$ be the subcomplex of Δ generated by facets of dimension at least m. The complex Δ is said to be *sequentially acyclic over* $\mathbf{k}$ if $\tilde{H}_i(\Delta^{\langle m \rangle}; \mathbf{k}) = 0$ for all $i < m$. We say that Δ is *sequentially Cohen-Macaulay over* $\mathbf{k}$ if $\mathrm{lk}_\Delta F$ is sequentially acyclic over $\mathbf{k}$ for all $F \in \Delta$.

Exercise 4.1.5. The *pure m-skeleton* $\Delta^{[m]}$ of a simplicial complex Δ is defined to be the subcomplex generated by all faces of dimension m. Show that the formulation of sequentially Cohen-Macaulay complex given in Definition 4.1.4 is equivalent to the following formulation of Duval [**64**]: Δ is sequentially Cohen-Macaulay if and only if $\Delta^{[m]}$ is Cohen-Macaulay for all $m = 1, 2, \ldots, \dim(\Delta)$.

It is clear that pure sequential Cohen-Macaulayness is the same thing as Cohen-Macaulayness. The following generalization of Theorem 4.1.2 shows that sequential Cohen-Macaulayness is a topological property, just as Cohen-Macaulayness is. Given a nonnegative integer m and a topological space X, define $X^{\langle m\rangle}$ to be the topological closure of the set

$$\{p \in X : p \text{ has a neighborhood homeomorphic to an open } d\text{-ball where } d \geq m\}.$$

Theorem 4.1.6. *The simplicial complex Δ is sequentially Cohen-Macaulay over $\mathbf{k}$ if and only if for all $i < m$ and $p \in \|\Delta\|^{\langle m\rangle}$,*

$$\tilde{H}_i(\|\Delta\|^{\langle m\rangle}; \mathbf{k}) = H_i(\|\Delta\|^{\langle m\rangle}, \|\Delta\|^{\langle m\rangle} - p; \mathbf{k}) = 0.$$

Exercise 4.1.7. Munkres proof of Theorem 4.1.2 is based on the following lemma: For any face F of a simplicial complex Δ, any point p in the interior of F, and any integer i,

$$\tilde{H}_i(\mathrm{lk} F; \mathbf{k}) \cong H_{i+\dim F+1}(\|\Delta\|, \|\Delta\| - p; \mathbf{k}).$$

Use this lemma to prove Theorem 4.1.6.

Exercise 4.1.8 (Wachs [**201**])**.** Show that if Δ is sequentially acyclic then $\tilde{H}_i(\Delta; \mathbb{Z})$ is free for all i and vanishes whenever there is no facet of dimension i.

Theorem 4.1.9 (Stanley [**170**])**.** *Every shellable simplicial complex is sequentially Cohen-Macaulay over $\mathbf{k}$ for all $\mathbf{k}$.*

Exercise 4.1.10. Prove Theorem 4.1.9 by first showing that if Δ is shellable then there is a shelling order of the facets of Δ in which the dimensions of the facets weakly decrease (see Björner and Wachs [**40**]).

We say that a poset is (sequentially) Cohen-Macaulay if its order complex is (sequentially) Cohen-Macaulay. Early work on Cohen-Macaulay posets can be found in the paper of Baclawski [**12**] and in the survey article of Björner, Garsia, and Stanley [**32**]. We have the following nice characterization of sequentially Cohen-Macaulay posets, which, in the pure case, is well-known and follows easily from Definition 4.1.1.

Theorem 4.1.11 (Björner, Wachs, and Welker [**44**])**.** *A poset P is sequentially Cohen-Macaulay if and only if every open interval of $\hat{P}$ is sequentially acyclic.*

One direction of the proof follows immediately from the fact that open intervals are links of chains. The other is a consequence of the fact that joins of sequentially acyclic simplicial complexes are sequentially acyclic, which is proved in [**44**] by using a fiber theorem of Quillen. Fiber theorems are discussed in Lecture 5. It is also shown in [**44**] that other poset operations such as product preserve sequential Cohen-Macaulayness, a fact that had been known for some time in the pure case [**207**]; see Exercise 5.1.6.

To compute the unique nonvanishing Betti number of a Cohen-Macaulay poset (or any poset in which homology is concentrated in a single dimension), one needs

only to compute its Möbius invariant. Indeed, it follows from the Phillip Hall Theorem (Proposition 1.2.6) and the Euler-Poincaré formula (Theorem 1.2.8) that if the homology of P vanishes below the top dimension then

$$\tilde{\beta}_{l(P)}(P) = (-1)^{l(P)}\mu(\hat{P}). \tag{4.1.1}$$

By applying the recursive definition of Möbius function we get the recursive formula observed by Sundaram in [**177**],

$$\tilde{\beta}_{l(P)-1}(P \setminus \{\hat{0}\}) = \sum_{x \in P} (-1)^{l(P)+r(x)}\tilde{\beta}_{r(x)-2}(\hat{0}, x), \tag{4.1.2}$$

where P is a Cohen-Macaulay poset with a bottom element $\hat{0}$ and $r(x)$ is the rank of x.

Since general sequentially Cohen-Macaulay posets can have homology in multiple dimensions, (4.1.1) and (4.1.2) do not hold in the nonpure setting. However, if the sequentially Cohen-Macaulay poset has a property known as semipure, then the following generalization of (4.1.2) shows that Möbius function can still be used to compute its Betti numbers. A poset P is said to be *semipure* if $P_{\leq y} := \{x \in P : x \leq y\}$ is pure for all $y \in P$. The proper part of the poset given in Figure 3.2.1 is semipure. Also the face poset of any simplicial complex is semipure.

Theorem 4.1.12 (Wachs [**201**]). *Let P be a semipure sequentially Cohen-Macaulay poset with a bottom element $\hat{0}$. Then for all m,*

$$\tilde{\beta}_{m-1}(P \setminus \{\hat{0}\}) = \sum_{\substack{x \in P \\ m(x) = m}} (-1)^{m+r(x)}\tilde{\beta}_{r(x)-2}(\hat{0}, x), \tag{4.1.3}$$

where $m(x)$ is the length of the longest chain of P containing x.

Exercise 4.1.13. Prove Theorem 4.1.12.

We will demonstrate the effectiveness of the formulas (4.1.2) and (4.1.3) in computing Betti numbers on concrete examples in Section 4.3. Equivariant versions of these formulas will be discussed in Section 4.4. For more general versions see [**201**].

There is a homotopy version of Cohen-Macaulay complexes due to Quillen [**137**] and a homotopy version of sequentially Cohen-Macaulay complexes studied by Björner, Wachs and Welker [**43, 44**]. The requirement that homology vanish below a certain dimension is replaced by the requirement that the homotopy groups vanish below that dimension. The homotopy versions are stronger than the homology versions, but are still still weaker than shellability. They are not topological properties, however. For instance there is a triangulation of the 5-sphere that is not homotopy Cohen-Macaulay, see [**137**, Section 8].

We summarize the implications for bounded posets in the following diagram.

$$\begin{array}{ccc}
\text{pure EL-shellable} & \Longrightarrow & \text{EL-shellable} \\
\Downarrow & & \Downarrow \\
\text{pure CL-shellable} & \Longrightarrow & \text{CL-shellable} \\
\Downarrow & & \Downarrow \\
\text{pure shellable} & \Longrightarrow & \text{shellable} \\
\Downarrow & & \Downarrow \\
\text{homotopy Cohen-Macaulay} & \Longrightarrow & \text{sequentially homotopy Cohen-Macaulay} \\
\Downarrow & & \Downarrow \\
\text{Cohen-Macaulay over } \mathbb{Z} & \Longrightarrow & \text{sequentially Cohen-Macaulay over } \mathbb{Z} \\
\Downarrow & & \Downarrow \\
\text{Cohen-Macaulay over } \mathbf{k} & \Longrightarrow & \text{sequentially Cohen-Macaulay over } \mathbf{k}
\end{array}$$

It is known that all the implications but

$$\text{(pure) EL-shellable} \Longrightarrow \text{(pure) CL-shellable}$$

are strict. Whether or not there are CL-shellable posets that are not EL-shellable is an open question. It is also unknown whether or not CL-shellability and dual CL-shellability are equivalent. Examples of shellable posets that are not CL-shellable were obtained by Vince and Wachs [**196**] and Walker [**208**].

There are other important properties of simplicial complexes which have recursive formulations such as vertex decomposability (introduced by Provan and Billera [**135**] for pure simplicial complexes and extended to nonpure simplicial complexes by Björner and Wachs [**41**]) and constructible complexes (introduced by Hochster [**97**] and extended to the nonpure case by Jonsson [**105**]). There are no special poset versions of these tools, however, as there are for shellability.

4.2. Recursive Atom Orderings

In this section we present a rather technical, but very useful tool for establishing poset shellability (and in turn sequential Cohen-Macaulayness). This technique, called recursive atom ordering, was introduced by Björner and Wachs in the early 1980's. It is equivalent to CL-shellability, but does not involve edge labelings.

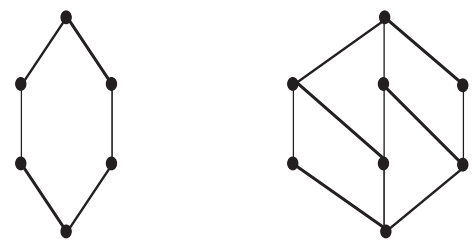
Definition 4.2.1. A bounded poset P is said to *admit a recursive atom ordering* if its length $l(P)$ is 1, or if $l(P) > 1$ and there is an ordering $a_1, a_2, \ldots, a_t$ of the atoms of P that satisfies:

(i) For all $j = 1, 2, \ldots, t$ the interval $[a_j, \hat{1}]$ admits a recursive atom ordering in which the atoms of $[a_j, \hat{1}]$ that belong to $[a_i, \hat{1}]$ for some $i < j$ come first.

(ii) For all $i < j$, if $a_i, a_j < y$ then there is a $k < j$ and an atom z of $[a_j, \hat{1}]$ such that $a_k < z \leq y$.

A *recursive coatom ordering* is a recursive atom ordering of the dual poset P^*.

Figure 4.2.1 (a) gives an example of a poset that does not admit a recursive atom ordering since condition (ii) fails for every ordering of the atoms. The poset in Figure 4.2.1 (b) does admit a recursive atom ordering. The left to right order in the drawing of the Hasse diagram gives an ordering of the atoms that satisfies (i) and (ii), for *each* interval.



(a) (b)

Figure 4.2.1

Theorem 4.2.2 (Björner and Wachs [**38, 40**]). *A bounded poset P is CL-shellable if and only if P admits a recursive atom ordering.*

Proof idea. We use a recursive procedure to obtain a recursive atom ordering from a CL-labeling and vice versa. Given a CL-labeling λ, order the atoms of P in increasing order of the labels $\lambda(c, \hat{0} < a)$, where c is any maximal chain containing the atom a (the label is independent of the maximal chain c since $\hat{0} < a$ is a bottom edge). Then recursively use the restriction of the CL-labeling to each interval $[a, \hat{1}]$ to obtain a recursive atom ordering of $[a, \hat{1}]$.

Conversely, given a recursive atom ordering of P, label the bottom edge of each maximum chain with the position of the atom in the atom ordering. Then recursively use the recursive atom ordering of $[a, \hat{1}]$ to obtain an appropriate chain-edge labeling of $[a, \hat{1}]$, for each atom a. □

A bounded poset P is said to be *semimodular* if for all $u, v \in P$ that cover some $x \in P$, there is an element $y \in P$ that covers both u and v. Semimodular posets are not necessarily shellable. A poset is said to be *totally semimodular* if every closed interval is semimodular. Note that semimodular lattices are totally semimodular. In particular, the boolean algebra and the partition lattice are totally semimodular.

Theorem 4.2.3 (Björner and Wachs [**38**]). *Every ordering of the atoms of a totally semimodular poset is a recursive atom ordering.*

Theorem 4.2.4 (Björner and Wachs [**38**]). *An ordering of the facets of a simplicial complex Δ is a shelling if and only if the ordering is a recursive coatom ordering of the face lattice $L(\Delta)$.*

Exercise 4.2.5. Prove

(a) Theorem 4.2.3.
(b) Theorem 4.2.4

Definition 4.2.6 (Wachs and Walker [**205**]). A meet semilattice P with rank function r is said to be a *geometric semilattice* if

(i) every interval is a geometric lattice
(ii) for all $x \in P$ and subset A of atoms whose join exists, if $r(x) < r(\bigvee A)$ then there is an $a \in A$ such that $a \not\leq x$ and $a \vee x$ exists.

Examples of geometric semilattices include

- geometric lattices

- intersection semilattices of affine hyperplane arrangements
- face posets of matroid complexes
- top truncated geometric lattices; i.e., L_S where L is a geometric lattice and
$$S = \{0, 1, 2, \ldots, j\}$$
for some $j \leq l(L)$
- $L - [x, \hat{1}]$, where L is a geometric lattice and $x \in L$ (this is a characterization).

A set A of atoms in a geometric semilattice P is said to be *independent* if $\bigvee A$ exists and $r(\bigvee A) = |A|$. A *basic* set of atoms is a maximal independent set of atoms.

Theorem 4.2.7 (Wachs and Walker [**205**]). *Any ordering of the atoms of a geometric semilattice P that begins with a basic set of atoms is a recursive atom ordering of $P \cup \{\hat{1}\}$.*

Exercise 4.2.8. Let P be a geometric semilattice.

(a) Show P is pure.
(b) Show every principal upper order ideal $P_{\geq x} := \{y \in P : y \geq x\}$ is a geometric semilattice.
(c) Prove Theorem 4.2.7.

From the proof of Theorem 4.2.2, we know that to every recursive atom ordering of a poset one can associate a CL-labeling. A natural question to ask, when given a particular recursive atom ordering, is whether there is a nice associated CL-labeling. Then one can obtain Betti numbers and bases from the decreasing chains. Ziegler [**217**] describes such a CL-labeling for one of the recursive atom orderings of Theorem 4.2.7 and shows that it is in fact an EL-labeling. He then constructs bases for homology and cohomology of geometric semilattices which generalize the Björner NBC bases for geometric lattices discussed in Section 3.2.3.

4.3. More Examples

In this section we present examples of semipure posets that admit recursive atom orderings and we demonstrate the use of the recursive formulas (4.1.2) and (4.1.3) in computing their Betti numbers. Let's begin with some *pure* examples.

Example 4.3.1. Let $k \leq n$. The injective word poset $\mathcal{I}_{n,k}$ is defined to be the poset of words of length at most k over alphabet $[n]$ with no repeated letters. The order relation is the subword relation. This poset is pure and is a lower order ideal of the normal word poset $\mathcal{N}_n$ (discussed in Exercise 3.3.8) which in turn is a subposet of the word poset $\mathcal{W}_n$ (also discussed in Exercise 3.3.8). It is left as an exercise to show that lexicographic order on the injective words of length k is a recursive coatom ordering of $\mathcal{I}_{n,k} \cup \{\hat{1}\}$. One can use the fact that all closed intervals of $\mathcal{I}_{n,k}$ are isomorphic to Boolean algebras. Since the duals of the closed intervals are totally semimodular, one needs only to verify condition (i) of Definition 4.2.1.

Exercise 4.3.2.

(a) Show that lexicographic order on the injective words of length k is a recursive coatom ordering of $\mathcal{I}_{n,k} \cup \{\hat{1}\}$.

(b) Use formula (4.1.2) to show that the top Betti number $\tilde{\beta}_{k-1}(\bar{I}_{n,n})$ is equal to d_n, the number of derangements of n elements. Consequently, $\bar{I}_{n,n}$ has the homotopy type of a wedge of d_n spheres of dimension $k-1$.

(c) Let $\mathcal{N}_{n,k}$ be the induced subposet of $\mathcal{N}_n$ consisting of words of length at most k. Use Exercise 3.3.8, Theorem 3.4.1, and formula (4.1.2) to show that $\bar{\mathcal{N}}_{n,k}$ has the homotopy type of a wedge of $(n-1)^k$ spheres of dimension $k-1$.

(d) Let $\mathcal{W}_{n,k}$ be the induced subposet of $\mathcal{W}_n$ consisting of words of length at most k. Show that $\bar{\mathcal{W}}_{n,k}$ also has the homotopy type of a wedge of $(n-1)^k$ spheres of dimension $k-1$.

Remark 4.3.3. The injective word poset and normal word poset of Example 4.3.1 were introduced by Farmer [**68**] who showed that $\bar{\mathcal{I}}_{n,k}$ and $\bar{\mathcal{N}}_{n,k}$ have the homology of a wedge of $(k-1)$-spheres. Björner and Wachs [**38**] recovered these results and strengthened them to homotopy by establishing shellability. The Betti number computations of Exercise 4.3.2 (b) and (c) are due to Reiner and Webb [**142**] and Farmer [**68**], respectively.

Recall that in Example 3.2.2, we used EL-shellability to show that the proper part of the partition lattice Π_n has the homotopy type of a wedge of $(n-1)!$ spheres of dimension $n-3$. In the next two examples we use the techniques of Sections 4.1 and 4.2 to obtain analogous results for the even and odd block size partition posets.

Example 4.3.4. Let Π_{2n}^{even} be the subposet of Π_{2n} consisting of partitions whose block sizes are even. The even block size partition poset is pure but lacks a bottom element. The upper intervals $[x,\hat{1}]$ are all partition lattices, which are totally semimodular; so every atom ordering of $[x,\hat{1}]$ is a recursive atom ordering. Associate a word with each atom of $\Pi_{2n}^{\text{even}} \cup \{\hat{0}\}$ by listing the elements of each block in increasing order and then listing the blocks in lexicographic order. We claim that lexicographic order on the words corresponds to a recursive atom ordering of $\Pi_{2n}^{\text{even}} \cup \{\hat{0}\}$. The verification is left as an exercise.

We next use (4.1.2) to compute the Betti numbers of $\bar{\Pi}_{2n}^{\text{even}}$. Since Π_{2n}^{even} is pure and the upper intervals are isomorphic to partition lattices, to compute the unique nonvanishing Betti number $\tilde{\beta}_{n-2}(\bar{\Pi}_{2n}^{\text{even}})$, we apply (4.1.2) to the dual of the poset. Let β_{2n} denote the Betti number $\tilde{\beta}_{n-2}(\bar{\Pi}_{2n}^{\text{even}})$. By (4.1.2) we have

$$\begin{aligned}
\beta_{2n} &= \sum_x (-1)^{b(x)+n} \tilde{\beta}_{b(x)-3}(x,\hat{1}) \\
&= \sum_x (-1)^{b(x)+n} (b(x)-1)! \\
&= \sum_{r=1}^{n} (-1)^{r+n} (r-1)!\ |\{x \in \Pi_{2n}^{\text{even}} : b(x)=r\}|,
\end{aligned}$$

where $b(x)$ denotes the number of blocks of x. Note that

$$|\{x \in \Pi_{2n}^{\text{even}} : b(x)=r\}| = \frac{1}{r!} \sum_{(j_1,j_2,\ldots,j_r)\vDash n} \binom{2n}{2j_1, 2j_2, \ldots, 2j_r},$$

where $\vDash n$ denotes composition of n. We now have

$$\beta_{2n} = \sum_{r=1}^{n} (-1)^{r+n} \frac{1}{r} \sum_{(j_1,j_2,\ldots,j_r)\vDash n} \binom{2n}{2j_1, 2j_2, \ldots, 2j_r}.$$

The exponential generating function for the Betti numbers is thus given by,

$$\begin{aligned}\sum_{n\geq 1}\beta_{2n}\frac{u^{2n}}{(2n)!} &= \sum_{r\geq 1}(-1)^r\frac{1}{r}\sum_{n\geq 1}\sum_{(j_1,j_2,\ldots,j_r)\vDash n}\binom{2n}{2j_1,2j_2,\ldots,2j_r}(-1)^n\frac{u^{2n}}{(2n)!}\\ &= \sum_{r\geq 1}(-1)^r\frac{1}{r}\Big(\sum_{j\geq 1}(-1)^j\frac{u^{2j}}{(2j)!}\Big)^r\\ &= -\ln(\cos u). \qquad (4.3.1)\end{aligned}$$

By taking derivatives we get,

$$\sum_{n\geq 1}\beta_{2n}\frac{u^{2n-1}}{(2n-1)!} = \tan u.$$

So β_{2n} is equal to the coefficient of $\frac{u^{2n-1}}{(2n-1)!}$ in the Taylor series expansion of $\tan u$. It is well known that this coefficient is equal to the Euler number E_{2n-1}, where E_m is defined to be the number of alternating permutations in $\mathfrak{S}_m$, i.e., permutations with descents at all the even positions and ascents at all the odd positions. We conclude that $\bar{\Pi}_{2n}^{\text{even}}$ has the homotopy type of a wedge of E_{2n-1} spheres of dimension $n-2$.

There is an alternative way to arrive at $\beta_{2n} = E_{2n-1}$. Replace the last line of (4.3.1) with

$$-\ln\Big(\sum_{j\geq 0}(-1)^j\frac{u^{2j}}{(2j)!}\Big)$$

and then take derivatives of both sides of the equation. This results in

$$\sum_{n\geq 1}\beta_{2n}\frac{u^{2n-1}}{(2n-1)!}\sum_{j\geq 0}(-1)^j\frac{u^{2j}}{(2j)!} = \sum_{j\geq 0}(-1)^{j-1}\frac{u^{2j-1}}{(2j-1)!}.$$

By equating coefficients we obtain the recurrence relation,

$$\sum_{r=1}^{j}\binom{2j-1}{2r-1}(-1)^r\beta_{2r} = -1$$

It is not difficult to check that E_{2r-1} satisfies the same recurrence relation.

Example 4.3.5. Let Π_{2n+1}^{odd} be the subposet of Π_{2n+1} consisting of partitions whose block sizes are odd. The odd block size partition poset is totally semimodular; so any atom ordering is a recursive atom ordering.

Now let β_{2n+1} denote the top Betti number $\tilde{\beta}_{n-2}(\bar{\Pi}_{2n+1}^{\text{odd}})$. An argument similar to the one used in deriving (4.3.1) yields: $\sum_{n\geq 0}\beta_{2n+1}\frac{u^{2n+1}}{(2n+1)!}$ is the compositional inverse of $\sum_{n\geq 0}(-1)^n\frac{u^{2n+1}}{(2n+1)!}$, i.e.,

$$\sum_{n\geq 0}\beta_{2n+1}\frac{1}{(2n+1)!}\Big(\sum_{j\geq 0}(-1)^j\frac{u^{2j+1}}{(2j+1)!}\Big)^{2n+1} = u. \qquad (4.3.2)$$

Since $\sum_{j\geq 0}(-1)^j\frac{u^{2j+1}}{(2j+1)!} = \sin u$, we have $\sum_{n\geq 0}\beta_{2n+1}\frac{u^{2n+1}}{(2n+1)!} = \sin^{-1} u$. The coefficient of $\frac{u^{2n+1}}{(2n+1)!}$ in the Taylor series expansion of $\sin^{-1} u$ is $(2n+1)!!^2$, where $(2n+1)!! := 1\cdot 3\cdot 5\cdots(2n+1)$. We conclude that $\bar{\Pi}_{2n+1}^{\text{odd}}$ has the homotopy type of a wedge of $(2n+1)!!^2$ spheres of dimension $n-2$.

Exercise 4.3.6. For integers $n, d \geq 1$ and $k \geq 0$, let $\Pi_{nd+k}^{k \bmod d}$ be the subposet of the partition lattice Π_{nd+k} consisting of partitions whose block sizes are congruent to $k \bmod d$. This poset is pure only for $k \equiv 0$ or $1 \mod d$, and is bounded only for $k \equiv 1 \mod d$.

(a) Show that lexicographical order on the words associated with the atoms as in Example 4.3.4 (i.e., list the elements of each block in increasing order and then list the blocks in lexicographical order) yields a recursive atom ordering of $\Pi_{nd}^{0 \bmod d} \cup \{\hat{0}\}$.

(b) Show that $\Pi_{nd+1}^{1 \bmod d}$ admits a recursive atom ordering.

(c) Find a recursive atom ordering of $\Pi_{nd+k}^{k \bmod d} \cup \{\hat{0}\}$ for $k \not\equiv 1 \mod d$.

(d) Show that $\bar{\Pi}_{nd}^{0 \bmod d}$ has the homotopy type of a wedge of E_{dn-1}^{d} spheres of dimension $n-2$, where

$$E_{dn-1}^{d} = |\{\sigma \in \mathfrak{S}_{nd-1} : \mathrm{des}(\sigma) = \{d, 2d, \ldots, (n-1)d\}\}|.$$

(e) Show that $\bar{\Pi}_{nd+1}^{1 \bmod d}$ has the homotopy type of a wedge of c_n spheres of dimension $n-2$, where $\sum_{n \geq 0} c_n \frac{u^{nd+1}}{(nd+1)!}$ is the compositional inverse of $\sum_{n \geq 0} (-1)^n \frac{u^{nd+1}}{(nd+1)!}$.

Remark 4.3.7. The result on the Möbius invariant of the even block size partition lattice derived in Example 4.3.4 first appeared in in the 1976 MIT thesis of Garrett Sylvestor on Ising ferromagnets [**188**]. Stanley [**167**] extended this result to the d-divisible partition lattice of Exercise 4.3.6 (d). The Möbius invariant results derived in Example 4.3.5 and Exercise 4.3.6 (e) are also due to Stanley. The recursive atom ordering for the d-divisible partition lattice (Example 4.3.4 and Exercise 4.3.6 (a)) is due to Wachs (see [**150**]), as is an EL-labeling of this poset [**198**]. The recursive atom ordering for the 1 mod d partition lattice (Example 4.3.5 and Exercise 4.3.6 (b)) is due to Björner (see [**54**] and [**38**]). The general recursive atom ordering of $\Pi_{nd+k}^{k \bmod d} \cup \{\hat{0}\}$ of Exercise 4.3.6 (c) is due to Wachs [**201**].

In the next example, we demonstrate the full power of Theorem 4.1.12 by finding the Betti numbers for a nonpure shellable poset.

Example 4.3.8. For $n \geq k \geq 3$, let $\Pi_n^{\geq k}$ be the subposet of Π_n consisting of partitions whose block sizes are at least k. It was shown by Björner and Wachs [**40**] that the poset $\Pi_n^{\geq k} \cup \{\hat{0}\}$ admits a recursive atom ordering similar to that of $\Pi_{2n}^{\mathrm{even}} \cup \{\hat{0}\}$. The following computation of Betti numbers appears in [**201**]. The dual of $\Pi_n^{\geq k}$ is semipure and the upper intervals are partition lattices; so we apply (4.1.3) to the dual. First note that

$$m(x) = \sum_{i=1}^{b(x)} \left\lfloor \frac{|B_i|}{k} \right\rfloor - 1, \tag{4.3.3}$$

where $B_1, B_2, \ldots, B_{b(x)}$ are the blocks of x. By (4.1.3) we have,

$$\begin{aligned} \tilde{\beta}_{m-2}(\bar{\Pi}_n^{\geq k}) &= \sum_{x : m(x) = m-1} (-1)^{b(x)+m} \tilde{\beta}_{b(x)-3}(x, \hat{1}) \\ &= \sum_{r \geq 1} (-1)^{r+m} (r-1)! \; |\{x \in \Pi_n^{\geq k} : b(x) = r,\, m(x) = m-1\}|. \end{aligned}$$

We have,

$$|\{x \in \Pi_n^{\geq n} : b(x) = r,\, m(x) = m-1\}| = \frac{1}{r!} \sum_{\substack{(j_1, j_2, \ldots, j_r) \vDash n \\ j_i \geq k\ \forall i \\ \sum_{i=1}^r \lfloor j_i/k \rfloor = m}} \binom{n}{j_1, j_2, \ldots, j_r}.$$

The two parameter exponential generating function is thus,

$$\begin{aligned}
&\sum_{m,n\geq 1} (-1)^m \tilde{\beta}_{m-2}(\bar{\Pi}_n^{\geq k})\, t^m \frac{u^n}{n!} \\
&\quad = \sum_{r\geq 1} (-1)^r \frac{1}{r} \sum_{n,m\geq 1} \sum_{\substack{(j_1, j_2, \ldots, j_r) \vDash n \\ j_i \geq k\ \forall i}} \binom{n}{j_1, j_2, \ldots, j_r}\, t^{\sum_{i=1}^r \lfloor j_i/k \rfloor} \frac{u^n}{n!} \\
&\quad = \sum_{r\geq 1} (-1)^r \frac{1}{r} \Big(\sum_{j\geq k} t^{\lfloor j/k \rfloor} \frac{u^j}{j!} \Big)^r \\
&\quad = -\ln\Big(1 + \sum_{j\geq k} t^{\lfloor j/k \rfloor} \frac{u^j}{j!}\Big).
\end{aligned}$$

Exercise 4.3.9. Show that $\Pi_n^{\geq k} \cup \{\hat{0}\}$ has a recursive atom ordering.

Problem 4.3.10. Linusson [**121**] computed the Möbius invariant of the poset $\Pi_n^{\geq k} \cup \{\hat{0}\}$ in order to compute lower bounds for the complexity of a problem similar to the k-equal problem of Section 3.2.4; namely that of determining whether a given list of real numbers has the property that the number of occurrences of each entry is at least k. It was shown by Björner and Lovász [**33**] that Betti number computations give better bounds than Möbius function computations. Can the Betti number computation of Example 4.3.8 be used to improve Linusson's lower bound for the complexity of the "at least k problem"?

Exercise 4.3.11 (Wachs [**201**])**.** In this exercise, we generalize the results of Examples 4.3.4 and 4.3.5 to the nonpure case of Exercise 4.3.6 . For positive integers n, d, k, with $k \leq d$, let $k_0 = k/\gcd(k,d)$ and $d_0 = d/\gcd(k,d)$. Show that

$$\sum_{\substack{m\,\geq\,1 \\ n\,\geq\,0}} (-1)^m \beta_{m-2}(\bar{\Pi}_{nd+k}^{k \bmod d})\, t^{md_0+1} \frac{u^{nd+k}}{(nd+k)!} = f\Big(\sum_{i\geq 0} t^{\lfloor i/k_0 \rfloor d_0+1} \frac{u^{id+k}}{(id+k)!} \Big),$$

where $f(y)$ is the compositional inverse of the formal power series

$$g(y) = \sum_{i\geq 0} \frac{y^{id_0+1}}{(id_0+1)!}.$$

4.4. The Whitney Homology Technique

In this section we discuss a technique for computing group representations on the homology of sequentially Cohen-Macaulay posets and demonstrate its use on the examples of Section 4.3. This technique was introduced by Sundaram [**177**] in the pure case and later generalized to semipure posets by Wachs [**201**]. The pure case is based on an equivariant version of (4.1.2) and the semipure version is based on the more general Theorem 4.1.12.

For any Cohen-Macaulay G-poset P with bottom element $\hat{0}$, *Whitney homology* of P is defined for each integer r as follows,

$$\mathrm{WH}_r(P) = \bigoplus_{x \in P_r} \tilde{H}_{r-2}(\hat{0}, x),$$

where $P_r := \{x \in P : r(x) = r\}$.

The action of G on P induces a representation of G on $\mathrm{WH}_r(P)$. Indeed $g \in G$ takes $(r-2)$-chains of $(\hat{0}, x)$ to $(r-2)$-chains of $(\hat{0}, gx)$. More precisely, as a G-module

$$\mathrm{WH}_r(P) = \bigoplus_{x \in P_r/G} \tilde{H}_{r-2}(\hat{0}, x) \uparrow_{G_x}^{G},$$

where P_r/G is a set of orbit representatives and G_x is the stabilizer of x.

Whitney homology for geometric lattices was introduced by Baclawski [**11**] as the homology of an algebraic complex whose Betti numbers are the signless Whitney numbers of the first kind. The formulation given here is due to Björner [**22**]. Whitney homology for intersection lattices of complex hyperplane arrangements forms an algebra isomorphic to an algebra that Orlik and Solomon used to give a combinatorial presentation of the cohomology algebra of the complement of the arrangement; see [**28**]. Sundaram [**177**] recognized that Whitney homology for any Cohen-Macaulay poset could be used as a tool in computing group representations on homology. Her technique is based on the following result, which she obtains as a consequence of the Hopf trace formula.

Theorem 4.4.1 (Sundaram [**177, 178**])**.** *Suppose P is a G-poset with a bottom element $\hat{0}$. Then*

$$\bigoplus_{r=-1}^{l(P)-1} (-1)^r \tilde{H}_r(P \setminus \{\hat{0}\}) \cong_G \bigoplus_{r=0}^{l(P)} (-1)^{r-1} \bigoplus_{x \in P/G} \tilde{H}_{r-2}(\hat{0}, x) \uparrow_{G_x}^{G}. \tag{4.4.1}$$

Consequently, if P is also Cohen-Macaulay,

$$\tilde{H}_{l(P)-1}(P \setminus \{\hat{0}\}) \cong_G \bigoplus_{r=0}^{l(P)} (-1)^{l(P)+r} \mathrm{WH}_r(P). \tag{4.4.2}$$

Exercise 4.4.2.

(a) Prove Theorem 4.4.1.

(b) Show that if P is the face poset of a simplicial complex then (4.4.1) reduces to the Hopf trace formula.

Example 4.4.3 (Reiner and Webb [**142**])**.** Consider the injective word poset $\mathcal{I}_{n,k}$ of Example 4.3.1. Let G be the symmetric group $\mathfrak{S}_n$, which acts on $\mathcal{I}_{n,k}$ in the obvious way. Since $\mathcal{I}_{n,k}$ is Cohen-Macaulay, we can apply (4.4.2). Let $x \in \mathcal{I}_{n,k}$ be a word of length r. Clearly G_x is isomorphic to the Young subgroup $(\mathfrak{S}_1)^{\times r} \times \mathfrak{S}_{n-r}$, since the letters of x must be fixed and the letters outside of x may be freely permuted. G_x acts trivially on the letters outside of x. The interval $(\hat{0}, x)$ is isomorphic to the proper part of the Boolean algebra B_r; so its top homology is 1-dimensional. Hence,

$$\tilde{H}_{r-2}(\hat{0}, x) \cong_{G_x} \underbrace{S^{(1)} \otimes \cdots \otimes S^{(1)}}_{r} \otimes S^{(n-r)}$$

Since $\mathfrak{S}_n$ acts transitively on each rank row of $\mathcal{I}_{n,k}$, we have

$$\begin{aligned} \mathrm{WH}_r(\mathcal{I}_{n,k} \cup \{\hat{1}\}) &= \tilde{H}_{r-2}(\hat{0}, x) \uparrow_{G_x}^{\mathfrak{S}_n} \\ &\cong_{\mathfrak{S}_n} (S^{(1)})^{\bullet r} \bullet S^{(n-r)}. \end{aligned}$$

It follows from (4.4.2) that

$$\begin{aligned} \tilde{H}_{k-1}(\bar{\mathcal{I}}_{n,k}) &\cong_{\mathfrak{S}_n} \bigoplus_{r=0}^{k} (-1)^{k-r} \ (S^{(1)})^{\bullet r} \bullet S^{(n-r)} \\ &\cong_{\mathfrak{S}_n} (-1)^k S^{(n)} \ \oplus \ S^{(1)} \bullet \bigoplus_{r=0}^{k-1} (-1)^{k-1-r} \ (S^{(1)})^{\bullet r} \bullet S^{(n-r-1)} \\ &\cong_{\mathfrak{S}_n} (-1)^k S^{(n)} \ \oplus \ S^{(1)} \bullet \tilde{H}_{k-2}(\bar{\mathcal{I}}_{n-1,k-1}). \end{aligned}$$

This recurrence relation, for $n = k$, is an equivariant version of the well-known recurrence relation for derangement numbers: $d_n = (-1)^n + nd_{n-1}$. From this recurrence relation, one can obtain the following decomposition of $\tilde{H}_{n-1}(\bar{\mathcal{I}}_{n,n})$ into irreducibles:

$$\tilde{H}_{n-1}(\bar{\mathcal{I}}_{n,n}) \cong_{\mathfrak{S}_n} \bigoplus_{\lambda \vdash n} c_\lambda S^\lambda, \tag{4.4.3}$$

where c_λ is the number of standard Young tableaux of shape λ whose first descent is even.

Exercise 4.4.4. Use Exercise 2.2.10 and the recurrence relation to prove (4.4.3).

Note that by taking dimensions on both sides of (4.4.3), and applying the well-known Robinson-Schensted-Knuth correspondence, one recovers an enumerative result of Désarménien [**59**] that the number of derangements in $\mathfrak{S}_n$ is equal to the number of permutations in $\mathfrak{S}_n$ with first descent even. Désarménien gives an elegant direct combinatorial proof of this result. The Frobenius characteristic of the representation $\bigoplus_{\lambda \vdash n} c_\lambda S^\lambda$ was used by Désarménien and Wachs [**60**] to obtain deeper enumerative connections between the two classes of permutations. A refinement of the Reiner-Webb decomposition was given by Hanlon and Hersh [**89**]. In [**141**], Reiner and Wachs use (4.4.3) to obtain a decomposition (into irreducible representations of $\mathfrak{S}_n$) of the eigenspaces of the so called "random to top" operator in card shuffling theory. Type B analogs of these enumerative and representation theoretic results can also be found in [**141**].

Exercise 4.4.5. Recall from Exercises 4.3.2 (c) and (d) that the top Betti number of both $\bar{\mathcal{N}}_{n+1,k}$ and $\bar{\mathcal{W}}_{n+1,k}$ is n^k.

(a) (Shareshian and Wachs) Use (4.4.2) to prove

$$\mathrm{ch}(\tilde{H}_n(\bar{\mathcal{N}}_{n+1,k}) \downarrow_{\mathfrak{S}_n}^{\mathfrak{S}_{n+1}}) = \sum_{t \geq 1} S(k,t) h_1^t h_{n-t},$$

where $S(n,k)$ is the Stirling number of the second kind.

(b) Conclude that $\tilde{H}_n(\bar{\mathcal{N}}_{n+1,k}) \downarrow_{\mathfrak{S}_n}^{\mathfrak{S}_{n+1}}$ is the kth tensor power of $S^{(n)} \oplus S^{(n-1,1)}$ by comparing characters.

(c) Show (a) and (b) hold for $\mathcal{W}_{n+1,k}$. (Part (b) for $\mathcal{N}_{n+1,k}$ and $\mathcal{W}_{n+1,k}$ was originally observed by Stanley by means of the fixed point Möbius invariant, see Section 4.6.)

Exercise 4.4.6 (Sundaram [**177**])**.** Show that by applying (4.4.2) to the Boolean algebra B_n, one obtains the well-known symmetric function identity,

$$\sum_{i=0}^{n}(-1)^i e_i h_{n-i} = 0.$$

Sundaram developed her Whitney homology technique in order to study representations of the symmetric group on various Cohen-Macaulay subposets of the partition lattice. In fact, she applies it to the full partition lattice and obtains a conceptual representation theoretic proof of the following classical result of Stanley. The original proof of Stanley used a computation, due to Hanlon [**85**], of the fixed point Möbius invariant of the partition lattice. This technique is discussed in Section 4.6.

Theorem 4.4.7 (Stanley [**169**])**.** *For all positive integers n,*

$$\tilde{H}_{n-3}(\bar{\Pi}_n)) \cong_{\mathfrak{S}_n} e^{2\pi i/n} \uparrow_{\mathfrak{C}_n}^{\mathfrak{S}_n} \otimes \mathrm{sgn}_n \tag{4.4.4}$$

where $\mathfrak{C}_n$ is the cyclic subgroup of $\mathfrak{S}_n$ generated by $\sigma := (1, 2, \ldots, n)$ and $e^{2\pi i/n}$ denotes the one dimensional representation of $\mathfrak{C}_n$ whose character value at σ is $e^{2\pi i/n}$.

Proof. (Sundaram [**177**]). Theorem 4.4.1 is applied to the dual of the partition lattice. By setting $T = \mathbb{Z}^+$, $b = r + 1$ and $z_i = 1$ in (2.4.3), one obtains

$$\sum_{n>r} \mathrm{ch}\mathrm{WH}_r((\Pi_n)^*) = \mathrm{ch}\tilde{H}_{r-2}(\bar{\Pi}_{r+1})\Big[\sum_{i\geq 1} h_i\Big], \tag{4.4.5}$$

Now (4.4.2) yields,

$$\begin{aligned}\sum_{n\geq 1}(-1)^{n-1}\mathrm{ch}\tilde{H}_{n-2}(\Pi_n \setminus \{\hat{1}\}) &= \sum_{n\geq 1}\sum_{r=0}^{n-1}(-1)^r \mathrm{ch}\mathrm{WH}_r((\Pi_n)^*) \\ &= \sum_{r\geq 0}(-1)^r \mathrm{ch}\tilde{H}_{r-2}(\bar{\Pi}_{r+1})\Big[\sum_{i\geq 1} h_i\Big].\end{aligned}$$

Since $\Delta(\Pi_n \setminus \{\hat{1}\})$ is contractible for all $n > 1$ and is $\{\emptyset\}$ when $n = 1$, it follows that

$$h_1 = \sum_{r\geq 0}(-1)^r \mathrm{ch}\tilde{H}_{r-2}(\bar{\Pi}_{r+1})\Big[\sum_{i\geq 1} h_i\Big].$$

Since h_1 is the plethystic identity,

$$\begin{aligned}\sum_{r\geq 1}(-1)^{r-1}\mathrm{ch}\tilde{H}_{r-3}(\bar{\Pi}_r) &= \Big(\sum_{i\geq 1} h_i\Big)^{[-1]} \\ &= \sum_{d\geq 1}\frac{1}{d}\mu(d)\log(1+p_d)\end{aligned} \tag{4.4.6}$$

where $[-1]$ denotes plethystic inverse, μ is the number theoretic Möbius function, and p_d is the power sum symmetric function. The last equation follows from a formula of Cadogan [**53**], which is also derived in [**177**]. By extracting the degree n term, we have

$$\mathrm{ch}\tilde{H}_{n-3}(\bar{\Pi}_n) = \frac{1}{n}\sum_{d|n}(-1)^{n-n/d}\mu(d)\, p_d^{n/d}. \tag{4.4.7}$$

A standard formula for the character of an induced representation yields

$$\operatorname{ch} e^{2\pi i/n} \uparrow_{\mathfrak{C}_n}^{\mathfrak{S}_n} = \frac{1}{n} \sum_{d|n} \mu(d)\, p_d^{n/d}.$$

By (2.4.1) and Theorem 2.4.3 (b), this together with (4.4.7) implies (4.4.4), as desired. □

The representation $e^{2\pi i/n} \uparrow_{\mathfrak{C}_n}^{\mathfrak{S}_n}$ is a well-studied representation called the *Lie representation* because it is isomorphic to the representation of the symmetric group on the multilinear component of the free Lie algebra on n generators, cf., Theorem 1.6.2.

Exercise 4.4.8 (Sundaram [**177**])**.** Use the Whitney homology technique to prove the following results of Calderbank, Hanlon and Robinson [**54**].

(a) (equivariant version of Exercise 4.3.6 (e))

$$\sum_{n\geq 0} (-1)^n \operatorname{ch} \tilde{H}_{n-2}(\bar{\Pi}_{nd+1}^{1 \bmod d}) = \Big(\sum_{i\geq 0} h_{id+1}\Big)^{[-1]}.$$

(b) (equivariant version of Exercise 4.3.6 (d))

$$\sum_{n\geq 1} (-1)^{n-1} \operatorname{ch} \tilde{H}_{n-2}(\bar{\Pi}_{nd}^{0 \bmod d}) = \Big(\sum_{i\geq 1} h_i\Big)^{[-1]} \Big[\sum_{i\geq 1} h_{id}\Big]. \tag{4.4.8}$$

We now present an equivariant version of Theorem 4.1.12, which extends Theorem 4.4.1 to the nonpure setting. For any semipure sequentially Cohen-Macaulay G-poset P with a bottom element $\hat{0}$, define r,m-Whitney homology to be the G-module

$$\mathrm{WH}_{r,m}(P) := \bigoplus_{\substack{x \in P_r \\ m(x) = m}} \tilde{H}_{r-2}(\hat{0}, x), \tag{4.4.9}$$

where $m(x)$ is the length of the longest chain of P containing x and $P_r := \{x \in P : r(x) = r\}$.

Theorem 4.4.9 (Wachs [**201**])**.** *Let P be a semipure sequentially Cohen-Macaulay G-poset. Then for all m,*

$$\tilde{H}_{m-1}(P \setminus \{\hat{0}\}) \cong_G \bigoplus_{r=0}^{m} (-1)^{m+r}\, \mathrm{WH}_{r,m}(P). \tag{4.4.10}$$

Example 4.4.10 (equivariant version of Example 4.3.8 [**201**])**.** We will use Theorem 4.4.9 and (2.4.3) to obtain the following formula for the two parameter generating function for the homology of $\bar{\Pi}_n^{\geq k}$,

$$\sum_{\substack{m\,\geq\, 1 \\ n\,\geq\, k}} (-1)^{m-1} \operatorname{ch} \tilde{H}_{m-2}(\bar{\Pi}_n^{\geq k})\, u^n\, t^m = \Big(\sum_{i\geq 1} h_i\Big)^{[-1]} \Big[\sum_{i\geq k} h_i\, u^i\ t^{\lfloor \frac{i}{k} \rfloor}\Big]. \tag{4.4.11}$$

We apply (4.4.9) to the dual of $\Pi_n^{\geq k}$,

$$\mathrm{WH}_{r,m}((\Pi_n^{\geq k})^*) = \bigoplus_{\substack{\lambda \,\in\, \mathrm{Par}(T,\, r+1) \\ \lambda \,\vdash\, n \\ m(\lambda) \,=\, m}} \ \bigoplus_{x \in \Pi(\lambda)} \tilde{H}_{r-2}(x, \hat{1}),$$

where $T = \{k, k+1, \dots\}$, $m(\lambda) = \sum_i \lfloor \lambda_i/k \rfloor - 1$ (recall (4.3.3) here), and the remaining notation is defined in Example 2.4.9 . By setting $z_i = u^i t^{\lfloor i/k \rfloor}$ in (2.4.3), we obtain

$$\begin{aligned}\sum_{m,n} \mathrm{ch}\mathrm{WH}_{r,m-1}((\Pi_n^{\geq k})^*)\, u^n\, t^m &= \sum_{\lambda \in \mathrm{Par}(T,r+1)} \Big(\mathrm{ch} \bigoplus_{x \in \Pi(\lambda)} \tilde{H}_{r-2}(x, \hat{1})\Big) u^{|\lambda|}\ t^{\sum_i \lfloor \lambda_i/k \rfloor} \\ &= \mathrm{ch}\tilde{H}_{r-2}(\Pi_{r+1}) \Big[\sum_{i \geq k} h_i\, u^i\ t^{\lfloor \frac{i}{k} \rfloor}\Big].\end{aligned}$$

Thus formula (4.4.11) follows from (4.4.10) and (4.4.6).

The following result generalizes (4.4.8) and (4.4.11). Its proof is similar to that of (4.4.11) described above.

Theorem 4.4.11 (Wachs [**201**])**.** *Suppose $S \subseteq \{2, 3, \dots\}$ is such that S and $\{s - \min S : s \in S\}$ are closed under addition. For $n \in S$, let Π_n^S be the subposet of Π_n consisting of partitions whose block sizes are in S. Then*

$$\sum_{\substack{m \geq 1 \\ n \in S}} (-1)^{m-1} \mathrm{ch}\tilde{H}_{m-2}(\bar{\Pi}_n^S)\, u^n\, t^m = \Big(\sum_{i \geq 1} h_i\Big)^{[-1]} \Big[\sum_{i \in S} h_i\, u^i\ t^{\phi(i)}\Big],$$

where $\phi(i) := \max\{j \in \mathbb{Z}^+ : i - (j-1)\min S \in S\}$.

We remark that the restricted block size partition poset $\Pi_n^S \cup \{\hat{0}\}$ of Theorem 4.4.11 is the intersection lattice of a subspace arrangement, which is discussed further in Section 5.4.

Example 4.4.12 (equivariant version of Exercise 4.3.11 [**201**])**.** Theorem 4.4.9 and a generalization of (2.4.3) given in [**201**] can be used to obtain the following formula for the two parameter generating function for the homology of the j mod d partition poset for $j = 2, 3, \dots, d$:

$$\begin{aligned}&\sum_{\substack{m \geq 1 \\ n \geq 0}} (-1)^m\ \ \mathrm{ch}\tilde{H}_{m-1}(\bar{\Pi}_{nd+j}^{j \bmod d})\, u^{nd+j} t^{md_0+1} \\ &\qquad = \Big(\sum_{i \geq 0} h_{id_0+1}\Big)^{[-1]} \Big[\sum_{i \geq 0} h_{id+j}\, u^{id+j} t^{\lfloor \frac{i}{j_0} \rfloor d_0+1}\Big],\end{aligned}$$

where $j_0 = \frac{j}{\gcd(j,d)}$ and $d_0 = \frac{d}{\gcd(j,d)}$.

Problem 4.4.13. Do the results of this section on restricted block size partition lattices have nice generalizations to Dowling lattices or intersection lattices of Coxeter arrangements?[1]

For other restricted block size partition posets with very interesting equivariant homology, see the work of Sundaram [**182**].

[1]For nice recent results on this see [**93**].

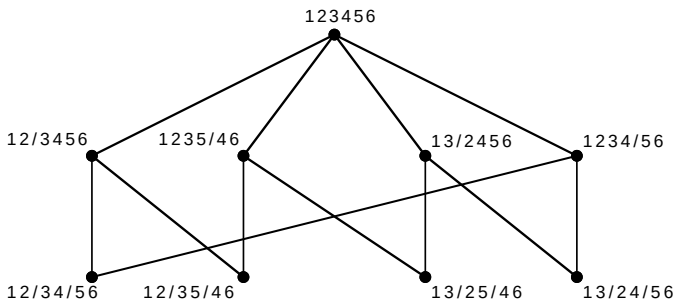


Figure 4.5.1. Π^2_{123456}

4.5. Bases for the Restricted Block Size Partition Posets

As we have seen in previous lectures, the construction of explicit bases for homology and cohomology is an effective tool in studying group representations on homology. In this section we construct bases for the homology and cohomology of the restricted block size partition posets studied in the previous section. These bases are used to obtain further results on the representations of the symmetric group on homology of the posets and to relate these representations to representations on homology of certain interesting graph complexes.

4.5.1. The d-divisible partition lattice $\Pi_{nd}^{0 \bmod d}$

We construct analogs of the splitting basis for the homology of the partition lattice given Section 3.2.2 and its dual basis for cohomology. To *switch-split* a permutation σ at position j is to form the partition

$$\sigma(1), \ldots, \sigma(j-1), \sigma(j+1) \ \mid \ \sigma(j), \sigma(j+2) \ldots, \sigma(n)$$

of $[n]$. Switch-splitting a permutation at two or more nonadjacent positions is defined similarly. Let $d \geq 2$. For each $\sigma \in \mathfrak{S}_{nd}$, let Π_σ^d be the induced subposet of $\Pi_{nd}^{0 \bmod d}$ consisting of partitions obtained by splitting or switch-splitting σ at any number of positions in $\{d, 2d, \ldots, nd\}$. The subposet Π^2_{123456} of $\Pi_6^{0 \bmod 2}$ is shown in Figure 4.5.1.

Each poset $\Pi_\sigma^d \cup \{\hat{0}\}$ is isomorphic to the face lattice C_{n-1} of the $(n-1)$-cross-polytope. Therefore $\Delta(\bar{\Pi}_\sigma^d)$ is an $(n-2)$-sphere embedded in $\Delta(\bar{\Pi}_{nd}^{0 \bmod d})$, and hence it determines a fundamental cycle $\rho_\sigma^d \in \tilde{H}_{n-2}(\bar{\Pi}_{nd}^{0 \bmod d}; \mathbb{Z})$. In each poset $\bar{\Pi}_\sigma^d$, we select a distinguished maximal chain c_σ^d whose k block partition is obtained by splitting σ at positions $d, 2d, \ldots, (k-1)d$, for $k = 2, 3, \ldots, n$. For example,

$$c^2_{123456} = (12/34/56 < 12/3456).$$

Let

$$A_n^d := \{\sigma \in \mathfrak{S}_{nd} : \mathrm{des}(\sigma) = \{d, 2d, \ldots, (n-1)d\}, \ \sigma(nd) = nd\}.$$

Recall that in Exercise 4.3.6 (d), it is stated that $\bar{\Pi}_{nd}^{0 \bmod d}$ has the homotopy type of a wedge of $|A_n^d|$ spheres of dimension $n-2$.

Theorem 4.5.1 (Wachs [**198**]). *The set $\{\rho_\sigma^d : \sigma \in A_n^d\}$ forms a basis for $\tilde{H}_{n-2}(\bar{\Pi}_{nd}^{0 \bmod d}; \mathbb{Z})$ and the set $\{c_\sigma^d : \sigma \in A_n^d\}$ forms a basis for $\tilde{H}^{n-2}(\bar{\Pi}_{nd}^{0 \bmod d}; \mathbb{Z})$.*

The theorem is proved by first showing that for all $\alpha, \beta \in A_n^d$, if $c_\alpha^d \in \Pi_\beta^d$ then $\alpha \leq \beta$ in lexicographic order. This is used to establish linear independence of both $\{\rho_\sigma^d : \sigma \in A_n^d\}$ and $\{c_\sigma^d : \sigma \in A_n^d\}$. The result then follows from Exercise 4.3.6 (d).

The "splitting basis" given in Theorem 4.5.1 is used in [**198**] to give a combinatorial proof of the following result of Calderbank, Hanlon and Robinson, which was first conjectured by Stanley.

Theorem 4.5.2 (Calderbank, Hanlon and Robinson [**54**]). *Let $H_{n,d}$ be the skew hook of size $nd-1$ and descent set $\{d, 2d, \ldots, (n-1)d\}$ (cf. Section 2.2). Then*

$$\tilde{H}_{n-2}(\bar{\Pi}_{nd}^{0 \bmod d}) \downarrow_{\mathfrak{S}_{nd-1}}^{\mathfrak{S}_{nd}} \cong_{\mathfrak{S}_{nd-1}} S^{H_{n,d}}$$

Calderbank, Hanlon, Robinson obtain this result as a consequence of (4.4.8). By mapping tableaux of shape $H_{n,d}$ to maximal chains of $\bar{\Pi}_{nd}^{0 \bmod d}$, one gets a combinatorial proof. Indeed, for each tableaux T of shape $H_{n,d}$, let σ_T be the permutation obtained by reading the entries of the skew hook tableaux T from the southwest end to the northeast end of T and then attaching nd. Now define

$$c_T := c_{\sigma_T}^d \quad \text{and} \quad \rho_T := \rho_{\sigma_T}^d.$$

Theorem 4.5.3 (Wachs [**198**]). *The map $T \mapsto c_T$ induces a well-defined $\mathfrak{S}_{nd-1}$-isomorphism from the skew hook Specht module $S^{H_{n,d}}$ to $\tilde{H}^{n-2}(\bar{\Pi}_{nd}^{0 \bmod d}) \downarrow_{\mathfrak{S}_{nd-1}}^{\mathfrak{S}_{nd}}$.*

To prove this, one first observes that the row permutations leave c_T invariant; then one shows that the Garnir relations map to cohomology relations.

There is a dual version of polytabloid defined for each tableaux T by

$$e_T^* := \sum_{\alpha \in R_\lambda} \sum_{\beta \in C_\lambda} \operatorname{sgn}(\beta)\, T\beta\alpha.$$

(This is actually closer to the traditional notion of polytabloid than the one we gave in Section 2.2.) For each skew or straight shape λ, it is known that

$$\langle e_T^* : T \in \mathcal{T}_\lambda \rangle \cong_{\mathfrak{S}_n} S^\lambda.$$

Theorem 4.5.4 (Wachs [**198**]). *The map $e_T^* \mapsto \rho_T$ induces a well-defined $\mathfrak{S}_{nd-1}$-isomorphism from $S^{H_{n,d}}$ to $\tilde{H}_{n-2}(\bar{\Pi}_{nd}^{0 \bmod d}) \downarrow_{\mathfrak{S}_{nd-1}}^{\mathfrak{S}_{nd}}$.*

All the results of this subsection were generalized to the restricted block size partition posets Π_n^S of Theorem 4.4.11 by Browdy and Wachs [**50, 51**]. The nonpurity of Π_n^S in the general case significantly increases the complexity of the results. The "at least k" partition poset $\Pi_n^{\geq k}$ is an example of such a nonpure poset.

4.5.2. The 1 mod d partition lattice $\Pi_{nd+1}^{1 \bmod d}$

First we describe a basis for top cohomology of $\bar{\Pi}_{nd+1}^{1 \bmod d}$, due to Hanlon and Wachs [**91**], which generalizes the decreasing chain basis $\{\bar{c}_\sigma : \sigma \in \mathfrak{S}_n, \sigma(n) = n\}$ for cohomology of $\bar{\Pi}_n$ given in Section 3.2.2. This basis is used in [**91**] to prove a generalization of Theorem 1.6.2, which relates the $\mathfrak{S}_{nd+1}$-module $\tilde{H}^{n-2}(\bar{\Pi}_{nd+1}^{1 \bmod d})$ to a $(d+1)$-ary version of the free Lie algebra. Then we describe a basis for top homology found about ten years later by Shareshian and Wachs [**159**], which generalizes the tree-splitting basis for homology of $\bar{\Pi}_n$ given in Section 3.2.2. This basis is used in [**159**] to relate the $\mathfrak{S}_{nd+1}$-module $\tilde{H}^{n-2}(\bar{\Pi}_{nd+1}^{1 \bmod d})$ to the homology of graph complexes studied by Linusson Shareshian and Welker [**122**] and Jonsson [**105**].

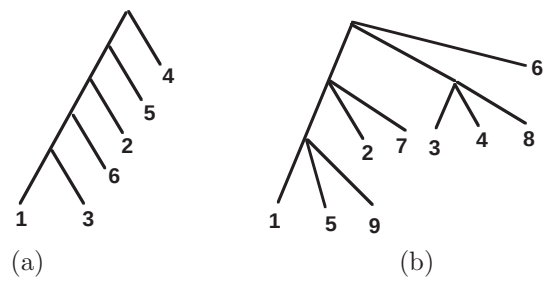


Figure 4.5.2

Let $\mathcal{T}^d_{nd+1}$ be the set of rooted planar $(d+1)$-ary trees on leaf set $[nd+1]$ (i.e., rooted trees in which each internal node has exactly $d+1$ children that are ordered from left to right). For any node x of $T \in \mathcal{T}^d_{nd+1}$, let $m(x)$ be the smallest leaf in the tree rooted at x. A tree T in $\mathcal{T}^d_{nd+1}$ is said to be a d-brush if for each node y of T, the m-values of the children of y increase from left to right, and the child with the largest m-value is a leaf. An example of a 1-brush is given in Figure 4.5.2 (a) and of a 2-brush is given in Figure 4.5.2 (b). Note that every 1-brush looks like a comb, which is the reason for the terminology "brush".

Exercise 4.5.5.

(a) Show that the number of 1-brushes on leaf set $[n+1]$ is $n!$.
(b) Show that the number of 2-brushes on leaf set $[2n+1]$ is $(2n-1)!!^2$.

Recall from Section 1.6 that the postorder traversal of a binary tree on leaf set $[n]$ yields a maximal chain of $\bar{\Pi}_n$. Now we consider a more general construction, which associates a maximal chain c_T of $\bar{\Pi}^{1 \bmod d}_{nd+1}$ to each tree T in $\mathcal{T}^d_{nd+1}$. Each internal node y of T corresponds to a merge of $d+1$ blocks that are the leaf sets of the trees rooted at the $d+1$ children of y. Postorder traversal of internal nodes of T yields a sequence of merges, which corresponds to a maximal chain c_T of $\bar{\Pi}^{1 \bmod d}_{nd+1}$. For example if T is the tree of Figure 4.5.2 (b) then

$$c_T = 159/2/7/3/4/8/6 \lessdot 15927/3/4/8/6 \lessdot 15927/348/6$$

Theorem 4.5.6 (Hanlon and Wachs [**91**])**.** *Let $\mathcal{B}^d_{nd+1}$ be the set of d-brushes in $\mathcal{T}^d_{nd+1}$. The set $\{c_T : T \in \mathcal{B}^d_{nd+1}\}$ forms a basis for $\tilde{H}^{n-2}(\bar{\Pi}^{1 \bmod d}_{nd+1}; \mathbb{Z})$.*

Exercise 4.5.7 (Hanlon and Wachs [**91**])**.** Prove Theorem 4.5.6 by showing

(a) The set $\{c_T : T \in \mathcal{B}^d_{nd+1}\}$ spans $\tilde{H}^{n-2}(\bar{\Pi}^{1 \bmod d}_{nd+1}; \mathbb{Z})$.
(b) $|\mathcal{B}^d_{nd+1}| = |\mu(\Pi^{1 \bmod d}_{nd+1})|$.

We now construct the Shareshian-Wachs basis for homology of $\bar{\Pi}^{1 \bmod d}_{nd+1}$. Just like the tree-splitting basis, which it generalizes, it consists of fundamental cycles of Boolean algebras embedded in $\Pi^{1 \bmod d}_{nd+1}$.

A connected graph G is said to be a *d-clique tree* if either G consists of a single node, or G contains a d-clique, the removal of whose edges results in a graph with d connected components that are all d-clique trees. Note that each edge of a d-clique tree is in a unique d-clique and the removal of the edges of *any* d-clique from a d-clique tree results in a graph with d connected components that are all d-clique

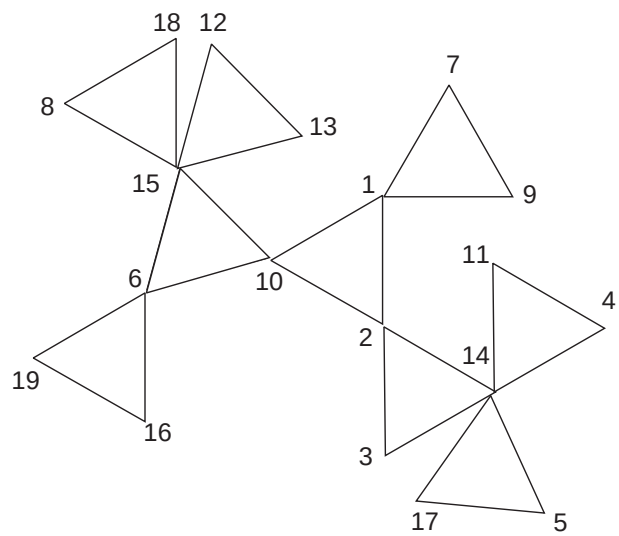


Figure 4.5.3

trees. Note also that a 2-clique tree is an ordinary tree. An example of a 3-clique tree, which we refer to as a triangle tree, is given in Figure 4.5.3.

Given a $(d+1)$-clique tree T on node set $[nd+1]$, one obtains a partition in $\Pi_{nd+1}^{1 \bmod d}$ by choosing any set of $(d+1)$-cliques of T and removing the edges of each clique in the set. The blocks of the partition are the node sets of the connected components of the resulting graph. We say that the partition is obtained by *splitting* the $(d+1)$-clique tree T at the chosen set of $(d+1)$-cliques. For example, the partition obtained by splitting the triangle tree at the shaded triangles in Figure 4.5.4 is

$$19, 16, 6 \ / \ 8, 15, 18, 12, 13 \ / \ 10, 2, 1, 9, 7 \ / \ 3 \ / \ 17, 5, 4, 11, 14.$$

Now let Π_T be the subposet of $\Pi_{nd+1}^{1 \bmod d}$ consisting of partitions obtained by splitting T. Clearly Π_T is isomorphic to the subset lattice B_n. Therefore $\Delta(\bar{\Pi}_T)$ is an $(n-2)$-sphere which determines a fundamental cycle ρ_T.

Theorem 4.5.8 (Shareshian and Wachs [**159**]). *The set of fundamental cycles ρ_T such that T is a $(d+1)$-clique tree on node set $[nd+1]$, spans $\tilde{H}_{n-2}(\bar{\Pi}_{nd+1}^{1 \bmod d}; \mathbb{Z})$.*

This is proved by identifying a set S of $(d+1)$-clique trees, called increasing $(d+1)$-clique trees, and establishing a bijection and unitriangular relationship between $(d+1)$-clique trees and d-brushes. This shows that $\{\rho_T : T \in S\}$ is a basis for $\tilde{H}_{n-2}(\bar{\Pi}_{nd+1}^{1 \bmod d}; \mathbb{Z})$. The increasing 2-clique trees are the increasing trees discussed in Section 3.2.2 and the homology basis is the tree splitting basis.

Next we discuss an application of Theorem 4.5.8 that led to the discovery of the clique tree splitting basis in the first place. Let NPM_{2n} be the poset of nonempty graphs on node set $[2n]$ that don't contain a perfect matching (i.e., a subgraph in which each of the $2n$ vertices has degree 1), ordered by inclusion of edge sets. Linusson, Shareshian, and Welker [**122**] show, using discrete Morse theory, that

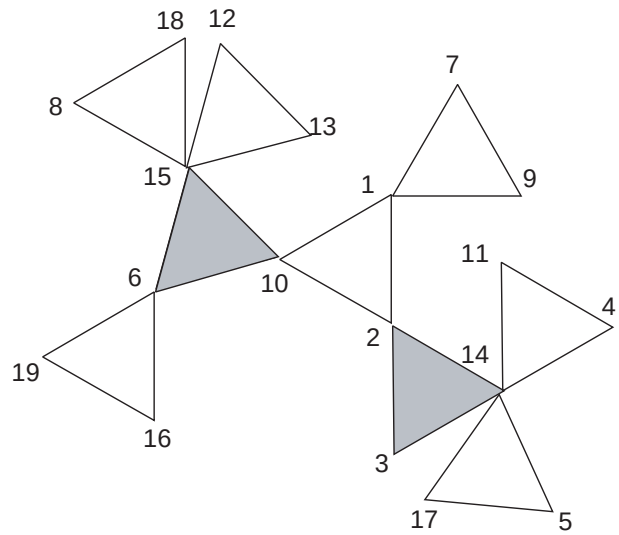


Figure 4.5.4

NPM_{2n} has the homotopy type of a wedge of $(2n-1)!!^2$ spheres of dimension $3n-4$. It was in this work that the increasing triangle trees first arose as the critical elements of a Morse matching on NPM_{2n} (increasing triangle trees are just called trees of triangles in [**122**]). See the chapter by Forman [**74**] in this volume to learn about discrete Morse theory and critical elements. Recall that the property of containing (or not containing) a perfect matching is an example of a monotone graph property (see Section 1.4). When we discuss Alexander duality in the next lecture, we will see how the homology of the poset of graphs that have a monotone graph property is related to the homology of the poset of graphs that don't have the property.

Theorem 4.5.8 and discrete Morse theory play essential roles in the proof of the following equivariant version of the Linusson-Shareshian-Welker result. Indeed, discrete Morse theory is used to show that there is a sign twisted $\mathfrak{S}_{2n-1}$-isomorphism between the homology of NPM_{2n} and the cohomology of another poset called the *factor critical graph poset*. Discrete Morse theory is also used to show that the cohomology of the factor critical graph poset is generated by certain maximal chains naturally indexed by triangle trees. This gives a natural map from the triangle tree generators ρ_T of homology of $\Pi_{2n-1}^{1 \bmod 2}$ to the maximal chains of the factor critical graph poset that are indexed by the triangle trees. Relations on the triangle tree generators of $\tilde{H}_{n-3}(\bar{\Pi}_{2n-1}^{1 \bmod 2})$ are derived, which are shown to map to coboundary relations in the factor critical graph poset.

Theorem 4.5.9 (Shareshian and Wachs [**159**]). *For all $n \geq 1$,*

$$\tilde{H}_{3n-4}(\mathrm{NPM}_{2n}) \downarrow_{\mathfrak{S}_{2n-1}}^{\mathfrak{S}_{2n}} \cong_{\mathfrak{S}_{2n-1}} \tilde{H}_{n-3}(\bar{\Pi}_{2n-1}^{1 \bmod 2}) \otimes \mathrm{sgn}_{2n-1}.$$

Another graph poset, recently studied by Jonsson [**105**], curiously has the same bottom nonvanishing homology as NPM_{2n} and $\bar{\Pi}_{2n-1}^{1 \bmod 2}$ (up to sign twists). A graph is said to be *d-edge-connected* if removal of any set of at most $d-1$ edges leaves

the graph connected. So a 1-edge-connected graph is just a connected graph. The d-clique trees are examples of graphs that are $(d-1)$-edge-connected but not d-edge-connected. In fact, the d-clique trees are minimal elements of the poset of $(d-1)$-edge-connected graphs. Jonsson studied the integral homology of the poset of graphs that are not 2-edge-connected. Let NEC_n^d be the poset of nonempty graphs on node set $[n]$ that are not d-edge-connected. Jonsson discovered that NEC_{2n+1}^2 has homology in multiple dimensions and that the bottom nonvanishing reduced integral homology is in dimension $3n-2$ and is of rank $(2n-1)!!^2$. His proof involves discrete Morse theory with the increasing triangle trees as the critical elements. Shareshian and Wachs [**159**] obtain the following equivariant version of Jonsson's result by giving presentations of the two homology groups in terms of triangle-trees.

Theorem 4.5.10 (Shareshian and Wachs [**159**]). *For all n,*

$$\tilde{H}_{3n-2}(\mathrm{NEC}_{2n+1}^2) \cong_{\mathfrak{S}_{2n+1}} \tilde{H}_{n-2}(\bar{\Pi}_{2n+1}^{1 \bmod 2}) \otimes \mathrm{sgn}_{2n+1}.$$

Conjecture 4.5.11 (Shareshian and Wachs [**159**]). *For all $n, d \geq 1$,*

$$\tilde{H}_{\binom{d+1}{2}n-2}(\mathrm{NEC}_{dn+1}^d) \cong_{\mathfrak{S}_{dn+1}} \tilde{H}_{n-2}(\bar{\Pi}_{dn+1}^{1 \bmod d}) \otimes \mathrm{sgn}_{dn+1}^{\otimes d+1},$$

and

$$\tilde{H}_i(\mathrm{NEC}_{dn+1}^d) = 0$$

if $i < \binom{d+1}{2}n - 2$.

The conjecture is true for $d = 1, 2$. Indeed, Jonsson's homology result and Theorem 4.5.10 comprise the $d = 2$ case. The conjecture for $d = 1$ says that that the poset of disconnected graphs on node set $[n+1]$ has homology isomorphic, as an $\mathfrak{S}_{n+1}$-module, to that of the partition lattice. A proof of this well-known result is discussed in Example 5.2.8.

There is another interesting poset with the same homotopy type and $\mathfrak{S}_{nd+1}$-equivariant homology as that of $\bar{\Pi}_{nd+1}^{1 \bmod d}$, worth mentioning here. This poset is the proper part of the poset $\mathcal{T}_{nd+2}^{2 \bmod d}$ of homeomorphically irreducible trees on leaf set $[nd+2]$ in which each internal node has degree congruent to $2 \bmod d$. By *homeomorphically irreducible* we mean nonrooted and no node has degree 2. The order relation is as follows: $T_1 < T_2$ if T_1 can be obtained from T_2 by contracting internal edges. So the bottom element of $\mathcal{T}_{nd+2}^{2 \bmod d}$ is the star tree (the tree with only one internal node), and the maximal elements are trees in which each internal node has degree exactly $d+2$. The $d = 1$ case of the tree poset $\mathcal{T}_{nd+2}^{2 \bmod d}$ has arisen in various areas such as algebraic geometry, homotopy theory, geometric group theory, mathematical physics and mathematical biology; see eg., [**46, 197, 148, 211, 147, 1, 18, 138**] and the references contained therein. Vogtmann [**197**] showed that the tree poset in the $d = 1$ case is homotopy Cohen-Macaulay. There is a natural action of $\mathfrak{S}_{nd+2}$ on $\mathcal{T}_{nd+2}^{2 \bmod d}$ whose representation on the homology of $\mathcal{T}_{nd+2}^{2 \bmod d} - \{\hat{0}\}$ was computed by Robinson and Whitehouse [**148**] in the $d = 1$ case. Hanlon [**87**] introduced the general tree poset $\mathcal{T}_{nd+2}^{2 \bmod d}$, proved that it is Cohen-Macaulay, and generalized the Robinson-Whitehouse result.

Theorem 4.5.12 (Robinson and Whitehouse, $d = 1$ [**148**], Hanlon [**87**]). *For all d, $n \geq 1$,*

$$\tilde{H}_{n-2}(\bar{\mathcal{T}}_{nd+2}^{2 \bmod d}) \quad \cong_{\mathfrak{S}_{nd+2}} \quad \tilde{H}_{n-2}(\bar{\Pi}_{nd+1}^{1 \bmod d}) \uparrow_{\mathfrak{S}_{nd+1}}^{\mathfrak{S}_{nd+2}} - \tilde{H}_{n-1}(\bar{\Pi}_{nd+2}^{1 \bmod d}).$$

The $d = 1$ case of the $\mathfrak{S}_{nd+2}$-module given in Theorem 4.5.12 has come to be known as the Whitehouse module. It has occurred in a variety of diverse contexts such as homotopy theory [**148, 211**], cyclic Lie operads [**79, 111**], homology of partition posets [**181, 182**], knot theory and graph complexes [**8**], and hyperplane arrangements and Lie algebra homology [**90**].

From Theorem 4.5.12, one can show that (see [**183**]),

$$\tilde{H}_{n-2}(\bar{\mathcal{T}}^{2 \bmod d}_{nd+2}) \downarrow^{\mathfrak{S}_{nd+2}}_{\mathfrak{S}_{nd+1}} \cong_{\mathfrak{S}_{nd+1}} \tilde{H}_{n-2}(\bar{\Pi}^{1 \bmod d}_{nd+1}). \tag{4.5.1}$$

A direct combinatorial proof of this is discussed in the following exercise.

Exercise 4.5.13. By removing the leaf $nd + 2$ from any tree T in $\mathcal{T}^{2 \bmod d}_{nd+2}$ and designating the internal node that had been adjacent to the leaf $nd+2$ as the root, one turns T into a rooted nonplanar tree on leaf set $[nd+1]$ in which each internal node has $id + 1$ children, for some $i \geq 1$. The maximal elements of $\mathcal{T}^{2 \bmod d}_{nd+2}$ are now rooted nonplanar $(d + 1)$-ary trees on leaf set $[nd + 1]$.

(a) Show that any linear extension of the following partial ordering of rooted nonplanar $(d + 1)$-ary trees on leaf set $[nd + 1]$ is a recursive coatom ordering of $\mathcal{T}^{2 \bmod d}_{nd+2} \cup \{\hat{1}\}$. Let $T_1 \wedge T_2 \wedge \cdots \wedge T_{d+1}$ denote the $(d+1)$-tree in which $T_1, T_2, \ldots, T_{d+1}$ are the subtrees of the root. The partial order is defined to be the transitive closure of the relation given by $T < T'$ if T' can be obtained from T by replacing some subtree

$$(T_1 \wedge \cdots \wedge T_{d+1}) \wedge T_{d+2} \wedge \cdots \wedge T_{2d+1}$$

with the subtree

$$(T_1 \wedge \cdots \wedge T_d \wedge T_{d+2}) \wedge T_{d+1} \wedge T_{d+3} \wedge T_{2d+1}$$

where the minimum leaf of T_{d+2} is less than the minimum leaf of T_{d+1}.

(b) The poset $\mathcal{T}^{2 \bmod d}_{nd+2}$ can be viewed as the face poset of a simplicial complex whose facets correspond to rooted nonplanar $(d + 1)$-ary trees on leaf set $[nd + 1]$. Hence the recursive coatom ordering of (a) is simply a shelling of the simplicial complex. Show that the homology facets of the shelling correspond to the d-brushes. Consequently,

$$\bar{\mathcal{T}}^{2 \bmod d}_{nd+2} \simeq \bar{\Pi}^{1 \bmod d}_{nd+1}. \tag{4.5.2}$$

(c) Prove (4.5.1) by first observing that the set of the rooted planar $(d + 1)$-ary trees on leaf set $[nd + 1]$ indexes respective sets of maximal chains of $\bar{\Pi}^{1 \bmod d}_{nd+1}$ and $\bar{\mathcal{T}}^{2 \bmod d}_{nd+2}$ that generate top cohomology, and then showing that the two cohomology groups have the same presentation in terms of rooted planar $(d + 1)$-ary trees on leaf set $[nd + 1]$.

Trappmann and Ziegler [**190**] established shellability of $\mathcal{T}^{2 \bmod d}_{nd+2}$ in a different way from Exercise 4.5.13 (a), but with the same homology facets as in Exercise 4.5.13 (b). Ardila and Klivans [**1**] and Robinson [**147**] have recently independently proved a result stronger than the homotopy result (4.5.2) and the homology result (4.5.1) in the $d = 1$ case; namely that $\bar{\mathcal{T}}^{2 \bmod 1}_{n+1}$ and $\bar{\Pi}_n$ are $\mathfrak{S}_n$-homeomorphic. Robinson's homeomorphism for the $d = 1$ case restricts to a homeomorphism for the general case. Hence $\bar{\mathcal{T}}^{2 \bmod d}_{nd+2}$ and $\bar{\Pi}^{1 \bmod d}_{nd+1}$ are $\mathfrak{S}_{nd+1}$-homeomorphic for all $n, d \geq 1$.

Problem 4.5.14. It is well-known that for each permutation in $\mathfrak{S}_n$, there is a distinct copy of the face poset of the $(n-1)$-dimensional associahedron embedded in $\mathcal{T}_{n+2}^{2 \bmod 1}$ and that by taking fundamental cycles, one obtains a basis for the homology of $\bar{\mathcal{T}}_{n+2}^{2 \bmod 1}$. The associahedron is discussed in the chapter of Fomin and Reading in this volumn [**71**]. Is there a nice basis for homology of $\bar{\mathcal{T}}_{nd+2}^{2 \bmod d}$ for general d, consisting of fundamental cycles of embedded face posets of polytopes, which are copies of some sort of d-analog of the associahedron?

By Theorem 4.5.9 and the isomorphism (4.5.1), the restriction of the $\mathfrak{S}_{2n}$-equivariant homology of NPM_{2n} and $\bar{\mathcal{T}}_{2n}^{2 \bmod 2}$ to $\mathfrak{S}_{2n-1}$ are isomorphic (up to tensoring with the sign representation). For the unrestricted homology modules, we have the following conjecture.

Conjecture 4.5.15 (Linusson, Shareshian and Welker [**122**])**.** *For all $n \geq 1$,*

$$\tilde{H}_{3n-4}(\mathrm{NPM}_{2n}) \cong_{\mathfrak{S}_{2n}} \tilde{H}_{n-4}(\bar{\mathcal{T}}_{2n}^{2 \bmod 2}) \otimes \mathrm{sgn}_{2n}.$$

4.6. Fixed Point Möbius Invariant

We conclude this lecture with a very brief discussion of another commonly used recursive technique for computing homology representations. Let P be a G-poset and for each $g \in G$, let P^g denote the induced subposet of P consisting of elements that are fixed by g. Let χ^P be the character of the virtual representation of G on $\oplus_i (-1)^i \tilde{H}_i(P)$. The following consequence of the Hopf-Lefschetz fixed point theorem first appeared in the combinatorics literature in the work of Baclawski and Björner [**13**].

Theorem 4.6.1. *For any G-poset P and $g \in G$,*

$$\chi^P(g) = \mu(\hat{P}^g).$$

There are many interesting applications of this theorem in the literature. We give a simple example here.

Example 4.6.2 (Stanley, see [**25**])**.** Consider the normal word poset $\mathcal{N}_{n,k}$ under the action of $\mathfrak{S}_n$. Since $P := \bar{\mathcal{N}}_{n,k}$ is Cohen-Macaulay, $(-1)^{k-1}\chi^P$ is the character of the representation of $\mathfrak{S}_n$ on $\tilde{H}_{k-1}(\bar{\mathcal{N}}_{n,k})$. The words that are fixed by g are the words whose letters are fixed points of g. So P^g is the poset of normal nonempty words of length at most k on alphabet $F(g)$, where $F(g)$ is the set of fixed points of g. By Exercise 4.3.2 (c), $\mu(\hat{P}^g) = (-1)^{k-1}(|F(g)|-1)^k$. Since the irreducible representation $S^{(n-1,1)}$ has the character $(|F(g)|-1)$, its kth tensor power has the character $(|F(g)|-1)^k$. By Theorem 4.6.1,

$$\tilde{H}_{k-1}(\bar{\mathcal{N}}_{n,k}) \cong_{\mathfrak{S}_n} (S^{(n-1,1)})^{\otimes k}.$$

Since the identical argument works for $\mathcal{W}_{n,k}$,

$$\tilde{H}_{k-1}(\bar{\mathcal{W}}_{n,k}) \cong_{\mathfrak{S}_n} (S^{(n-1,1)})^{\otimes k}.$$

Exercise 4.6.3. Show that if χ is the character of the representation of $\mathfrak{S}_n$ on the homology of the injective word poset $\bar{\mathcal{I}}_{n,n}$ then

$$\chi(g) = d_{|F(g)|},$$

where d_j is the number of derangements of j letters.

LECTURE 5
Poset Operations and Maps

5.1. Operations: Alexander Duality and Direct Product

In this section we consider some fundamental operations on posets and their affect on homology. The operations are Alexander duality, join and direct product.

Theorem 5.1.1 (Poset Alexander duality, Stanley [**169, 171**])**.** *Let P be a G-poset whose order complex triangulates an n-sphere. If Q is any induced G-subposet of P then for all i,*

$$\tilde{H}_i(Q) \cong_G \tilde{H}^{n-i-1}(P-Q) \otimes \tilde{H}_n(P).$$

If Q is any induced subposet of P then for all i,

$$\tilde{H}_i(Q;\mathbb{Z}) \cong \tilde{H}^{n-i-1}(P-Q,\mathbb{Z}).$$

Exercise 5.1.2. Let P be the poset of all graphs on node set $[n]$ ordered by inclusion of edge sets. The symmetric group $\mathfrak{S}_n$ acts on P in the obvious way, making P an $\mathfrak{S}_n$-poset. Let Q be an induced subposet of P invariant under the action of $\mathfrak{S}_n$. Show that for all i,

$$\tilde{H}_i(\bar{Q}) \cong_{\mathfrak{S}_n} \tilde{H}_{\binom{n}{2}-i-3}(\bar{P}-\bar{Q}) \otimes \mathrm{sgn}_n^{\otimes n}.$$

For example, if Q is the poset of connected graphs on node set $[n]$ and R is the poset of disconnected graphs on node set $[n]$ then

$$\tilde{H}_i(\bar{Q}) \cong_{\mathfrak{S}_n} \tilde{H}_{\binom{n}{2}-i-3}(\bar{R}) \otimes \mathrm{sgn}_n^{\otimes n}.$$

Exercise 5.1.3. Given a simplicial complex Δ on vertex set V, its Alexander dual $\Delta^\vee$ is defined to be the simplicial complex on V consisting of complements of nonfaces of Δ, i.e.,

$$\Delta^\vee = \{V - F : F \subseteq V \text{ and } F \notin \Delta\}.$$

How are the homology of Δ and its Alexander dual related?

Let P be a G-poset and let Q be an H-poset. Then the join $P * Q$ and the product $P \times Q$ are $(G \times H)$-posets with respective $(G \times H)$ actions given by

$$(g,h)x = \begin{cases} gx \text{ if } x \in P \\ hx \text{ if } x \in Q \end{cases},$$

and

$$(g,h)(p,q) = (gp, hq).$$

We have the following poset version of the Künneth theorem of algebraic topology.

Theorem 5.1.4. *Let P be a G poset and let Q be an H-poset. Then*

$$\tilde{H}_r(P * Q) \cong_{G\times H} \bigoplus_i \tilde{H}_i(P) \otimes \tilde{H}_{r-i-1}(Q),$$

for all r.

A result of Quillen [**137**] states that if P and Q have bottom elements $\hat{0}_P$ and $\hat{0}_Q$, respectively, then there is a $(G \times H)$-homeomorphism

$$P \times Q \setminus \{(\hat{0}_P, \hat{0}_Q)\} \cong_{G\times H} P \setminus \{\hat{0}_P\} * Q \setminus \{\hat{0}_Q\}.$$

Walker [**206, 207**] proves the similar result that if P and Q have top elements as well as bottom elements then there is a $(G \times H)$-homeomorphism

$$\overline{P \times Q} \cong_{G\times H} \bar{P} * \bar{Q} * A_2,$$

where A_2 is a two element antichain on which the trivial group acts. Theorem 5.1.4 and Quillen's and Walker's results yield,

Theorem 5.1.5. *Let P be a G-poset with a bottom element $\hat{0}_P$ and let Q be an H-poset with a bottom element $\hat{0}_Q$. Then for all r,*

$$\tilde{H}_r(P \times Q \setminus \{(\hat{0}_P, \hat{0}_Q)\} \cong_{G\times H} \bigoplus_i \tilde{H}_i(P \setminus \{\hat{0}_P\}) \otimes \tilde{H}_{r-i-1}(Q \setminus \{\hat{0}_Q\}).$$

If P and Q also have top elements then for all r,

$$\tilde{H}_r(\overline{P \times Q}) \cong_{G\times H} \bigoplus_i \tilde{H}_i(\bar{P}) \otimes \tilde{H}_{r-i-2}(\bar{Q}).$$

Exercise 5.1.6. Let P and Q be posets with bottom elements. Show that $P \times Q$ is Cohen-Macaulay if and only if P and Q are Cohen-Macaulay. This result also holds in the nonpure case but is more difficult to prove; see [**44**].

The products in Theorem 5.1.5 are known as *reduced products*. There is a similar formula for the homology of ordinary direct products which follows from the Künneth theorem and another $(G \times H)$-homeomorphism

$$P \times Q \cong \|P\| \times \|Q\|$$

of Quillen [**137**] and Walker [**207**].

Theorem 5.1.7. *Let P be a G-poset and let Q be an H poset. Then for all r,*

$$H_r(P \times Q) \cong_{G\times H} \bigoplus_i H_i(P) \otimes H_{r-i}(Q).$$

Next we consider the n-fold product $P^{\times n}$ of a G-poset P. The group that acts on $P^{\times n}$ is the wreath product $\mathfrak{S}_n[G]$, defined in Section 2.4. The action on $P^{\times n}$ is given by

$$(g_1, g_2, \ldots, g_n; \sigma)(p_1, p_2, \ldots, p_n) = (g_1 p_{\sigma^{-1}(1)}, g_2 p_{\sigma^{-1}(2)}, \ldots, g_n p_{\sigma^{-1}(n)}).$$

For each pair of positive integers i and m, define the $\mathfrak{S}_m[G]$-module

$$H(P,i,m) := \begin{cases} S^{(m)}[H_{i-1}(P)] & \text{if } i \text{ is odd} \\ \\ S^{(1^m)}[H_{i-1}(P)] & \text{if } i \text{ is even.} \end{cases}$$

For a partition λ with m_i parts equal to i for each i, let

$$\mathfrak{S}_\lambda[G] := \mathop{\times}_{i:m_i>0} \mathfrak{S}_{m_i}[G]$$

and define the $\mathfrak{S}_\lambda[G]$-module

$$H(P,\lambda) := \bigotimes_{i:m_i>0} H(P,i,m_i).$$

We can now state an analog of Theorem 5.1.7.

Theorem 5.1.8 (Sundaram and Welker [**185**]). *Let P be a G-poset. Then for all r,*

$$H_r(P^{\times n}) \cong_{\mathfrak{S}_n[G]} \bigoplus_{\substack{\lambda \vdash r+n \\ l(\lambda)=n}} H(P,\lambda) \uparrow_{\mathfrak{S}_\lambda[G]}^{\mathfrak{S}_n[G]}.$$

For an analog of Theorem 5.1.5, define the $\mathfrak{S}_m[G]$-modules

$$\tilde{H}(P,i,m,1) := \begin{cases} S^{(m)}[\tilde{H}_{i-2}(P)] & \text{if } i \text{ is odd} \\ \\ S^{(1^m)}[\tilde{H}_{i-2}(P)] & \text{if } i \text{ is even} \end{cases}$$

and

$$\tilde{H}(P,i,m,2) := \begin{cases} S^{(1^m)}[\tilde{H}_{i-2}(P)] & \text{if } i \text{ is odd} \\ \\ S^{(m)}[\tilde{H}_{i-2}(P)] & \text{if } i \text{ is even} \end{cases}$$

and the $\mathfrak{S}_\lambda[G]$-modules

$$\tilde{H}(P,\lambda,1) := \bigotimes_{i:m_i>0} \tilde{H}(P,i,m_i,1)$$

and

$$\tilde{H}(P,\lambda,2) := \bigotimes_{i:m_i>0} \tilde{H}(P,i,m_i,2).$$

Theorem 5.1.9 (Sundaram and Welker [**185**]). *Let P be a G-poset with a bottom element $\hat{0}$. Then for all r,*

$$\tilde{H}_r(P^{\times n} - \{(\hat{0},\ldots,\hat{0})\}) \cong_{\mathfrak{S}_n[G]} \bigoplus_{\substack{\lambda \vdash r+n+1 \\ l(\lambda)=n}} \tilde{H}(P-\{\hat{0}\},\lambda,1) \uparrow_{\mathfrak{S}_\lambda[G]}^{\mathfrak{S}_n[G]}.$$

If P also has a top element then for all r,

$$\tilde{H}_r(\overline{P^{\times n}}) \cong_{\mathfrak{S}_n[G]} \bigoplus_{\substack{\lambda \vdash r+2 \\ l(\lambda)=n}} \tilde{H}(\bar{P},\lambda,2) \uparrow_{\mathfrak{S}_\lambda[G]}^{\mathfrak{S}_n[G]}.$$

We remark that we have stated Theorems 5.1.8 and 5.1.9 in a form different from the original given in [**185**] but completely equivalent.

Example 5.1.10 (The Boolean algebra B_n)**.** We apply the second part of Theorem 5.1.9 to $B_n = B_1^{\times n}$. The trivial group $\mathfrak{S}_1$ acts on B_1 and this induces an action of $\mathfrak{S}_n[\mathfrak{S}_1] = \mathfrak{S}_n$ on B_n, which is precisely the action given in (2.3.1). Clearly $\tilde{H}_{i-2}(\bar{B}_1) = 0$ unless $i = 1$, in which case $H_{-1}(\bar{B}_1)$ is the trivial representation of S_1. It follows that

$$\tilde{H}(B_1, 1, m, 2) = S^{(1^m)}[S^{(1)}] = S^{(1^m)},$$

and $\tilde{H}(B_1, i, m, 2) = 0$ for $i \neq 1$. Hence

$$\tilde{H}(B_1, \lambda, 2) = \begin{cases} S^{(1^n)} & \text{if } \lambda = (1^n) \\ 0 & \text{otherwise.} \end{cases}$$

It follows that the only nonzero term in the decomposition given in Theorem 5.1.9 is the term corresponding to $\lambda = (1^n)$ and $r = n-2$. Hence we recover the fact that $\tilde{H}_r(\bar{B}_n)$ is the sign representation for $r = n-2$ and is 0 otherwise.

Exercise 5.1.11. The face lattice C_n of the n-cross-polytope with its top element removed can be expressed as a product,

$$C_n \setminus \hat{1} = (C_1 \setminus \hat{1})^{\times n}.$$

Since $\mathfrak{S}_2$ acts on $C_1 \setminus \{\hat{1}\}$, the product induces an action of the hyperoctahedral group $\mathfrak{S}_n[\mathfrak{S}_2]$ on $C_n \setminus \hat{1}$.

(a) Show that this action is the action given in Example 2.3.3.

(b) Use Theorem 5.1.9 to prove that

$$\tilde{H}_r(\bar{C}_n) \cong_{\mathfrak{S}_n[\mathfrak{S}_2]} \begin{cases} S^{(1^n)}[S^{(1^2)}] & \text{if } r = n-1 \\ 0 & \text{otherwise.} \end{cases}$$

(c) Use Theorem 5.1.9 to obtain a generalization of the formula in (b) for the poset $X_m^{\times n}$, where X_m is the poset consisting of a bottom element $\hat{0}$ and m atoms.

Example 5.1.12 ([**185**])**.** In this example we compute the homology of lower intervals in the partition lattice Π_n. Each lower interval is isomorphic to a product of smaller partition lattices,

$$[\hat{0}, x] \cong \bigtimes_i \Pi_i^{\times m_i},$$

where m_i is equal to the number of blocks of x of size i. The stabilizer $(\mathfrak{S}_n)_x$ of x under the action of $\mathfrak{S}_n$ on Π_n is isomorphic to $\bigtimes_i \mathfrak{S}_{m_i}[\mathfrak{S}_i]$. The $(\mathfrak{S}_n)_x$-poset $[\hat{0}, x]$ is isomorphic to the $\bigtimes_i \mathfrak{S}_{m_i}[\mathfrak{S}_i]$-poset $\bigtimes_i \Pi_i^{\times m_i}$. We use the second parts of Theorems 5.1.5 and 5.1.9 to compute the representation of $\bigtimes_i \mathfrak{S}_{m_i}[\mathfrak{S}_i]$ on the homology of $(\hat{0}, x)$. This yields the following formula of Lehrer and Solomon [**119**] for the only nonvanishing homology,

$$\tilde{H}_{r(x)-2}(\hat{0}, x) \cong_{\bigtimes_i \mathfrak{S}_{m_i}[\mathfrak{S}_i]} S^{(m_1)}[\tilde{H}_{-2}(\bar{\Pi}_1)] \otimes S^{(1^{m_2})}[\tilde{H}_{-1}(\bar{\Pi}_2)] \otimes S^{(m_3)}[\tilde{H}_0(\bar{\Pi}_3)] \otimes \cdots .$$

By summing over all set partitions of rank r and taking Frobenius characteristic one gets,

$$\mathrm{ch} \bigoplus_{\substack{x \in \Pi_n \\ r(x) = r}} \tilde{H}_{r-2}(\hat{0}, x) = \text{degree } n \text{ term in } (-1)^r h_{n-r} \Big[\sum_{m \geq 1} (-1)^{m-1} \mathrm{ch} \tilde{H}_{m-3}(\bar{\Pi}_m) \Big].$$

Exercise 5.1.13 ([**177, 178**]). Use Theorems 5.1.5 and 5.1.9 to compute the representation of $\mathsf{X}_i\, \mathfrak{S}_{m_i}[\mathfrak{S}_i]$ on $H_i(\hat{0}, x)$, where m_i is equal to the number of blocks of x of size i and $(\hat{0}, x)$ is an interval in the

(a) d-divisible partition lattice $\Pi_{nd}^{0 \bmod d}$,
(b) 1 mod d partition poset $\Pi_{nd+1}^{1 \bmod d}$.

Computations such as those of Example 5.1.12 and Exercise 5.1.13 are used in applications of equivariant versions of the Orlik-Solomon formula and the Goresky-MacPherson formula. This is discussed further in Section 5.4.

The examples above don't adequately demonstrate the power of the product theorems because in these examples homology occurs in only one dimension. A demonstration of the full power can be found in [**178**], [**184**] and [**201**], where representations on the homology of intervals of the k-equal partition lattice $\Pi_{n,k}$, the at least k partition lattice $\Pi_n^{\geq k}$, the general j mod d partition poset $\Pi_{nd+j}^{j \bmod d}$, and the other restricted block size partition posets of Section 4.4, are computed.

5.2. Quillen Fiber Lemma

In a seminal 1978 paper, Quillen introduced several poset fiber theorems. In this section we shall present the most basic of these, known as the "the Quillen fiber lemma", which has proved to be one of the most useful tools in poset topology.

Given two posets P and Q, a map $f : P \to Q$ is called a *poset map* if it is order preserving, i.e., $x \leq_P y$ implies $f(x) \leq_Q f(y)$. Given two G-posets P and Q, a poset map $f : P \to Q$ is called a *G-poset map* if $gf(x) = f(gx)$ for all $x \in P$. Given two simplicial complexes Δ and Γ, a *simplicial map* is a map $f : \Delta \to \Gamma$ that takes 0-faces of Δ to 0-faces of Γ and preserves inclusions, i.e., f is a poset map from $P(\Delta)$ to $P(\Gamma)$. If Δ and Γ are G-simplicial complexes and f commutes with the G-action then f is said to be a G-simplicial map. A G-continuous map $f : X \to Y$ from G-space X to G-space Y is a continuous map that commutes with the G-action. Clearly, a G-poset map $f : P \to Q$ induces a G-simplicial map $f : \Delta(P) \to \Delta(Q)$, which in turn induces a G-continuous map $f : \|\Delta(P)\| \to \|\Delta(Q)\|$.

For any element x of a poset P, let

$$P_{>x} := \{y \in P : y > x\} \quad \text{and} \quad P_{\geq x} := \{y \in P : y \geq x\}.$$

The subsets $P_{<x}$ and $P_{\leq x}$ are defined similarly.

The basic version of the Quillen fiber lemma pertains to homotopy type. Quillen also gave an integral homology version and a equivariant homology version. There is also an equivariant homotopy version due to Thévenaz and Webb [**189**]. We present only the homotopy version and the equivariant homology version. Recall that $\simeq$ denotes homotopy equivalence.

Theorem 5.2.1 (Quillen fiber lemma [**137**]). *Let $f : P \to Q$ be a poset map. If the fiber $f^{-1}(Q_{\leq y})$ is contractible for all $y \in Q$ then*

$$P \simeq Q.$$

Theorem 5.2.2 (Equivariant homology version [**137**]). *Let $f : P \to Q$ be a G-poset map. If the fiber $f^{-1}(Q_{\leq y})$ is acyclic (over $\mathbb{C}$) for all $y \in Q$ then*

$$\tilde{H}_r(P) \cong_G \tilde{H}_r(Q)$$

for all r.

Let's look at Quillen's original example [**137**]. For a finite group G, let $\mathcal{S}_p(G)$ be the poset of nontrivial p-subgroups of G ordered by inclusion and let $\mathcal{A}_p(G)$ be the induced subposet of nontrivial elementary abelian p-subgroups of G. The poset $\mathcal{S}_p(G)$ and its order complex were studied by Brown in his work on group cohomology. Quillen proposed the smaller poset $\mathcal{A}_p(G)$ as a way of studying the homotopy invariants of the Brown complex. He used the Quillen fiber lemma to establish homotopy equivalence between the Brown complex and the Quillen complex. Both posets are G-posets, where G acts by conjugation.

Theorem 5.2.3 (Quillen [**137**]). *For any finite group G and prime p,*

$$\mathcal{A}_p(G) \simeq \mathcal{S}_p(G), \tag{5.2.1}$$

and

$$\tilde{H}_j(\mathcal{A}_p(G)) \cong_G \tilde{H}_j(\mathcal{S}_p(G)) \quad \forall j. \tag{5.2.2}$$

Proof. The inclusion map

$$i : \mathcal{A}_p(G) \to \mathcal{S}_p(G)$$

is a G-poset map. Let us show that the fiber $i^{-1}(S_p(G)_{\leq H})$ is contractible for all $H \in \mathcal{S}_p(G)$ so that we can apply the Quillen fiber lemma. Note that

$$i^{-1}(S_p(G)_{\leq H}) = \mathcal{A}_p(H).$$

Since H is a nontrivial p-group, it has a nontrivial center. Let B be the subgroup of the center consisting of all elements of order 1 or p. Consider the poset map $f : \mathcal{A}_p(H) \to \{BA : A \in \mathcal{A}_p(H)\}$ defined by $f(A) = BA$. The fibers of this map are contractible since they all have maximum elements. So by the Quillen fiber lemma

$$\mathcal{A}_p(H) \simeq \{BA : A \in \mathcal{A}_p(H)\}.$$

Since $\{BA : A \in \mathcal{A}_p(H)\}$ has minimum element B, it is contractible. So $\mathcal{A}_p(H)$ and thus $i^{-1}(S_p(G)_{\leq H})$ is contractible. Hence by the Quillen fiber lemma, (5.2.1) holds and by the equivariant homology version, (5.2.2) holds. □

For certain finite groups G, the homology of the Quillen complex $\Delta(\mathcal{A}_p(G))$ is well-understood; namely for groups of Lie type. Quillen [**137**] shows that for groups G of Lie type in characteristic p, the Quillen complex $\Delta(\mathcal{A}_p(G))$ is homotopic to a simplicial complex known as the building for G, which has the homotopy type of a wedge of spheres of a single dimension. Webb [**209**] shows that the unique nonvanishing G-equivariant homology of $\mathcal{A}_p(G)$ is the same as that of the building, which is known as the Steinberg representation. For general finite groups, the Quillen complex is not nearly as well-behaved, nor well-understood. In fact, it was only recently shown by Shareshian [**156**] that for certain primes the integral homology of the Quillen complex of the symmetric group has torsion.

The following long-standing conjecture of Quillen imparts a sense of the significance of poset topology in group theory.

Conjecture 5.2.4 (Quillen Conjecture). *For any finite group G and prime p, the poset $\mathcal{A}_p(G)$ is contractible if and only if G has a nontrivial normal p-subgroup.*

The necessity of contractibility was proved by Quillen; the sufficiency is still open. However significant progress has been made by Aschbacher and Smith [**4**].

In order to gain understanding of the Quillen complex for the symmetric group, Bouc [**47**] considered the induced subposet T_n of $\mathcal{S}_2(\mathfrak{S}_n)$ consisting of nontrivial

2-subgroups of $\mathfrak{S}_n$ that contain a transposition, and the induced subposet T'_n of $\mathcal{A}_2(\mathfrak{S}_n)$ consisting of nontrivial elementary abelian 2-subgroups generated by transpositions. It is easy to see that T'_n is $\mathfrak{S}_n$-homeomorphic to the matching complex M_n discussed in Section 2.4.3. Bouc used the Quillen fiber lemma to prove

$$T_n \simeq T'_n \simeq M_n \tag{5.2.3}$$

and for all i,

$$\tilde{H}_i(T_n) \cong_{\mathfrak{S}_n} \tilde{H}_i(T'_n) \cong_{\mathfrak{S}_n} \tilde{H}_i(M_n). \tag{5.2.4}$$

In addition to computing the representation of the symmetric group on the homology of the matching complex, Bouc discovered torsion in the integral homology of the matching complex. See [**157**], [**202**] and [**105**] for further results on torsion in the matching complex.

Exercise 5.2.5. Prove (5.2.3) and (5.2.4).

The usefulness of the matching complex in understanding the topology of the Quillen and Brown complexes was recently demonstrated by Ksontini [**116, 117**]. He used simple connectivity of M_n for $n \geq 8$, which was proved by Bouc, to establish simple connectivity of $\mathcal{S}_2(\mathfrak{S}_n)$ for $n \geq 8$. It was also shown by Ksontini [**116, 117, 118**], Shareshian [**156**], and Shareshian and Wachs [**158**] that a hypergraph version of the matching complex is useful in studying the topology of $\Delta(\mathcal{S}_p(\mathfrak{S}_n))$ when $p \geq 3$.

Exercise 5.2.6. Let $\bar{x}$ be a closure operator on P, i.e. $x \leq \bar{x}$ and $\bar{\bar{x}} = \bar{x}$ for all $x \in P$. Define $\mathrm{cl}(P) := \{x \in P : \bar{x} = x\}$.

(a) Use the Quillen fiber lemma to prove:

$$P \simeq \mathrm{cl}(P).$$

(b) Use (a) to show that

$$\bar{\mathcal{W}}(n,k) \simeq \bar{\mathcal{N}}(n,k).$$

(We already know this and (c) from Exercise 4.3.2.)

(c) Derive an equivariant homology version of the homotopy result in (a) and use it to prove

$$\tilde{H}_i(\bar{\mathcal{W}}(n,k)) \cong_{\mathfrak{S}_n} \tilde{H}_i(\bar{\mathcal{N}}(n,k)) \quad \forall i.$$

Exercise 5.2.7 (Homology version of Rota's crosscut theorem)**.** Let L be a lattice and let M be the subposet of $\bar{L}$ consisting of non-$\hat{0}$ meets of coatoms. Prove:

(a) $\bar{L} \simeq M$.
(b) If G is a group acting on L then for all i, $\tilde{H}_i(\bar{L}) \cong_G \tilde{H}_i(M)$.
(c) Let $\Gamma(L)$ be the simplicial complex whose vertex set is the set of coatoms of L and whose faces are sets of coatoms whose meet is not $\hat{0}$ (this is known as the cross-cut complex of L). Show $\bar{L} \simeq \Gamma(L)$ and if G is a group acting on L then for all i, $\tilde{H}_i(\bar{L}) \cong_G \tilde{H}_i(\Gamma(L))$.

The next two examples are connected with the combinatorics of knot spaces and arose in the work of Vassiliev [**193, 194, 195**].

Example 5.2.8. Let NCG_n be the poset of disconnected graphs on node set $[n]$ ordered by inclusion of edge sets. Let

$$f : \overline{\mathrm{NCG}}_n \to \bar{\Pi}_n$$

be the poset map such that $f(G)$ is the partition of $[n]$ whose blocks are the node sets of the connected components of G. The fibers of f are given by

$$f^{-1}((\hat{0}, \pi]) = (\hat{0}, G_\pi],$$

where G_π is the graph whose connected components are cliques on the blocks of π. Since the fibers have maximal elements they are contractible. Hence by the Quillen fiber lemma

$$\overline{\mathrm{NCG}}_n \simeq \overline{\Pi}_n,$$

which implies that $\overline{\mathrm{NCG}}_n$ has the homotopy type of a wedge of $(n-1)!$ spheres of dimension $n-3$. By the equivariant homology version of the Quillen fiber lemma the only nonvanishing homology of $\overline{\mathrm{NCG}}_n$ is given by

$$\tilde{H}_{n-3}(\overline{\mathrm{NCG}}_n) \cong_{\mathfrak{S}_n} \tilde{H}_{n-3}(\Pi_n). \tag{5.2.5}$$

Exercise 5.2.9. Let CG_n be the poset of connected graphs on node set $[n]$ ordered by inclusion of edge sets. From Alexander duality and (5.2.5), we have that he only nonvanishing homology of $\overline{\mathrm{CG}}_n$ is given by

$$\tilde{H}_{\binom{n-1}{2}-1}(\overline{\mathrm{CG}}_n) \cong_{\mathfrak{S}_n} \mathrm{sgn}^{\otimes n} \otimes \tilde{H}_{n-3}(\Pi_n).$$

Show that for $n \geq 3$, the poset $\overline{\mathrm{CG}}_n$ has the homotopy type of a wedge of $(n-1)!$ spheres of dimension $\binom{n-1}{2} - 1$.

Example 5.2.10. A graph is said to be *k-connected* if removal of any set of at most $k-1$ vertices leaves the graph connected. Note that, unless $k = 1$, this is a different graph property from that of being k-edge-connected, which was discussed in Section 4.5.2. Every triangle tree is 2-edge-connected, but not 2-connected if the number of nodes is greater than 3. Let NCG_n^k be the poset of graphs on node set $[n]$ that are not k-connected, ordered by inclusion of edge sets. The following result solves a problem of Vassiliev [**195**] which arose from his study of knots.

Theorem 5.2.11 (Babson, Björner, Linusson, Shareshian, Welker [**8**], and Turchin [**192**])**.**

(a) $\overline{NCG}_n^2$ *has the homotopy type of a wedge of* $(n-2)!$ *spheres of dimension* $2n-5$.
(b) $\tilde{H}_{n-3}(\overline{NCG}_n^2) \downarrow_{\mathfrak{S}_{n-1}}^{\mathfrak{S}_n} \cong_{\mathfrak{S}_{n-1}} \tilde{H}_{n-4}(\Pi_{n-1}) \otimes \mathrm{sgn}_{n-1}$
(c) $\tilde{H}_{n-3}(\overline{NCG}_n^2) \cong_{\mathfrak{S}_n} (\tilde{H}_{n-4}(\Pi_{n-1}) \uparrow_{\mathfrak{S}_{n-1}}^{\mathfrak{S}_n} - \tilde{H}_{n-3}(\Pi_n)) \otimes \mathrm{sgn}_n$

Note that the representation on the right side of (c) is the Whitehouse module discussed in Section 4.5.2. We describe the proof method of Babson, Björner, Linusson, Shareshian, and Welker. Define the poset map

$$f : \overline{\mathrm{NCG}}_n^2 \to \overline{B_{n-1} \times \Pi_{n-1}}$$

by letting

$$f(G) = (S, \pi),$$

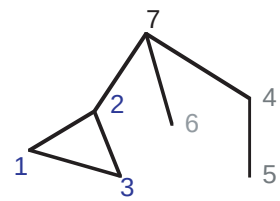


Figure 5.2.1. Not 2-connected graph

where S is the set of nodes joined to node n by an edge, and π is the partition whose blocks correspond to the connected components of $G \setminus \{n\}$. For example, if G is the not 2-connected graph in Figure 5.2.1 then

$$f(G) = (\{2, 4, 6\}, 123/45/6).$$

Some of the fibers have maximum elements; so one sees immediately that they are contractible. For example,

$$f^{-1}((\hat{0}, (\{2, 4, 6\}, 123/45/6)]) = (\hat{0}, G],$$

where G is the graph in Figure 5.2.1. But it is not so easy to see that the fibers without maximum elements are contractible. For example, the fiber $f^{-1}((\hat{0}, (\{1, 2\}, 123456)])$ has no maximum. To show that such fibers are contractible, Babson, Björner, Linusson, Shareshian, and Welker used discrete Morse theory, which had just been introduced by Forman [**72**]. In fact, this was the first of many applications of discrete Morse theory in topological combinatorics.

Now by the Quillen fiber lemma, we have

$$\overline{\mathrm{NCG}}_n^2 \simeq \overline{B_{n-1} \times \Pi_{n-1}}.$$

Since the product of shellable bounded posets is shellable (Theorem 3.1.10), and the proper part of a shellable poset is shellable (Corollary 3.1.9), $\overline{B_{n-1} \times \Pi_{n-1}}$ has the homotopy type of a wedge of $(2n-5)$-spheres. The number of spheres is the absolute value of the Möbius invariant, which by Proposition 1.2.1 and (3.2.1) is $(n-2)!$. Hence Part (a) of the theorem holds.

For Part (b), we use the equivariant homology version of the Quillen fiber lemma. Clearly the map f commutes with the permutations that fix n. Hence,

$$\tilde{H}_{2n-5}(\overline{\mathrm{NCG}}_n^2) \downarrow_{\mathfrak{S}_{n-1}}^{\mathfrak{S}_n} \cong_{\mathfrak{S}_{n-1}} \tilde{H}_{2n-5}(\overline{B_{n-1} \times \Pi_{n-1}}).$$

Theorem 5.1.5 completes the proof of Part (b).

The proof of Part (c) uses Part (b) and a computation of the fixed point Möbius invariant (Section 4.6).

Exercise 5.2.12. Let CG_n^k be the poset of k-connected graphs on node set $[n]$ ordered by inclusion of edge sets.

(a) Use Theorem 5.2.11 to determine the homotopy type of $\overline{\mathrm{CG}}_n^2$.

(b) ([**8**]) Show that $\overline{\mathrm{CG}}_n^{n-2}$ is dual to the face poset of the matching complex M_n. Conclude that neither $\overline{\mathrm{CG}}_n^{n-2}$ nor $\overline{\mathrm{NCG}}_n^{n-2}$ has the homotopy type of a wedge of spheres and that their integral homology has torsion.

Shareshian [**154**] uses discrete Morse theory to determine the homotopy type of $\overline{\mathrm{CG}}_n^2$ directly without resorting to Theorem 5.2.11, and to obtain bases for homology of $\overline{\mathrm{CG}}_n^2$, thereby solving another problem of Vassiliev [**195**]. The only value of k other than $k = 1, 2, n-2$, for which anything significant is known about the topology of the not k-connected graph complex is $k = 3$. Discrete Morse theory is used to obtain the following result.

Theorem 5.2.13 (Jonsson [**104**]). *$\overline{NCG}_n^3$ has the homotopy type of a wedge of $(n-3)(n-2)!/2$ spheres of dimension $2n-4$.*

Problem 5.2.14. What can be said about the representation of $\mathfrak{S}_n$ on the homology of $\overline{\mathrm{NCG}}_n^3$?

Problem 5.2.15. ([**8**]) What can be said about the topology of $\overline{\mathrm{NCG}}_n^k$ for $3 < k < n-2$? When does $\overline{\mathrm{NCG}}_n^k$ fail to be a wedge of spheres? When does its integral homology have torsion? Recall that the integral homology of $\overline{\mathrm{NCG}}_n^{n-2}$ has torsion (Exercise 5.2.12 (b)).

5.3. General Poset Fiber Theorems

In this section we present homotopy and homology versions of a fiber theorem of Björner, Wachs and Welker, which generalizes the Quillen fiber lemma and several other fiber theorems. Recall the definitions of k-connectivity and k-acyclicity, which appear after Theorem 3.1.7.

Theorem 5.3.1 (Björner, Wachs, Welker [**43**]). *Let $f : P \to Q$ be a poset map such that for all $q \in Q$ the fiber $\Delta(f^{-1}(Q_{\leq q}))$ is $l(f^{-1}(Q_{<q}))$-connected. Then*

$$\Delta(P) \simeq \Delta(Q) \ \vee \ \left\{\Delta(f^{-1}(Q_{\leq q})) * \Delta(Q_{>q}) : q \in Q\right\}, \tag{5.3.1}$$

where $*$ denotes join defined in (1.1.1) and $\vee$ denotes the wedge (of $\Delta(Q)$ and all $\Delta(f^{-1}(Q_{\leq q})) * \Delta(Q_{>q})$) formed by identifying the vertex q in $\Delta(Q)$ with any element of $f^{-1}(Q_{\leq q})$, for each $q \in Q$.

For clarity, let us remark that if $\Delta(Q)$ is connected then the space described on the right-hand side of (5.3.1), which has $|Q|$ wedge-points, is homotopy equivalent to a *one-point* wedge, where arbitrarily chosen points of $f^{-1}(Q_{\leq q})$, one for each $q \in Q$, are identified with some (arbitrarily chosen) point of Q. Thus (5.3.1) becomes

$$\Delta(P) \simeq \Delta(Q) \vee \bigvee_{q \in Q} \Delta(f^{-1}(Q_{\leq q})) * \Delta(Q_{>q}). \tag{5.3.2}$$

We will refer to a poset map $f : P \to Q$ such that for all $q \in Q$ the fiber $\Delta(f^{-1}(Q_{\leq q}))$ is $l(f^{-1}(Q_{<q}))$-connected as being *well-connected*. Note that the connectivity condition implies that each fiber $f^{-1}(Q_{\leq q})$ is nonempty.

Example 5.3.2. Let $f : P \to Q$ be the poset map depicted in Figure 5.3.1. For the two maximal elements of Q the fiber $\Delta(f^{-1}(Q_{\leq q}))$ is a 1-sphere. For the bottom element of Q the fiber $\Delta(f^{-1}(Q_{\leq q}))$ is a 0-sphere, and $\Delta(Q_{>q})$ is a 0-sphere too. So in either case $\Delta(f^{-1}(Q_{\leq q})) * \Delta(Q_{>q})$ is homeomorphic to a 1-sphere. Hence the simplicial complex on the right side of (5.3.1) has a 1-sphere attached to each element of Q. Thus Theorem 5.3.1 determines $\Delta(P)$ to have the homotopy type of a wedge of three 1-spheres. One can see this directly by observing that $\Delta(P)$ is homeomorphic to two 1-spheres intersecting in two points.

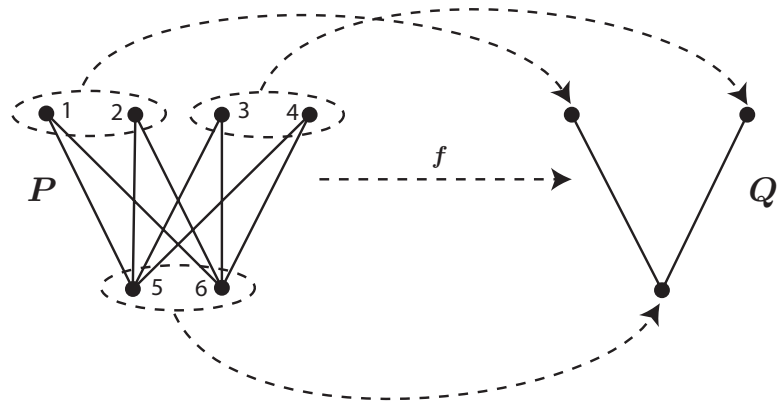


Figure 5.3.1. A well-connected poset map

In [**43**], a version of the general fiber theorem for homology over the integers or over any field is also given, as are equivariant homotopy and homology versions. We state only the equivariant homology version here.

Theorem 5.3.3 (Björner, Wachs, Welker [**43**]). *Let $f : P \to Q$ be a G-poset map. If $f^{-1}(Q_{\leq q})$ is $l(f^{-1}(Q_{<y}))$-acyclic for all $q \in Q$ then*

(5.3.3)
$$\tilde{H}_r(P) \;\cong_G\; \tilde{H}_r(Q) \;\oplus\; \bigoplus_{q\in Q/G} \;\bigoplus_{i=-1}^{r} \Big(\tilde{H}_i(f^{-1}(Q_{\leq q})) \otimes \tilde{H}_{r-i-1}(Q_{>q})\Big)\Big\uparrow_{G_q}^{G}.$$

The general fiber theorems are proved using techniques of [**210, 222**] involving diagrams of spaces and spectral sequences, which were developed to study subspace arrangements. In Section 5.4, we discuss the connection between the general fiber theorems and subspace arrangements.

Example 5.3.4. Theorem 5.3.3 can be applied to the well-connected poset map $f : P \to Q$ given in Example 5.3.2. Let G be the cyclic group $\mathfrak{S}_2$ whose non-identity element acts by $(1,2)(3,4)$ on P and trivially on Q. The map f is clearly a G-poset map. For each $q \in Q$, we have $G_q = \mathfrak{S}_2$. If q is the bottom element of Q then the fiber $f^{-1}(Q_{\leq q})$ is $\mathfrak{S}_2$-homeomorphic to a 0-sphere on which $\mathfrak{S}_2$ acts trivially. The same is true for $Q_{>q}$. It follows that the representation of $\mathfrak{S}_2$ on $\tilde{H}_0(f^{-1}(Q_{\leq q})) \otimes \tilde{H}_0(Q_{>q})$ is the trivial representation $S^{(2)}$. If q is one of the maximal elements of Q then the fiber $f^{-1}(Q_{\leq q})$ is $\mathfrak{S}_2$-homeomorphic to a circle with $(1,2)(3,4)$ acting by reflecting the circle about the line spanned by a pair of antipodal points. Hence the representation of $\mathfrak{S}_2$ on $\tilde{H}_1(f^{-1}(Q_{\leq q}))$ is the sign representation $S^{(1^2)}$. Since $Q_{>q}$ is the empty simplicial complex, the representation of $\mathfrak{S}_2$ on $\tilde{H}_{-1}(Q_{>q})$ is the trivial representation. It follows that the representation of $\mathfrak{S}_2$ on $\tilde{H}_1(f^{-1}(Q_{\leq q})) \otimes \tilde{H}_{-1}(Q_{>q})$ is $S^{(1^2)}$. We conclude from (5.3.3) that the $\mathfrak{S}_2$-module $\tilde{H}_1(P)$ decomposes into $S^{(2)} \oplus S^{(1^2)} \oplus S^{(1^2)}$ and that $\tilde{H}_i(P) = 0$ for $i \neq 1$.

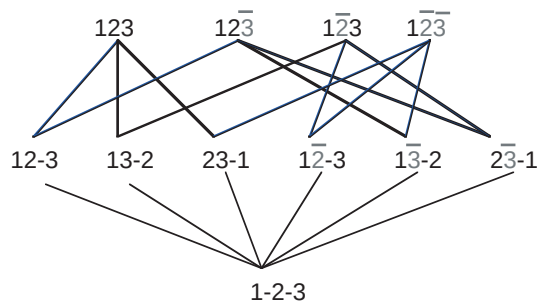


Figure 5.3.2. P_3

Example 5.3.5. Consider the action of the hyperoctahedral group $\mathfrak{S}_n[\mathbb{Z}_2]$ on the type B partition lattice Π_n^B. Define the bar-erasing map

$$f : \bar{\Pi}_n^B \to \bar{\Pi}_{\{0,1,\ldots,n\}} \tag{5.3.4}$$

by letting $f(\pi)$ be the partition obtained by erasing the bars from the barred partition π. For example

$$f(0\,4\,7\,/1\,\bar{5}\,6\,\bar{9}\,/2\,\bar{3}) = 0\,4\,7\,/1\,5\,6\,9\,/2\,3$$

It is clear that f is a $\mathfrak{S}_n[\mathbb{Z}_2]$-poset map. We establish well-connectedness in the following exercise.

Exercise 5.3.6 (Wachs [**204**])**.** Let P_n be the subposet of Π_n^B consisting of barred partitions whose zero-block is a singleton, where the *zero-block* is the block that contains 0. The poset P_3 with the zero-block suppressed is given in Figure 5.3.2. Show the following.

(a) The fibers of the bar-erasing map have the form

$$\Pi_{b_0}^B \times P_{b_1} \times \cdots \times P_{b_k} - \{\hat{0}\},$$

where $b_0 + b_1 + \cdots + b_k = n + 1$.

(b) P_n is a geometric semilattice

(c) The bar-erasing map (5.3.4) is well-connected.

(d) $|\mu(P_n \cup \{\hat{1}\}| = (2(n-1)-1)!!$.

It follows from Exercise 5.3.6 that the general fiber theorems can be used to compute the homotopy type and $\mathfrak{S}_n[\mathbb{Z}_2]$-equivariant homology of Π_n^B. Since $\overline{\Pi_n^B}$ has the homotopy type of a wedge of $(2n-1)!!$ spheres of dimension $n-2$ (see Exercise 3.2.11), the homotopy equivalence (5.3.2) reduces to a combinatorial identity, which turns out to be easy to prove directly and therefore not very interesting. The equivariant homology version of the general fiber theorem does yield an interesting identity. Theorems 5.1.5 and 5.1.9 can be applied to the fibers given in Part (a) of Exercise 5.3.6. This and the homology version of the general fiber theorem are used in [**204**] to prove

$$\tilde{H}_{n-2}(\overline{\Pi_n^B}) \cong_{\mathfrak{S}_n[\mathbb{Z}_2]} \bigoplus_{(\gamma_1,\ldots,\gamma_k):\sum_i \gamma_i = n} \tilde{H}_{\gamma_1-2}(\overline{P_{\gamma_1}}) \bullet \cdots \bullet \tilde{H}_{\gamma_k-2}(\overline{P_{\gamma_k}}), \tag{5.3.5}$$

where $\bullet$ denotes induction product:

$$X \bullet Y := (X \otimes Y) \uparrow_{\mathfrak{S}_n[\mathbb{Z}_2]\times\mathfrak{S}_m[\mathbb{Z}_2]}^{\mathfrak{S}_{n+m}[\mathbb{Z}_2]} .$$

This example is considered in the more general setting of rank-selected Dowling lattices in [**204**]. The homotopy type of the Dowling lattice version of $P_n \setminus \{\hat{0}\}$ was also computed by Hultman [**100**].

Nonpure versions of Cohen-Macaulay fiber theorems of Baclawski [**12**] and of Quillen [**137**] follow from the general fiber theorems. We state only the nonpure version of Baclawski's result.

Theorem 5.3.7 (Björner, Wachs, Welker [**43**]). *Let P and Q be semipure posets and let $f : P \to Q$ be a surjective rank-preserving poset map. Assume that for all $q \in Q$, the fiber $\Delta(f^{-1}(Q_{\leq q}))$ is Cohen-Macaulay over $\mathbf{k}$. If Q is sequentially Cohen-Macaulay over $\mathbf{k}$, then so is P.*

5.4. Fiber Theorems and Subspace Arrangements

Shortly after the Goresky-MacPherson formula (Theorem 1.3.8) appeared, Ziegler and Živaljević obtained a homotopy version, and after that Sundaram and Welker obtained an equivariant homology version. In this section we show that the Ziegler-Živaljević and Sundaram-Welker formulas for linear subspace arrangements can be easily derived from the general fiber theorems of Section 5.3. This connection should come as no surprise, since the method of proof of the general fiber theorems is based on the proofs of Ziegler-Živaljević and Sundaram-Welker; namely diagrams of spaces and spectral sequences.

Let $\mathcal{A}$ be a linear subspace arrangement i.e., a finite collection of linear subspaces in Euclidean space $\mathbb{R}^d$. The *singularity link* $V_{\mathcal{A}}^o$ is defined as

$$V_{\mathcal{A}}^o = \mathbb{S}^{(d-1)} \cap \bigcup_{X\in\mathcal{A}} X,$$

where $\mathbb{S}^{(d-1)}$ is the unit $(d-1)$-sphere in $\mathbb{R}^d$. Recall that $L(\mathcal{A})$ denotes the intersection lattice of $\mathcal{A}$ (ordered by reverse inclusion). Let $\bar{L}(\mathcal{A})$ denote $L(\mathcal{A}) \setminus \{\hat{0}, \hat{1}\}$ if $\mathcal{A}$ is essential (i.e., $\cap\mathcal{A} = \{0\}$), and $L(\mathcal{A}) \setminus \{\hat{0}\}$ otherwise.

Theorem 5.4.1 (Ziegler & Živaljević [**222**]). *Let $\mathcal{A}$ be a linear subspace arrangement. Then*

$$V_{\mathcal{A}}^o \;\simeq\; \Delta(\bar{L}(\mathcal{A})) \;\vee\; \left\{\mathrm{susp}^{\dim x}(\Delta(\hat{0}, x)) : x \in \bar{L}(\mathcal{A})\right\}, \tag{5.4.1}$$

where the wedge is formed by identifying each vertex x in $\Delta(\bar{L}(\mathcal{A}))$ with any point in $\mathrm{susp}^{\dim x}(\Delta(\hat{0}, x))$. Consequently, for each i,

$$\tilde{H}_i(V_{\mathcal{A}}^o; \mathbb{Z}) \cong \bigoplus_{x\in L(\mathcal{A})\setminus\{\hat{0}\}} \tilde{H}_{i-\dim x}((\hat{0}, x); \mathbb{Z}). \tag{5.4.2}$$

Proof. (given in [**43**]). Suppose $\mathcal{A} = \{X_1, \ldots, X_n\}$. Let $\mathcal{H}$ be an essential hyperplane arrangement in $\mathbb{R}^d$ such that each X_i is the intersection of a collection of hyperplanes in $\mathcal{H}$. The hyperplane arrangement $\mathcal{H}$ partitions $\mathbb{S}^{(d-1)}$ into open cells of each dimension from 0 to $d-1$. Let $\mathcal{F}(\mathcal{H})$ be the poset of closures of the nonempty cells ordered by inclusion. Clearly $\Delta(\mathcal{F}(\mathcal{H}))$ is a triangulation of $\mathbb{S}^{(d-1)}$. For $X \in \bar{L}(\mathcal{A})$, let $\mathcal{F}_{\mathcal{H}}(X)$ be the order ideal of $\mathcal{F}(\mathcal{H})$ consisting of closed cells contained in X, and let $\mathcal{F}_{\mathcal{H}}(\mathcal{A})$ be the order ideal of $\mathcal{F}(\mathcal{H})$ consisting of closed cells

contained in $V_{\mathcal{A}}^o$. Note that $\Delta(\mathcal{F}_{\mathcal{H}}(X))$ is a triangulation of the $(\dim X - 1)$-sphere $X \cap \mathbb{S}^{(d-1)}$ and $\Delta(\mathcal{F}_{\mathcal{H}}(\mathcal{A}))$ is a triangulation of $V_{\mathcal{A}}^o$.

Now let

$$f : \mathcal{F}_{\mathcal{H}}(\mathcal{A}) \to \bar{L}(\mathcal{A})^*$$

be defined by

$$f(\tau) = \bigcap_{i : X_i \supseteq \tau} X_i.$$

Clearly f is order preserving. We claim that f is well-connected. Observe that for all $X \in \bar{L}(\mathcal{A})$,

$$f^{-1}((\bar{L}(\mathcal{A})^*)_{\leq X}) = \mathcal{F}_{\mathcal{H}}(X).$$

So $f^{-1}((\bar{L}(\mathcal{A})^*)_{\leq X})$ is a $(\dim X - 1)$-sphere, which is $(\dim X - 2)$-connected. Since $f^{-1}((\bar{L}(\mathcal{A})^*)_{< X})$ has length at most $\dim X - 2$, f is indeed well-connected, and thus (5.4.1) follows from Theorem 5.3.1. □

The reason we refer to Theorem 5.4.1 as a homotopy version of the Goresky-MacPherson formula is that the Goresky-MacPherson formula (for linear subspace arrangements) follows from (5.4.2) by Alexander duality. Indeed, one can simply apply Theorem 5.1.1 to the subposet $\mathcal{F}_{\mathcal{H}}(\mathcal{A})$ of $\mathcal{F}(\mathcal{H})$. A homotopy version of the Goresky-MacPherson formula for general affine subspace arrangements is also given in [**222**].

Now let G be a finite subgroup of the orthogonal group O_d that maps subspaces in $\mathcal{A}$ to subspaces in $\mathcal{A}$. We say that $\mathcal{A}$ is a G-arrangement. Clearly G acts as a group of poset maps on $L(\mathcal{A})$ and as a group of homomorphisms on $\tilde{H}_i(V_{\mathcal{A}}^o)$ and on $\tilde{H}_i(M_{\mathcal{A}})$. For each $x \in L(\mathcal{A})$, let S_x be the $(\dim x - 1)$-sphere $x \cap \mathbb{S}^{(d-1)}$. Each g in the stabilizer G_x acts as an orientation preserving or reversing homeomorphism on S_x. Define

$$\mathrm{sgn}_x(g) := \begin{cases} 1 & \text{if } g \text{ is orientation preserving} \\ -1 & \text{if } g \text{ is orientation reversing.} \end{cases}$$

By viewing g as an element of $GL(x)$, we have $\mathrm{sgn}_x(g) = \det(g)$. Note that $\tilde{H}_{\dim x - 1}(S_x)$ is a one dimensional representation of G_x whose character is given by sgn_x.

Theorem 5.4.2 (Sundaram and Welker [**186**])**.** *Let $\mathcal{A}$ be a G-arrangement of linear subspaces in $\mathbb{R}^d$. Then for all i,*

$$\tilde{H}_i(V_{\mathcal{A}}^o) \cong_G \bigoplus_{x \in L(\mathcal{A}) \setminus \{\hat{0}\}/G} (\tilde{H}_{i - \dim x}(\hat{0}, x) \otimes \mathrm{sgn}_x) \uparrow_{G_x}^G . \tag{5.4.3}$$

Consequently,

$$\tilde{H}^i(M_{\mathcal{A}}) \cong_G \bigoplus_{x \in L(\mathcal{A}) \setminus \{\hat{0}\}/G} (\tilde{H}_{d-2-i-\dim x}(\hat{0}, x) \otimes \mathrm{sgn}_x \otimes \mathrm{sgn}_{\hat{0}}) \uparrow_{G_x}^G . \tag{5.4.4}$$

Exercise 5.4.3. Use Theorems 5.3.3 and 5.1.1 to prove Theorem 5.4.2.

In [**186**] Theorem 5.4.2 is applied to subspace arrangements whose intersection lattices are the k-equal partition lattice discussed in Section 3.2.4 and the d-divisible partition lattice discussed in Exercises 4.3.6 and 4.4.8. By computing the multiplicity of the trivial representation in $\tilde{H}^i(M_{\mathcal{A}})$, Sundaram and Welker obtain the Betti number formula of Arnol'd given in (3.2.3) and an analogous formula for the space

of monic polynomials of degree n, with at least one root multiplicity not divisible by d. For another recent approach to studying these Betti numbers, see [**113**].

In [**201**], Theorem 5.4.2 is applied to the subspace arrangement whose intersection lattice is the restricted block size partition lattice $\Pi_n^S \cup \{\hat{0}\}$ of Theorem 4.4.11. Since each interval $(\hat{0}, x)$ of Π_n^S is the product of smaller Π_m^S, one uses Theorem 4.4.11, and the product formulas of Section 5.1 to compute the homology of $(\hat{0}, x)$ (see Example 5.1.12 and Exercise 5.1.13). This yields,

Theorem 5.4.4 (Wachs [**201**]). *Suppose $S \subseteq \{2, 3, \dots\}$ is such that S and $\{s - \min S : s \in S\}$ are closed under addition (eg., $S = \{k, k+1, \dots\}$ or more generally $S = \{kd, (k+1)d, \dots\}$). For $n \in S$, let M_n^S be the manifold*

$$\{\mathbf{z} := (z_1, \dots, z_n) \in \mathbb{C}^n : \text{ for some } i \text{ the number of occurrences of } z_i \text{ is not in } S\}.$$

Then

$$\sum_{\substack{m \in \mathbb{Z} \\ n \in S}} \mathrm{ch}\tilde{H}^m(M_n^S)\, u^n\, (-t)^{2n-m-1} = \sum_{i \geq 1} h_i t^i \Big[\Big(\sum_{i \geq 1} h_i\Big)^{[-1]}\Big]\Big[\sum_{i \in S} h_i\, u^i\ t^{\phi(i)}\Big],$$

where $\phi(i) := \max\{j \in \mathbb{Z}^+ : i - (j-1)\min S \in S\}$. (Our notation reflects the fact that plethysm is associative.)

We leave it as an open problem to use this formula to compute the multiplicity of the trivial representation in $\tilde{H}^m(M_n^S)$, and thereby obtain a formula (which generalizes the above mentioned Sundaram-Welker formula) for the Betti numbers of the space of monic polynomials of degree n with at least one root multiplicity not in S.

5.5. Inflations of Simplicial Complexes

Let Δ be a simplicial complex on vertex set $[n]$ and let $\mathbf{m} = (m_1, \dots, m_n)$ be an n-tuple of positive integers. We form a new simplicial complex $\Delta_{\mathbf{m}}$, called the $\mathbf{m}$-*inflation* of Δ, as follows. The vertex set of $\Delta_{\mathbf{m}}$ is $\{(i, c) : i \in [n], c \in [m_i]\}$ and the faces of $\Delta_{\mathbf{m}}$ are of the form $\{(i_1, c_1), \dots, (i_k, c_k)\}$ where $\{i_1, \dots, i_k\}$ is a $(k-1)$-face of Δ and $c_j \in [m_{i_j}]$ for all $j = 1, \dots, k$. We can think of c_j as a color assigned to vertex i_j and of $\{(i_1, c_1), \dots, (i_k, c_k)\}$ as a coloring of the vertices of face $\{i_1, \dots, i_k\}$. A color for vertex i is chosen from m_i colors.

Example 5.5.1. Let P and Q be the posets depicted in Figure 5.3.1. Clearly $\Delta(P)$ is the $(2, 2, 2)$-inflation of $\Delta(Q)$.

Theorem 5.5.2 (Björner, Wachs, and Welker [**43**]). *Let Δ be a simplicial complex on vertex set $[n]$ and let $\mathbf{m}$ be an n-tuple positive integers. If Δ is connected then*

$$\Delta_{\mathbf{m}} \simeq \bigvee_{F \in \Delta} (susp^{|F|}(lk_\Delta F))^{\vee \nu(F, \mathbf{m})},$$

where $\nu(F, \mathbf{m}) = \prod_{i \in F}(m_i - 1)$, $susp^k(\Delta)$ denotes the k-fold suspension of Δ, and $\Delta^{\vee k}$ denotes the k-fold wedge of Δ.

We prove this theorem in the following exercise.

Exercise 5.5.3. Let $f : \Delta_{\mathbf{m}} \to \Delta$ be the simplicial map that sends each vertex (i, c) of $\Delta_{\mathbf{m}}$ to vertex i of Δ. We call this the *deflating map*. This can be viewed as a poset map $f : P(\Delta_{\mathbf{m}}) \to P(\Delta)$.

(a) Show that the fiber $f^{-1}(P(\Delta)_{\leq F})$ is a geometric semilattice.
(b) Show that each $f^{-1}(\overline{P(\Delta)}_{\leq F})$ is a wedge of $\nu(F, \mathbf{m})$ spheres of dimension $\dim F$.
(c) Use (b) and the fact that the join operation is distributive over the wedge operation to prove Theorem 5.5.2.
(d) Show that $\Delta_{\mathbf{m}}$ is sequentially Cohen-Macaulay if Δ is.

Theorem 5.5.4. *Let Δ be a G-simplicial complex on vertex set $[n]$ and let $\mathbf{m}$ be an n-tuple of positive integers. If G acts on the inflation $\Delta_{\mathbf{m}}$ and this action commutes with the deflating map, then for all $r \in \mathbb{Z}$,*

$$\tilde{H}_r(\Delta_{\mathbf{m}}) \cong_G \bigoplus_{F \in \Delta/G} \left(\tilde{H}_{|F|-1}(\langle F \rangle_{\mathbf{m}(F)}) \otimes \tilde{H}_{r-|F|}(lk_\Delta F) \right) \uparrow_{G_F}^G,$$

where $\langle F \rangle$ is the subcomplex generated by F and $\mathbf{m}(F)$ is the subsequence $(m_{i_1}, \ldots, m_{i_t})$ of $\mathbf{m} = (m_1, \ldots, m_n)$ for $F = \{i_1 < \cdots < i_t\}$.

Shareshian [**156**] and Shareshian and Wachs [**158**] use Theorems 5.5.2 and 5.5.4, to derive information about the homology of the Quillen complex $\mathcal{A}_p(\mathfrak{S}_n)$ from a hypergraph version of the matching complex. Another application of Theorems 5.5.2 and 5.5.4 can be found in [**133**]. Here we discuss the original example that led to the general fiber theorem, namely the colored chessboard complex [**203**].

Recall from Section 2.4.3 that the chessboard complex $M_{m,n}$ is the collection of rook placements on an $m \times n$ chessboard such that there is at most one rook in each row and each column. This complex arose in the work of Garst [**78**] on Tits coset complexes, in the work of Živaljević and Vrećica [**223**] on the colored Tverberg problem, in the work of Reiner and Roberts [**140**] on Segre algebras, and in the work of Babson and Reiner [**10**] on generalizations of Coxeter complexes. Like the matching complex, the chessboard complex has been extensively studied in the literature; see the survey article of Wachs [**202**]. We state just a few of the results on chessboard complexes here.

Theorem 5.5.5. *Let $1 \leq m \leq n$ and $\nu_{m,n} = \min\{m, \lfloor \frac{m+n+1}{3} \rfloor\} - 1$.*
(a) (Garst [**78**]) *If $n \geq 2m - 1$ then $M_{m,n}$ is Cohen-Macaulay.*
(b) (Björner, Lovász, Vrećica and Živaljević [**35**]) *The complex $M_{m,n}$ is $(\nu_{m,n} - 1)$-connected.*
(c) (Shareshian and Wachs [**157**]) *For $n \leq 2m - 5$ and $m + n \equiv 1 \bmod 3$,*

$$\tilde{H}_{\nu_{m,n}}(M_{m,n}; \mathbb{Z}) \cong \mathbb{Z}_3.$$

Part (a) follows from part (b), and Ziegler [**219**] gives an alternative proof of part (b) by establishing shellability of the $\nu_{m,n}$-skeleton of $M_{m,n}$. Part (c) is one case of a more general torsion result, whose proof uses a chessboard complex analog of Bouc's representation theoretic formula (2.4.5) (due to Garst [**78**] for top homology, and to Friedman and Hanlon [**75**] in general) to construct a basis for the vector space $\tilde{H}_{m-1}(M_{m,n})$. This basis consists of fundamental cycles constructed from pairs of standard Young tableaux via a classical combinatorial algorithm known as the Knuth-Robinson-Schensted correspondence. We remark that Parts (b) and (c) have analogs for the matching complex [**35**] [**47**] [**157**].

Let $M_{m,n}^r$ be the simplicial complex of placements of colored rooks on an $m \times n$ chessboard such that there is at most one rook in each row and each column, and the colors come from the set $[r]$. The colored chessboard complex arose in Garst's

work as a more general example of a Tits coset complex. One can easily see that $M^r_{m,n}$ is the $(r, r, \ldots, r)$-inflation of $M_{m,n}$.

Exercise 5.5.6 (Wachs [**203**])**.** Use Theorems 5.5.2, 5.5.4, and 5.5.5 to obtain a colored version of Theorem 5.5.5, i.e. a generalization to $M^r_{m,n}$, where $r \geq 1$.

For open problems on the topology of chessboard complexes, matching complexes and related complexes, see [**202**] and [**157**].

BIBLIOGRAPHY

1. F. Ardila and C. Klivans, *The Bergman complex of a matroid and phylogenetic trees*, J. Combin. Theory (B), **96** (2006), 38-49.
2. D. Armstrong, *Braid groups, clusters, and free probability: an outline from the AIM workshop*, Jan. 2005, http://www.aimath.org/WWN/braidgroups/.
3. V.I. Arnol'd, *Topological invariants of algebraic functions*, Trans. Moscow Math. Soc. **21** (1970), 30-52.
4. M. Aschbacher and S. Smith, *On Quillen's conjecture for the p-groups complex*, Ann. of Math. **2** (1993), 473–529.
5. C. Athanasiadis, *Decompositions and connectivity of matching and chessboard complexes*, Discrete Compute. Geom. **31** (2004), 395–403.
6. C. Athanasiadis, T. Brady and C. Watt, *Shellability of noncrossing partition lattices*, Proc. Amer. Math. Soc. **135** (2007), 939-949.
7. C. Athanasiadis and V. Reiner, *Noncrossing partitions for the group D_n*, SIAM J. Discrete Math. **18** (2004), 397–417.
8. E. Babson, A. Björner, S. Linusson, J. Shareshian, and V. Welker, *Complexes of not i-connected graphs*, Topology **38** (1999), 271–299.
9. E. Babson and P. Hersh, *Discrete Morse functions from lexicographic orders*, Trans. AMS **357** (2005), 509–534.
10. E. Babson and V. Reiner, *Coxeter-like complexes*, Discrete Math. Theor. Comput. Sci. **6** (2004), 223–252.
11. K. Baclawski, *Whitney numbers of geometric lattices*, Advances in Math. **16** (1975), 125–138.
12. K. Baclawski, *Cohen-Macaulay ordered sets*, J. Algebra **63** (1980), 226–258.
13. K. Baclawski and A. Björner, *Fixed points and complements in finite lattices*, J. Combin. Theory (A) **30** (1981), 335–338.
14. H. Barcelo, *On the action of the symmetric group on the free Lie algebra and the partition lattice*, J. Combin. Theory (A) **55** (1990), 93–129.
15. M.M. Bayer and L.J. Billera, *Generalized Dehn-Sommerville relations for polytopes, spheres and Eulerian partially ordered sets*, Invent. Math. **79** (1985), 143–157.
16. N. Bergeron, *A hyperoctahedral analogue of the free Lie algebra*, J. Combin. Theory (A) **58** (1991), 256-278.
17. D. Bessis, *The dual braid monoid*, Annales Scientifiques de l'École Normale Suprieure, **36** (2003), 647–683.

18. L.J. Billera, S. Holmes, and K. Vogtmann, *Geometry of the space of phylogenetic trees*, Advances in Appl. Math. **27** (2001), 733–767.
19. L.J. Billera and C.W. Lee, *A proof of sufficiency of McMullen's conditions for f-vectors of simplicial complexes*, J. Combin. Th. (A) **31** (1981), 237–255.
20. L.J. Billera and A.N. Myers, *Shellability of interval orders*, Order **15** (1998/99), 113–117.
21. A. Björner, *Shellable and Cohen-Macaulay partially ordered sets*, Trans. AMS **260** (1980), 159–183.
22. A. Björner, *On the homology of geometric lattices*, Alg. Univ. **14** (1982), 107–128.
23. A. Björner, *Posets, regular CW-complexes and Bruhat order*, Europ. J. Combin. **5** (1984), 7–16.
24. A. Björner, *Orderings of Coxeter groups*, combinatorics and algebra (Boulder, Colo., 1983), Contemp. Math. **34** (1984), 175–195.
25. A. Björner, *The Möbius function of subword order*, in Invariant Theory and Tableaux, ed. D. Stanton, IMA Volumes in Math. and its Applic., Vol 19, Springer-Verlag, New York, 1990, 118–124.
26. A. Björner, *The homology and shellability of matroids and geometric lattices*, in Matroid Applications, ed. N. White, Cambridge University Press, 1992, pp. 226–283.
27. A. Björner, *The Möbius function of factor order*, Theoretical Computer Science **117** (1993), 91-98.
28. A. Björner, *Subspace arrangements*, First European Congress of Mathematics, Paris 1992, A. Joseph et al. (Eds.), Progress in Math., Vol. 119, Birkhäuser, 1994, pp. 321–370.
29. A. Björner, *Topological Methods*, Handbook of Combinatorics, R. Graham, M. Grötschel and L. Lovász, (Eds), North-Holland, Amsterdam, 1995, pp. 1819-1872.
30. A. Björner and F. Brenti, *Combinatorics of Coxeter Groups*, Graduate Texts in Mathematics, Vol. 231, Springer-Verlag, New York, 2005.
31. A. Björner and K. Eriksson, *Extendable shellability for rank 3 matroid complexes*, Discrete Math. **132** (1994), 373–376.
32. A. Björner, A. Garsia, and R. Stanley, *An introduction to the theory of Cohen-Macaulay posets*, in Ordered Sets (I. Rival, ed.), Reidel, Dordrecht/Boston/London, 1982, pp. 583-615.
33. A. Björner and L. Lovász, *Linear decision trees, subspace arrangements and Möbius functions*, J. AMS **7** (1994), 677–706.
34. A. Björner, L. Lovász and A. Yao, *Linear decision trees: volume estimates and topological bounds*, Proc. 24th ACM Symp. on Theory of Computing (May 1992), ACM Press , New York, 1992, 170–177.
35. A. Björner, L. Lovász, S.T. Vrećica and R.T. Živaljević, *Chessboard complexes and matching complexes*, J. London Math. Soc. **49** (1994), 25–39.
36. A. Björner and B.E. Sagan, *Subspace arrangements of type B_n and D_n*, J. Algebraic Combin. **5** (1996), 291–314.
37. A. Björner and M.L. Wachs, *Bruhat order of Coxeter groups and shellability*, Advances in Math. **43** (1982), 87–100.
38. A. Björner and M.L. Wachs, *On lexicographically shellable posets*, Trans. AMS **277** (1983), 323–341.

39. A. Björner and M.L. Wachs, *Generalized quotients in Coxeter groups*, Trans. AMS **308** (1988), 1-37.
40. A. Björner and M.L. Wachs, *Nonpure shellable complexes and posets I*, Trans. AMS **348** (1996), 1299–1327.
41. A. Björner and M.L. Wachs, *Nonpure shellable complexes and posets II*, Trans. AMS **349** (1997), 3945–3975.
42. A. Björner and M.L. Wachs, *Geometrically constructed bases for homology of partition lattices of type A, B and D*, special issue in honor of Richard Stanley, Electron. J. Combin. **11** (2004/05), Research Paper 3, 26 pp.
43. A. Björner, M.L. Wachs and V. Welker, *Poset fiber theorems*, Trans. AMS **357** (2005), 1877–1899.
44. A. Björner, M.L. Wachs and V. Welker, *On sequentially Cohen-Macaulay complexes and posets*, preprint 2007.
45. A. Björner and V. Welker, *The homology of "k-equal" manifolds and related partition lattices*, Advances in Math. **110** (1995), 277–313.
46. J.M. Boardman, *Homotopy structures and the language of trees*, Algebraic topology (Wisconsin, 1970) Proc. Symp. Pure Math. **22** (1971), 37–58.
47. S. Bouc, *Homologie de certains ensembles de* 2*-sous-groupes des groupes symétriques*, J. Algebra **150** (1992), 158–186.
48. T. Brady, *A partial order on the symmetric group and new $K(\pi,1)$'s for the braid groups*, Advances in Math. **161** (2001), 20–40.
49. T. Brady and C. Watt, *$K(\pi,1)$'s for Artin groups of finite type*, Proceedings of the Conference on Geometric and Combinatorial Group Theory, Part I, Geom. Dedicata **94** (2002), 225–250.
50. A.E. Browdy *The (co)homology of lattices of partitions with restricted block size*, Ph.D. dissertation, University of Miami, 1996.
51. A. Browdy and M.L. Wachs, *On the (co)homology of the lattice of partitions with restricted block size*, in preparation.
52. H. Bruggesser and P. Mani, *Shellable decompositions of cells and spheres*, Math. Scand. **29** (1971), 197–205.
53. C.C. Cadogan, *The Möbius function and connnected graphs*, J. Combin. Theory (B) **11** (1971), 193–200.
54. A.R. Calderbank, P. Hanlon and R.W. Robinson, *Partitions into even and odd block size and some unusual characters of the symmetric groups*, Proc. London Math. Soc. (3) **53** (1986), 288–320.
55. A. Chakrabarti, S. Khot and Y. Shi, *Evasiveness of subgraph containment and related properties*, SIAM J. Comput. **31** (2001/02), 866–875.
56. G. Danaraj and V. Klee, *Shellings of spheres and polytopes*, Duke Math. J. **41** (1974), 443–451.
57. C. de Concini and V. Lakshmibai, *Arithmetic Cohen-Macaulayness and arithmetic normality for Schubert varieties*, Amer. J. Math. **103** (1981), 835–850.
58. C. de Concini and C. Procesi, *Wonderful models of subspace arrangements*, Selecta Math., New Series **1** (1995), 459-494.
59. J. Désarménian, *Une autre interprétation du nombre de dérangements*, Sém. Lotharing. Combin. **9** (1984), 11-16.
60. J. Désarménian and M.L. Wachs, *Descentes des derangements et mot circulaires*, Sém Lotharing. Combin. **19** (1988), 13–21.

61. X. Dong, *Topology of bounded-degree graph complexes*, J. Algebra **262** (2003), 287–312.
62. X. Dong and M.L. Wachs, *Combinatorial Laplacian of the matching complex*, Electron. J. Combin. **9** (2002), R17.
63. J. G. Dumas, F. Heckenbach, D. Saunders and V. Welker, *Computing simplicial homology based on efficient Smith normal form algorithms*, Algebra, Geometry, and Software Systems, 177–206, Springer, Berlin, 2003.
64. A.M. Duval, *Algebraic shifting and sequentially Cohen-Macaulay simplicial complexes*, Electron. J. Combin. **3** (1996).
65. M.J. Dyer, *Hecke algebras and shellings of Bruhat intervals*, Compositio Math. **89** (1993), 91–115.
66. J.A. Eagon and V. Reiner, *Resolutions of Stanley-Reisner rings and Alexander duality*, J. Pure and Appl. Algebra **130** (1998), 265-275.
67. P.H. Edelman, *The Bruhat order of the symmetric group is lexicographically shellable*, Proc. AMS **82** (1981), 355–358.
68. F.D. Farmer, *Cellular homology for posets*, Math. Japonica **23** (1979), 607–613.
69. E.M. Feichtner and G.M. Ziegler, *On cohomology algebras of complex subspace arrangements*, Trans. AMS **352** (2000), 3523-3555.
70. J. Folkman, *The homology groups of a lattice*, J. Math. Mech. **15** (1966), 631–636.
71. S. Fomin and N. Reading, *Root systems and generalized associahedra*, PCMI Lecture Notes, 2004.
72. R. Forman, *Morse theory for cell complexes*, Advances in Math. **134** (1998), 90–145.
73. R. Forman, *Morse theory and evasiveness*, Combinatorica **20** (2000), 498–504.
74. R. Forman, *Discrete Morse theory*, PCMI Lecture Notes, 2004.
75. J. Friedman and P. Hanlon, *On the Betti numbers of chessboard complexes*, J. Algebraic Combin. **8** (1998), 193-203.
76. W. Fulton, *Young Tableaux, with Applications to Representation Theory and Geometry*, LMS Student Texts **35**, Cambridge University Press, Cambridge, 1997.
77. A.M. Garsia, *Combinatorial methods in the theory of Cohen-Macaulay rings*, Advances in Math. **38** (1980), 229–266.
78. P.F. Garst, *Cohen-Macaulay complexes and group actions*, Ph.D. Thesis, University of Wisconsin-Madison, 1979.
79. E. Getzler, and M.M. Kapranov, *Cyclic operads and cyclic homology* in Geometry, Topology and Physics, International Press, Cambridge, MA, 1995, pp. 167–201.
80. M. Goresky and R.D. MacPherson *Stratified Morse theory*, Ergebnisse der Mathematic und ihrer Grenzgebiete **14**, Springer-Verlag, 1988.
81. E. Gottlieb, *Cohomology of Dowling lattices and Lie (super)algebras*, Ph.D. Thesis, University of Miami, 1998.
82. E. Gottlieb, *On the homology of the h, k-equal Dowling lattice*, SIAM J. Discrete Math. **17** (2003), 50–71.
83. E. Gottlieb and M. L. Wachs, *Cohomology of Dowling lattices and Lie (super)algebras*, Advances in Appl. Math. **24** (2000), 301–336.

84. H.T. Hall, *Counterexamples in Discrete Geometry*, Ph. D. Thesis, University of California, Berkeley, 2004.
85. P. Hanlon, *The fixed-point partition lattices*, Pacific J. Math. **96** (1981), 319–341.
86. P. Hanlon, *The characters of the wreath product group acting on the homology groups of the Dowling lattices*, J. Algebra **91** (1984), 430–463.
87. P. Hanlon, *Otter's method and the homology of homeomorphically irreducible k-trees*, Jour. Combin. Theory (A), **74** (1996), 301–320.
88. P. Hanlon and P. Hersh, *Multiplicity of the trivial representation in rank-selected homology of the partition lattice*, J. Algebra **266** (2003), 521–538.
89. P. Hanlon and P. Hersh, *A hodge decomposition for the complex of injective words*, Pacific J. Math. **214** (2004), 109–125.
90. P. Hanlon and R.P. Stanley, *A q-deformation of a trivial symmetric group action*, Trans. AMS **350** (1998), 4445–4459.
91. P. Hanlon and M.L. Wachs, *On Lie k-algebras*, Advances in Math. **113** (1995), 206-236.
92. A. Hatcher, *Algebraic Topology*, Cambridge University Press, 2001.
93. A. Henderson, *Plethysm for wreath products and homology of sub-posets of Dowling lattices*, preprint 2006.
94. P. Hersh, *Lexicographic shellability for balanced complexes*, J. Algebraic Combin. **17** (2003), 225–254.
95. J. Herzog and T. Hibi, *Componentwise linear ideals*, Nagoya Math. J. **153** (1999), 141 153.
96. J. Herzog, V. Reiner, and V. Welker, *Componentwise linear ideals and Golod rings*, Michigan Math. J. **46** (1999), 211–223.
97. M. Hochster, *Rings of invariants of tori, Cohen-Macaulay rings generated by monomials, and polytopes*, Annals of Math. **96** (1972), 318–337.
98. A. Hultman, *Lexicographic shellability and quotient complexes*, J. Algebraic Combin. **16** (2002), 83–86.
99. A. Hultman, *The combinatorics of twisted involutions in Coxeter groups*, Trans. AMS, **359** (2007), 2787-2798.
100. A. Hultman, *The topology of spaces of phylogenetic trees with symmetry*, Discrete Math., to appear.
101. F. Incitti, *The Bruhat order on the involutions of the symmetric group*, J. Algebraic Combin. **20** (2004), 243–261.
102. F. Incitti, *The Bruhat order on the involutions of the hyperoctahedral group*, Europ. J. Combin. **24** (2003), 825–848.
103. F. Incitti, *Bruhat order on the involutions of classical Weyl groups*, Advances in Appl. Math., **37** (2006), 68–111.
104. J. Jonsson, *On the topology of simplicial complexes related to 3-connected and Hamiltonian graphs*, J. Combin. Theory (A) **104** (2003), 169–199.
105. J. Jonsson, *Simplicial Complexes of Graphs*, Ph.D. thesis, Royal Institute of Technology, Sweden, 2005.
106. A. Joyal, *Foncteurs analytiques et espèces de structure*, Springer Lecture Notes, No. 1234, Springer-Verlag, Berlin, Heidelberg, New York, 1986.
107. D.B. Karaguezian, V. Reiner and M.L. Wachs, *Matching complexes, bounded degree graph complexes, and weight spaces of GL_n-complexes*, J. Algebra **239** (2001), 77–92.

108. J. Kahn, M. Saks and D. Sturtevant, *A topological approach to evasiveness*, Combinatorica **4** (1984), 297–306.
109. B. Kind and P. Kleinschmidt, *Schälbare Cohen-Macauley-Komplexe und ihre Parametrisierung*, Math. Z. **167** (1979), 173-179.
110. A.A. Klyachko, *Lie elements in the tensor algebra*, Siberian Math. J. **15** (1971), 1296–1304.
111. M. Kontsevich, *Formal (non)-commutative symplectic geometry*, in the Gelfand Mathematical Seminar, 1990–1992, eds L. Corwin et al., Birkhäuser, Boston, 1993, pp. 173–187.
112. D. Kozlov, *General lexicographic shellability and orbit arrangements*, Ann. of Combin. **1** (1997), 67–90.
113. D. Kozlov, *Rational homology of spaces of complex monic polynomials with multiple roots*, Mathematika **49** (2002), 77–91.
114. C. Kratzer and J. Thévenaz, *Type d'homotopie des treillis et treillis des sous-groupes d'un groupe fini*, Comment. Math. Helv. **60** (1985), 86–106.
115. G. Kreweras, *Sur les partitions non croisées d'un cycle*, Discrete Math. **1** (1972), 333–350.
116. R. Ksontini, *Propriétés homotopiques du complexe de Quillen du groupe symétrique*, Thèse de Doctorate, Université de Lausanne, 2000.
117. R. Ksontini, *Simple connectivity of the Quillen complex of the symmetric group*, J. Combin. Theory (A) **103** (2003), 257–279.
118. R. Ksontini, *The fundamental group of the Quillen complex of the symmetric group*, J. Algebra **282** (2004), 33–57.
119. G. Lehrer and L. Solomon, *On the action of the symmetric group on the cohomology of the complement of its reflecting hyperplanes*, J. Algebra **104** (1986), 410–424.
120. W.B. Lickorish, *Unshellable triangulations of spheres*, European J. Combin. **12** (1991) 527–530.
121. S. Linusson, *Partitions with restricted block sizes, Möbius functions, and the k-of-each problem*, SIAM J. Discrete Math. **10** (1997), 18–29.
122. S. Linusson, J. Shareshian and V. Welker, *Complexes of graphs with bounded matching size*, preprint 2003.
123. D.E. Littlewood, *The Theory of Group Characters and Matrix Representations of Groups*, 2nd ed., Oxford University Press, 1950.
124. G. Lusztig, *Total positivity in partial flag manifolds*, Representation Theory, **2** (1998), 70–78.
125. F.H. Lutz, *Small examples of nonconstructible simplicial balls and spheres*, SIAM J. Discrete Math. **18** (2004), 103–109.
126. I.G. Macdonald, *Symmetric Functions and Hall Polynomials*, Oxford Mathematical Monographs, Oxford University Press, New York, 1979.
127. J. McCammond, *Noncrossing partitions in surprising locations*, Amer. Math. Monthly, **113** (2006), 598-610.
128. P. McMullen, *The maximum number of faces of a convex polytope*, Mathematika **17** (1970), 179–184.
129. E. Miller and B. Sturmfels, *Combinatorial Commutative Algebra*, Grad. Texts in Math., Springer-Verlag, New York 2004.
130. J. Munkres, *Topological results in combinatorics*, Michigan Math. J. **31** (1984), 113–128.

131. P. Orlik and L. Solomon, *Combinatorics and topology of complements of hyperplanes*, Invent. Math. **56** (1980), 167–189.
132. P. Orlik and H. Terao, *Arrangements of Hyperplanes*, Grundlehren Series **300**, Springer Verlag, Berlin-Heidelberg.
133. J. Pakianathan and E. Yalçin, *On commuting and non-commuting complexes*, J. Algebra **236** (2001), 396–418.
134. R. Proctor, *Classical Bruhat orders and lexicographically shellability*, J. Algebra **77** (1982), 104–126.
135. J.S. Provan and L.J. Billera, *Decompositions of simplicial complexes related to diameters of convex polyhedra*, Math. Oper. Res. **5** (1980), 576–594.
136. P. Pudlák and J. Tuma, *Every finite lattice can be embedded in a finite partition lattice*, Alg. Univ. **10** (1980), 74–95.
137. D. Quillen, *Homotopy properties of the poset of non-trivial p-subgroups of a group*, Advances in Math. **28** (1978), 101–128.
138. M. Readdy, *The pre-WDVV ring of physics and its topology*, Ramanujan J., **10** (2005), 269-281.
139. V. Reiner, *Non-crossing partitions for classical reflection groups*, Discrete Math. **177** (1997), 195–222.
140. V. Reiner and J. Roberts, *Minimal resolutions and homology of chessboard and matching complexes*, J. Algebraic Combin. **11** (2000), 135–154.
141. V. Reiner and M.L. Wachs, *A decomposition of the eigenspaces of the random to top card shuffling operator*, in preparation.
142. V. Reiner and P. Webb, *The combinatorics of the bar resolution in group cohomology*, J. Pure Appl. Algebra **190** (2004), 291–327.
143. V. Reiner and V. Welker, *A homological lower bound for order dimension of lattices*, Order **16** (1999), 165–170.
144. G. Reisner, *Cohen-Macaulay quotients of polynomial rings*, Advances in Math. **21** (1976), 30–49.
145. R.W. Richardson and T.A. Springer, *The Bruhat order on symmetric varieties*, Geom. Dedicata **49** (1990), 389–436.
146. K. Rietsch, *Total positivity and real flag varieties*, Ph.D. Thesis, MIT, 1998.
147. A. Robinson, *Partition complexes, duality and integral tree representations*, Algebr. Geom. Topol. **4** (2004), 943–960.
148. C.A. Robinson and S. Whitehouse *The tree representation of* $\sum_{n+1}$, J. Pure and App. Algebra **111** (1996), 245–253.
149. G.-C. Rota, *On the foundations of combinatorial theory, I. Theory of Möbius functions*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **2** (1964), 340–368.
150. B.E. Sagan, *Shellability of exponential structures*, Order **3** (1986), 47–54.
151. B.E. Sagan, *The Symmetric Group. Representations, Combinatorial Algorithms, and Symmetric Functions*, second edition, Graduate Texts in Mathematics, 203, Springer-Verlag, New York, 2001.
152. L. Schläfli, *Theorie der vielfachen Kontiutät*, written 1850-1852; Zürcher und Furrer, Zürich 1901; Denkschriften der Schweizerischen naturforschenden Gesellschaft **38** (1901), 1–237.
153. J. Shareshian, *On the shellability of the order complex of the subgroup lattice of a finite group*, Trans. AMS **353** (2001), 2689–2703.

154. J. Shareshian, *Discrete Morse theory for complexes of* 2*-connected graphs*, Topology **40** (2001), 681–701.
155. J. Shareshian, *Topology of order complexes of intervals in subgroup lattices*, J. Algebra **268** (2003), 677-686.
156. J. Shareshian, *Hypergraph matching complexes and Quillen complexes of symmetric groups*, J. Combin. Theory (A) **106** (2004), 299–314.
157. J. Shareshian and M.L. Wachs, *Torsion in the matching complex and chessboard complex*, Advances in Math., to appear.
158. J. Shareshian and M.L. Wachs, *Top homology of hypergraph matching complexes, p-cycle complexes and Quillen complexes of symmetric groups*, in preparation.
159. J. Shareshian and M.L. Wachs, *On the* 1 mod k *partition poset and graph complexes*, in preparation.
160. R. Simion, *On q-analogues of partially ordered sets*, J. Combin. Theory (A) **72** (1995), 135–183.
161. R. Simion, *Noncrossing partitions*, Discrete Math. **217** (2000), 367–409.
162. R.S. Simon, *Combinatorial properties of "cleanness"*, J. Algebra **167** (1994), 361–38.
163. L. Solomon, *A decomposition of the group algebra of a finite Coxeter group*, J. Algebra **9** (1968), 220–239.
164. R.P. Stanley, *Ordered structures and partitions*, Memoirs AMS no. 119 (1972).
165. R.P. Stanley, *Supersolvable lattices*, Alg. Univ. **2** (1972), 197–217.
166. R.P. Stanley, *Finite lattices and Jordan-Hölder sets*, Alg. Univ. **4** (1974), 361-371.
167. R.P. Stanley, *Exponential structures*, Stud. Appl. Math. **59** (1978), 73–82.
168. R.P. Stanley, *The number of faces of a simplicial convex polytope*, Advances in Math. **35** (1980), 236-238.
169. R.P. Stanley, *Some aspects of groups acting on finite posets*, J. Combin. Theory (A) **32** (1982), 132–161.
170. R.P. Stanley, *Combinatorics and Commutative Algebra, Second Edition*, Progress in Mathematics **41**, Birkhäuser, Boston 1983, 2nd ed. 1996.
171. R.P. Stanley, *Enumerative Combinatorics, Vol. 1*, Wadsworth & Brooks/Cole, Monterey, CA, 1986, Second edition, Birkhäuser, Boston, 1995.
172. R.P. Stanley, *Flag-symmetric and locally rank-symmetric partially ordered sets*, Foata Festschrift, Electron. J. Combin. **3** (1996), RP 6, 22 pp.
173. R.P. Stanley, *Parking functions and noncrossing partitions*, Wilf Festschrift, Electron. J. Combin. **4** (1997), RP 20, 14 pp.
174. R.P. Stanley, *Enumerative Combinatorics, Vol. 2*, Cambridge Studies in Advanced Mathematics **62**, Cambridge University Press, Cambridge, 1999.
175. R.P. Stanley *Recent developments in algebraic combinatorics*, Israel J. of Math. **143** (2004), 317–339.
176. R.P. Stanley, *An introduction to hyperplane arrangements*, PCMI Lecture Notes, 2004.
177. S. Sundaram, *The homology representations of the symmetric group on Cohen-Macaulay subposets of the partition lattice*, Advances in Math. **104** (1994), 225–296.
178. S. Sundaram, *Applications of the Hopf trace formula to computing homology representations*, Jerusalem combinatorics '93, 277–309, Contemp. Math., **178**,

Amer. Math. Soc., Providence, RI, 1994.

179. S. Sundaram, *The homology of partitions with an even number of blocks*, J. Algebra **4** (1995). 69–92.
180. S. Sundaram, *Plethysm, partitions with an even number of blocks and Euler numbers*, in "Formal Power Series and Algebraic Combinatorics 1994", 171–198, DIMACS Series in Discrete Mathematics and Theoretical Computer Science **24**, Amer. Math. Soc., Providence, RI, 1996.
181. S. Sundaram, *Homotopy of non-modular partitions and the Whitehouse module*, J. Algebraic Combin. **9** (1999), 251–269.
182. S. Sundaram, *On the topology of two partition posets with forbidden block sizes*, J. Pure Appl. Algebra **155** (2001), 271–304.
183. S. Sundaram, *A homotopy equivalence for partition posets related to liftings of S_{n-1}-modules to S_n*, J. Combin. Theory (A) **94** (2001), 156–168.
184. S. Sundaram and M.L. Wachs, *The homology representations of the k-equal partition lattice*, Trans. AMS **349** (1997), 935–954.
185. S. Sundaram and V. Welker, *Group representations on the homology of products of posets*, J. Combin. Theory (A) **73** (1996), 174–180.
186. S. Sundaram and V. Welker, *Group actions on arrangements and applications to configuration spaces*, Trans. AMS **349** (1997), 1389–1420.
187. E. Swartz, *g-Elements, finite buildings and higher Cohen-Macaulay connectivity*, J. Combin. Theory Ser. A **113** (2006), 1305-1320.
188. G. Sylvester, *Continuous spin ising ferromagnets*, Ph.D. Thesis, MIT, 1976.
189. J. Thévenaz and P.J. Webb, *Homotopy equivalence of posets with a group action*, J. Combin. Theory (A) **56** (1991), 173–181.
190. H. Trappmann and G. Ziegler, *Shellability of complexes of trees*, J. Combin. Theory (A) **82** (1998), 168–178.
191. W. Trotter, *Combinatorics and Partially Ordered Sets: Dimension Theory*, Johns Hopkins Series in the Mathematical Sciences, Johns Hopkins University, Press, Baltimore, 1992.
192. V. Turchin, *Homologies of complexes of doubly connected graphs*, Russian Mathematical Surveys (Uspekhi) **52** (1997), 426–427.
193. V.A. Vassiliev, *Complexes of connected graphs* in the Gelfand Mathematical Seminar, 1990–1992, eds. L. Corwin et al., Birkhäuser, Boston, 1993, pp. 223–235.
194. V.A. Vassiliev, *Complements of Discriminants of Smooth Maps: Topology and Applications*, Revised Edition, Vol. 98 of Translations of Mathematical Monographs., Amer. Math. Soc., Providence, RI, 1994.
195. V.A. Vassiliev, *Topology of two-connected graphs and homology of spaces of knots*, in: S. L. Tabachnikov (ed.), "Differential and Symplectic Topology of Knots and Curves", AMS Transl. Ser. 2, **190**, Amer. Math. Soc., Providence, RI, 1999.
196. A. Vince and M.L. Wachs, *A shellable poset that is not lexicographically shellable*, Combinatorica **5** (1985), 257–260.
197. K. Vogtmann, *Local Structure of some out (F_n)-complexes*, Proc. Edinburgh Math. Soc. (2) **33** (1990), 367–379.
198. M. L. Wachs, *A basis for the homology of the d-divisible partition lattice*, Advances in Math. **117** (1996), 294–318.

199. M. L. Wachs, *On the (co)homology of the partition lattice and the free Lie algebra*, Special issue in honor of Adriano Garsia, Discrete Math. **193** (1998), 287–319.
200. M. L. Wachs, *Obstructions to shellability*, Discrete Compute. Geom. **22** (1999), 95-103.
201. M.L. Wachs, *Whitney homology of semipure shellable posets*, J. Algebraic Combin. **9** (1999), 173–207.
202. M.L. Wachs, *Topology of matching, chessboard, and general bounded degree graph complexes*, Special issue in memory of Gian-Carlo Rota, Algebra Universalis **49** (2003), 345–385.
203. M.L. Wachs, *Bounded degree digraph and multigraph matching complexes*, in preparation.
204. M.L. Wachs, *Poset fiber theorems and Dowling lattices*, in preparation.
205. M.L. Wachs and J.W. Walker, *On geometric semilattices*, Order **2** (1986), 367-385.
206. J.W. Walker, *Topology and Combinatorics of Ordered Sets*, MIT Ph.D. thesis, 1981.
207. J.W. Walker, *Canonical homeomorphisms of posets*, European J. Combin. **9** (1988), 97–107.
208. J.W. Walker, *A poset which is shellable but not lexicographically shellable*, European J. Combin. **6** (1985), 287–288.
209. P. J. Webb, *Subgroup complexes*, Arcata Conference on Representations of Finite Groups (Arcata, Calif., 1986), 349-365, Proc. Sympos. Pure Math., **47**, Part 1, Amer. Math. Soc., Providence, RI, 1987.
210. V. Welker, G.M. Ziegler and R.T. Živaljević, *Homotopy colimits – comparison lemmas for combinatorial applications*, J. Reine Angew. Mathematik (Crelles Journal) **509** (1999), 117–149.
211. S. Whitehouse, *The Eulerian representations of $\sum_n$ as restrictions of representations of $\sum_{n+1}$*, J. Pure Appl. Algebra **115** (1996), 309–321.
212. L.K. Williams, *Shelling totally nonnegative flag varieties*, preprint 2005.
213. A. Yao, *Monotone bipartite graph properties are evasive*, SIAM J. Comput. **17** (1988), 517–520.
214. S. Yuzvinsky, *Small rational model of subspace complement*, Trans. AMS **354** (2002), 1921–1945.
215. T. Zaslavsky, *Facing up to arrangements: Face-count formulas for partitions of space by hyperplanes*, Memoirs AMS **154** (1975).
216. G.M. Ziegler, *Combinatorial models for subspace arrangements*, Habilitations-Schrift, Techn. Univ., Berlin 1992.
217. G.M. Ziegler, *Matroid shellability, β-systems, and affine hyperplane arrangements*, J. Algebraic Combin. **1** (1992), 283–300.
218. G.M. Ziegler, *On the difference between real and complex arrangements*, Math. Zeitschrift **212** (1993), 1–11.
219. G.M. Ziegler, *Shellability of chessboard complexes*, Israel J. Math. **87** (1994), 97–110.
220. G.M. Ziegler, *Lectures on Polytopes*, Grad. Texts in Math. **152**, Springer-Verlag, New York 1995.
221. G.M. Ziegler, *Shelling polyhedral 3-balls and 4-polytopes*, Discrete Compute. Geom. **19** (1998), 159–174.

222. G.M. Ziegler and R.T. Živaljević, *Homotopy type of arrangements via diagrams of spaces*, Math. Ann. **295** (1993), 527–548.
223. R.T. Živaljević and S.T. Vrećica, *The colored Tverberg problem and complexes of injective functions*, J. Combin. Theory (A) **61** (1992), 309–318.

Convex Polytopes: Extremal Constructions and f-Vector Shapes

Günter M. Ziegler

IAS/Park City Mathematics Series
Volume 14, 2004

Convex Polytopes: Extremal Constructions and f-Vector Shapes

Günter M. Ziegler

Introduction

These lecture notes treat some current aspects of two closely interrelated topics from the theory of convex polytopes: the shapes of f-vectors, and extremal constructions.

The study of *f-vectors* has had huge successes in the last forty years. The most fundamental one is undoubtedly the "g-theorem," conjectured by McMullen in 1971 and proved by Billera & Lee and Stanley in 1980, which characterizes the f-vectors of simplicial and of simple polytopes combinatorially. See also Section 5.2 of Forman's article in this volume, where h-vectors are discussed in connection with the Charney–Davis conjecture. Nevertheless, on some fundamental problems embarassingly little progress was made; one notable such problem concerns the shapes of f-vectors of 4-polytopes.

A number of striking and fascinating polytope *constructions* have been proposed and analyzed over the years. In particular, the Billera–Lee construction produces "all possible f-vectors" of simplicial polytopes. Less visible progress was made outside the range of simple or simplicial polytopes — where our measure of progress is that new polytopes "with interesting f-vectors" should be produced. Thus, still "it seems that overall, we are short of examples. The methods for coming up with useful examples in mathematics (or counterexamples to commonly believed conjectures) are even less clear than the methods for proving mathematical statements" (Gil Kalai, 2000).

These lecture notes are meant to display a fruitful interplay of these two areas of study: The discussion of f-vector shapes suggests the notion of "extremal" polytopes, that is, of polytopes with "extremal f-vector shapes." Our choice of constructions to be discussed here is guided by this: We will be looking at constructions that produce interesting f-vector shapes.

[1]Institute of Mathematics, MA 6–2, TU Berlin, D-10623 Berlin, Germany.
E-mail address: `ziegler@math.tu-berlin.de`.
Partially supported by the Deutsche Forschungs-Gemeinschaft (DFG), via the Research Center Matheon "Mathematics in the Key Technologies", the Research Groups "Algorithms, Structure, Randomness" and "Polyhedral Surfaces", and a Leibniz grant.

©2007 Günter M. Ziegler

After treating 3-polytopes in the first lecture and the f-vector shapes of very high-dimensional polytopes in the second one, we will start to analyze the case of 4-dimensional polytopes in detail. Thus the third lecture will explain a surprisingly simple construction for 2-simple 2-simplicial 4-polytopes, which have symmetric f-vectors. Lecture four sketches the geometry of the cone of f-vectors for 4-polytopes, and thus identifies the existence/construction of 4-polytopes of high "fatness" as a key problem. In this direction, the last lecture presents a very recent construction of "projected products of polygons," whose fatness reaches $9 - \varepsilon$. This shows that, on the topic of f-vectors of 4-polytopes, there is a narrowing gap between "the constraints we know" and "the examples we can construct."

Sources and Acknowledgements

The main sources, and the basis for the presentation in these lecture notes, are as follows. The proof of Steinitz' theorem in Lecture 1 is due to Alexander Bobenko and Boris Springborn [**17**]. A detailed report about the work by Ludwig Danzer, Anders Björner, Carl Lee, Jürgen Eckhoff and many others (Lecture 2) appears in [**79**, Sect. 8.6]; see also [**13**] and [**15**]. The "deep vertex truncation" construction presented in Lecture 3 is from joint work with Andreas Paffenholz [**57**]; my understanding of 2-simple 2-simplicial polytopes benefits also from my previous work with David Eppstein and Greg Kuperberg [**26**]. Lecture 4 draws heavily on my 2002 Beijing ICM report [**80**], which was relying on previous studies by Marge Bayer [**11**], and with Andrea Höppner [**42**]. Finally, the construction presented in Lecture 5 was announced in [**81**]; the intuition for it was built in previous joint work with Nina Amenta [**4**] and Michael Joswig [**43**].

Nikolaus Witte and Thilo Schröder have forcefully directed the problem sessions for my Utah lectures, and suggested a number of exercises. Nikolaus Witte has prepared many, and the nicest, figures for these notes. The construction of many of the examples and the beautiful Schlegel diagram graphics are based on the `polymake` system by Ewgenij Gawilow and Michael Joswig [**31**] [**32**] [**33**], which everyone is invited and recommended to try out, and use. An introduction to `polymake`, by Nikolaus Witte and Thilo Schröder, appears as an appendix, pp. 681–685.

I am grateful to all these colleagues for their work, for their explanations and critical comments, and for support on these lectures as well as in general. I have benefitted a lot from the lively discussions at the PCMI after my lectures, and from the many interesting questions and diverse feedback I got. Boris Springborn, Nikolaus Witte, Günter Rote, and many others provided very helpful comments on the draft version of these lecture notes. Thank you all!

More than usually, for the trip to Utah I have depended on the support, care, and love of Torsten Heldmann. Without him, I wouldn't have been able to go.

LECTURE 1
Constructing 3-Dimensional Polytopes

All the *polytopes* considered in these lecture notes are convex. A *d-polytope* is a d-dimensional polytope; thus the 3-dimensional polytopes to be discussed in this lecture are plainly 3-*polytopes*.*

How many 3-dimensional polytopes "do we know"? When pressed for examples, we will perhaps start with the platonic solids: the regular tetrahedron, cube and octahedron, icosahedron and dodecahedron.

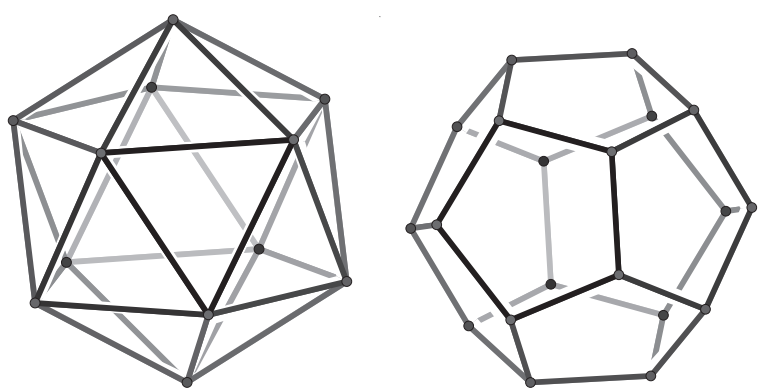


Figure 1.1. The regular icosahedron and dodecahedron

The classes of *stacked* and *cyclic* polytopes are of great importance for high-dimensional polytope theory because of their extremal f-vectors (according to the *lower bound theorem* and the *upper bound theorem*): Stacked polytopes arise from a simplex by repeatedly stacking pyramids onto the facets (cf. Lecture 2); cyclic polytopes are constructed as the convex hull of $n > d$ points on a curve of order d. However, neither of these constructions produces particularly impressive objects in dimension 3 (compare Figure 1.2, and Exercise 1.8).

The same must be said about pyramids and bipyramids over n-gons ($n \geq 3$) — see Figure 1.3.

*We assume that the readers are familiar with the basic terminology and discrete geometric concepts; see e.g. [**79**, Lect. 0] or [**40**].

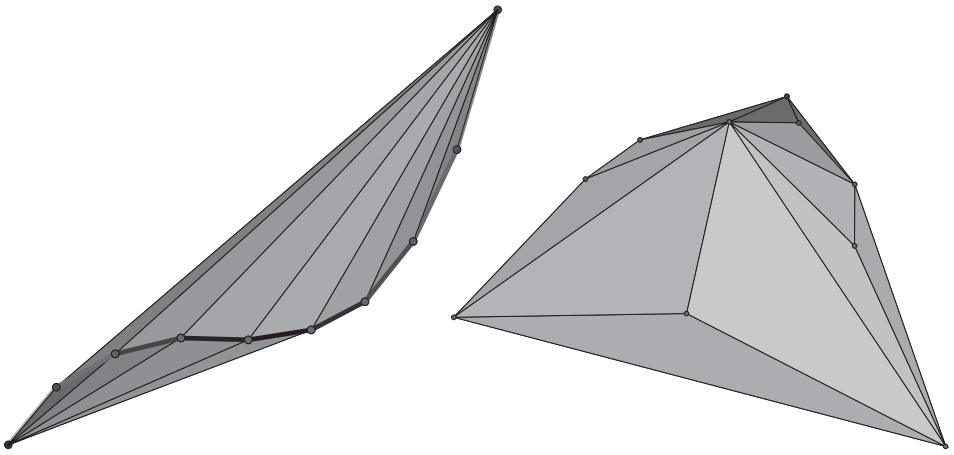


Figure 1.2. A cyclic 3-polytope $C_3(10)$ and a stacked 3-polytope, with 10 vertices each

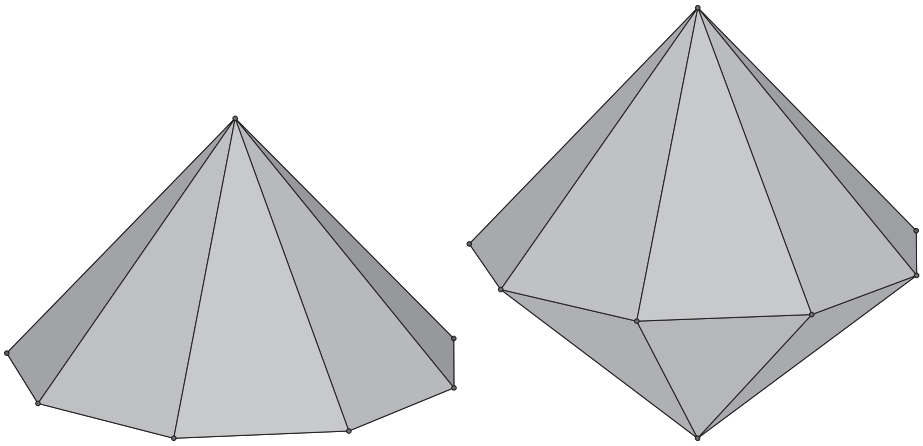


Figure 1.3. The pyramid and the bipyramid over a regular 10-gon

How do we get a "random" 3-polytope with lots of vertices? An obvious thing to look at is the convex hull of n random points on a 2-sphere.

Why is this not satisfactory? First, it produces only simplicial polytopes (with probability 1), and secondly it does not even produce all possible combinatorial types of simplicial 3-polytopes — see [**39**, Sect. 13.5]. It is a quite non-trivial problem to randomly produce all combinatorial types of polytopes of specified size (say, with a given number of edges). With the Steinitz theorem discussed below this reduces to a search for a random planar 3-connected graph with a given number of edges, say. See Schaeffer [**64**] for a recent treatment of this problem.

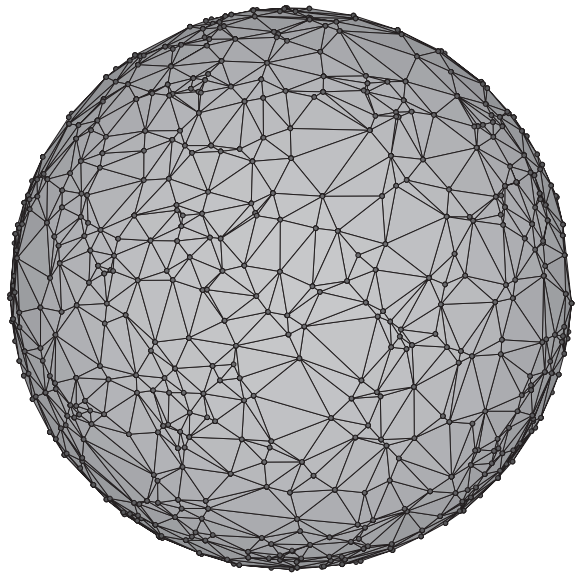


Figure 1.4. A random 3-polytope, with 1000 vertices on a sphere

1.1. The Cone of f-vectors

The f-*vector* of a 3-polytope P is the triplet of integers

$$f(P) \;=\; (f_0, f_1, f_2) \;\in\; \mathbb{Z}^3,$$

where f_0 is the number of vertices, f_1 is the number of edges, and f_2 denotes the number of facets (2-dimensional faces). In view of Euler's equation $f_0 - f_1 + f_2 = 2$ (which we take for granted here; but see Federico [**27**], Eppstein [**25**], and [**2**, Chap. 11]), the set of all f-vectors of 3-polytopes,

$$\mathcal{F}_3 \;:= \{(f_0, f_1, f_2) \;\in\; \mathbb{Z}^3 : f(P) = (f_0, f_1, f_2) \text{ is the } f\text{-vector of a 3-polytope } P\}$$

is a 2-dimensional set. Thus $\mathcal{F}_3$ is faithfully represented by the (f_0, f_2)-pairs of 3-polytopes,

$$\bar{\mathcal{F}}_3 \;:= \{(f_0, f_2) \;\in\; \mathbb{Z}^3 : f(P) = (f_0, f_1, f_2) \text{ for some 3-polytope } P\},$$

as shown in Figure 1.5: The missing f_1-component is given by $f_1 = f_0 + f_2 - 2$.

The set of all f-vectors of 3-polytopes was completely characterized by a young Privatdozent at the Technische Hochschule Berlin-Charlottenburg (now TU Berlin), Ernst Steinitz, in 1906: In a simple two-and-a-half-page paper he obtained the following result, whose proof we leave to you (Exercise 1.3).

Lemma 1.1 (Steinitz' lemma [**73**]). *The set of all f-vectors of 3-polytopes is given by*

$$\mathcal{F}_3 \;:= \{(f_0, f_1, f_2) \;\in\; \mathbb{Z}^3 : f_0 - f_1 + f_2 = 2,\; f_2 \le 2f_0 - 4,\; f_0 \le 2f_2 - 4\}.$$

This answer to the f-vector problem for 3-polytopes is remarkably simple: $\mathcal{F}_3$ is the set of *all* integral points in a 2-dimensional convex polyhedral cone. The three constraints that define the cone have clear interpretations: They are the Euler equation $f_0 - f_1 + f_2 = 2$, the upper bound inequality $f_2 \le 2f_0 - 4$,

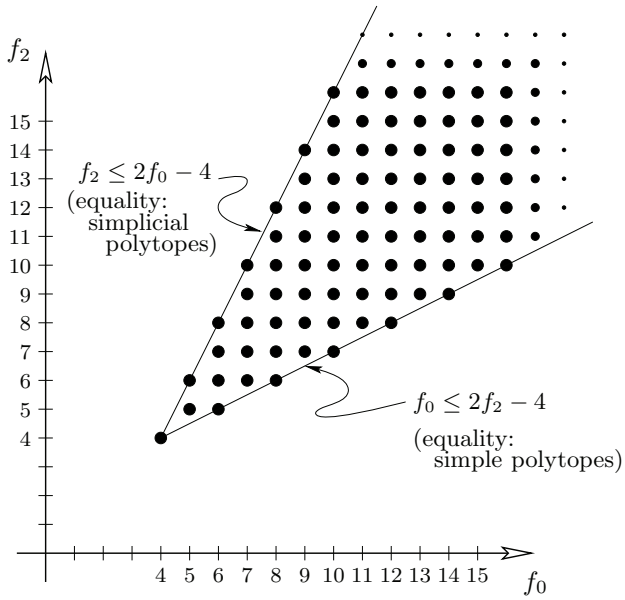


Figure 1.5. The set $\bar{\mathcal{F}}_3$, according to Steinitz' Lemma 1.1

which is tight exactly for the f-vectors of simplicial polytopes, and its dual, $f_0 \leq 2f_2 - 4$, which in the case of equality characterizes the f-vectors of simple 3-polytopes.

For the centennial of Steinitz' lemma, in 2006, let's strive for a characterization of the *cone* spanned by the f-vectors of 4-dimensional polytopes, cone($\mathcal{F}_4$). As we will see at the beginning of Lecture 4, this is a much more modest goal than a characterization of $\mathcal{F}_4$, which is not the set of all integral points in a convex set: It has "concavities" and even "holes."

Steinitz' lemma, as graphed in Figure 1.5, also shows that all (f-vectors of) convex 3-polytopes lie between the extremes of simple and of simplicial polytopes. And indeed, there seems to be the misconception that an analogous statement should be true in higher dimensions as well — it isn't. As we will see, there are additional interesting extreme cases in dimension 4, which are by far not as well understood as the simple and simplicial cases.

For any 3-polytope that is not a simplex, we may compute the "slope"

$$\phi(P) \;:=\; \frac{f_2 - 4}{f_0 - 4}$$

it generates in the graph of Figure 1.5, with respect to the apex $(4, 4)$ of the cone, which corresponds to a simplex. This slope satisfies

$$\tfrac{1}{2} \;\leq\; \phi(P) \;\leq\; 2,$$

where the lower bound characterizes simple polytopes, while the upper bound is tight for simplicial polytopes. Another interpretation of the parameter ϕ is that it is a homogeneous coordinate for the cone, where the denominator $f_0 - 4$ measures the "size" of the f-vector. (ϕ is homogeneous, so it yields $\frac{0}{0}$ for the f-vector of a simplex, which is the apex of the cone. Compare Exercise 1.5.)

1.2. The Steinitz Theorem

While Steinitz' lemma from 1906 is a very simple result, his theorem from 1922, characterizing the graphs of 3-polytopes, is substantial and deep. He knew that: He called it the "Fundamentalsatz der konvexen Typen," the fundamental theorem of convex types. Here is an informal version of it.

Theorem 1.2 (Steinitz' theorem [**74, 75**]). *There is a bijection*

$$\{3\text{-}\textit{connected planar graphs}\} \quad \longleftrightarrow \quad \{\textit{combinatorial types of } 3\text{-}\textit{polytopes}\}.$$

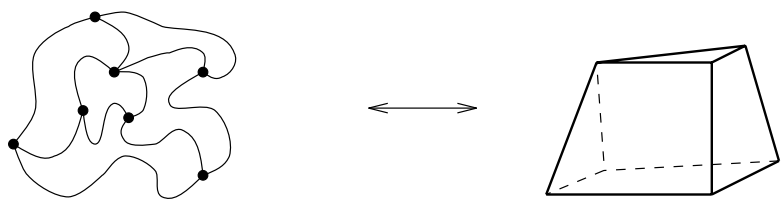


Figure 1.6. Graphs $\longleftrightarrow$ polytopes, according to Steinitz' Theorem 1.2

The direction "$\longleftarrow$" of Steinitz' theorem is not hard to establish.

Indeed, we do get a graph for any 3-polytope, namely the abstract graph whose nodes are the vertices of the polytope, and whose arcs are given by the edges of the polytope. This graph is indeed planar: To see this, one may first produce a radial projection of the polytope boundary (and thus of the vertices and edges) onto a sphere that contains the polytope, and then apply a stereographic projection [**41**, §36] to the plane. Or one may directly generate the "Schlegel diagram" and thus a straight-edge drawing of the graph in the plane. (In Lecture 3 we will see more of this tool, which shows its true power in the visualization of 4-polytopes.)

To see that the graph of any 3-polytope is 3-connected is also easy, using Menger's characterization of a d-connected graph as a graph that cannot be disconnected by removing or blocking less than d of its vertices. A powerful extension of this result is Balinski's theorem [**6**] [**79**, Thm. 3.14], that the graph of any d-polytope is d-connected.

Thus the hard and interesting part of Steinitz' theorem is the direction "$\longrightarrow$." It poses a non-trivial *construction problem*: To produce a convex 3-polytope with a prescribed graph (a geometric object) from an abstract planar graph (that is, from purely combinatorial data).

The first (easy) step for this is to convince oneself that the graph characterizes the complete combinatorial structure of the polytope. This follows from the simple observation (due to Whitney) that the faces of the polytope correspond exactly to the *non-separating induced cycles* in the graph.

Thus we have to construct convex 3-polytopes with prescribed combinatorics (face lattice), as given by a 3-connected planar graph. The importance of this step may be seen from the fact that three completely different types of proofs (and construction methods!) have been designed for it: Let's call them Steinitz type proofs, Tutte–Maxwell type proofs, and Koebe–Thurston type proofs.

Steinitz type proofs. Such proofs (of which Steinitz gave details on one in [**74**], and three are given in the Steinitz–Rademacher book [**75**] that appeared after Steinitz' death), are based on the following principle. Any planar 3-connected graph can be "reduced" to the complete graph K_4 by local operations, which yields a sequence

$$G = G_0 \;\to\; G_1 \;\to\; G_2 \;\to\; \ldots \;\to\; G_{N-1} \;\to\; G_N = K_4.$$

of 3-connected planar graphs.

This reduction sequence should then be reversed: Starting with a simplex Δ_3 (with graph K_4) we build up a sequence of polytopes,

$$P = P_0 \;\leftarrow\; P_1 \;\leftarrow\; P_2 \;\leftarrow\; \ldots \;\leftarrow\; P_{N-1} \;\leftarrow\; P_N = \Delta_3,$$

where P_i is a 3-polytope with graph G_i, again by simple/local construction steps.

Such a proof is presented in detail in [**79**, Lect. 4], so there is no need to do this here. We just mention that a number of interesting extensions and corollaries may be derived from Steinitz type proofs. Indeed, Barnette & Grünbaum [**9**] proved that in the construction of the polytope P, the shape of one face of the polytope may be prescribed. For example, some hexagon face may be required to be a regular hexagon, which imposes a non-trivial additional constraint. Similarly, Barnette [**7**] proved with a Steinitz type argument that a "shadow boundary" may be prescribed: P may be constructed in such a way that from some view-point outside the polytope, the edges that bound the visible part of the surface of the polytope correspond to a prescribed simple cycle in the graph of the polytope (which need not be induced). Equivalently, we may construct $P \subset \mathbb{R}^3$ so that the image $\pi(P)$ of P under the orthogonal projection $\pi : \mathbb{R}^3 \to \mathbb{R}^2$ is a polygon whose edges are given exactly by the edges of P that realize the prescribed cycle. Indeed, the edges must be "strictly preserved" by the projection, in the terminology that we will develop and use in Lecture 5.

Tutte–Maxwell type proofs. The Tutte–Maxwell approach to realizing 3-polytopes works in two stages: First one gets a "correct" drawing of the graph in the plane, then this drawing is lifted to 3-space.

For the first stage, one may assume that the graph contains a triangle face (if not, one dualizes; see Exercise 1.1). Then the vertices of this triangle are fixed in the plane, the edges are interpreted as ideal rubber bands, and the other vertices are placed according to the unique and easy-to-compute energy minimum, for which the sum of all squared edge lengths is minimal. This produces a correct, planar drawing of the graph without intersections — this is the (non-trivial) claim of Tutte's (1963) "rubber band method" [**77**]; moreover, any such drawing can be lifted to three-space according to Maxwell–Cremona theory, which may be traced back to work by Maxwell [**49**] nearly one hundred years earlier (1864). We refer to Richter-Gebert [**60**, Sect. 13.1] for a modern treatment, with all the proofs.

The Tutte–Maxwell proofs also buy us non-trivial corollaries: Indeed, each combinatorial type of 3-polytope can be realized with rational coordinates, and thus even with integral vertex coordinates (by clearing denominators). One can derive from a Tutte–Maxwell proof that singly-exponential vertex coordinates suffice for this: After a number of improvements on the original estimates by Onn & Sturmfels [**52**] we now know that each type of an n vertex 3-polytope with a triangle face can be represented with vertex coordinates in $\{0, 1, 2, \ldots, \lfloor 28.45^n \rfloor\}$ (see [**72**], [**61**] and [**59**]). It is not clear whether polynomial-size vertex coordinates can be achieved.

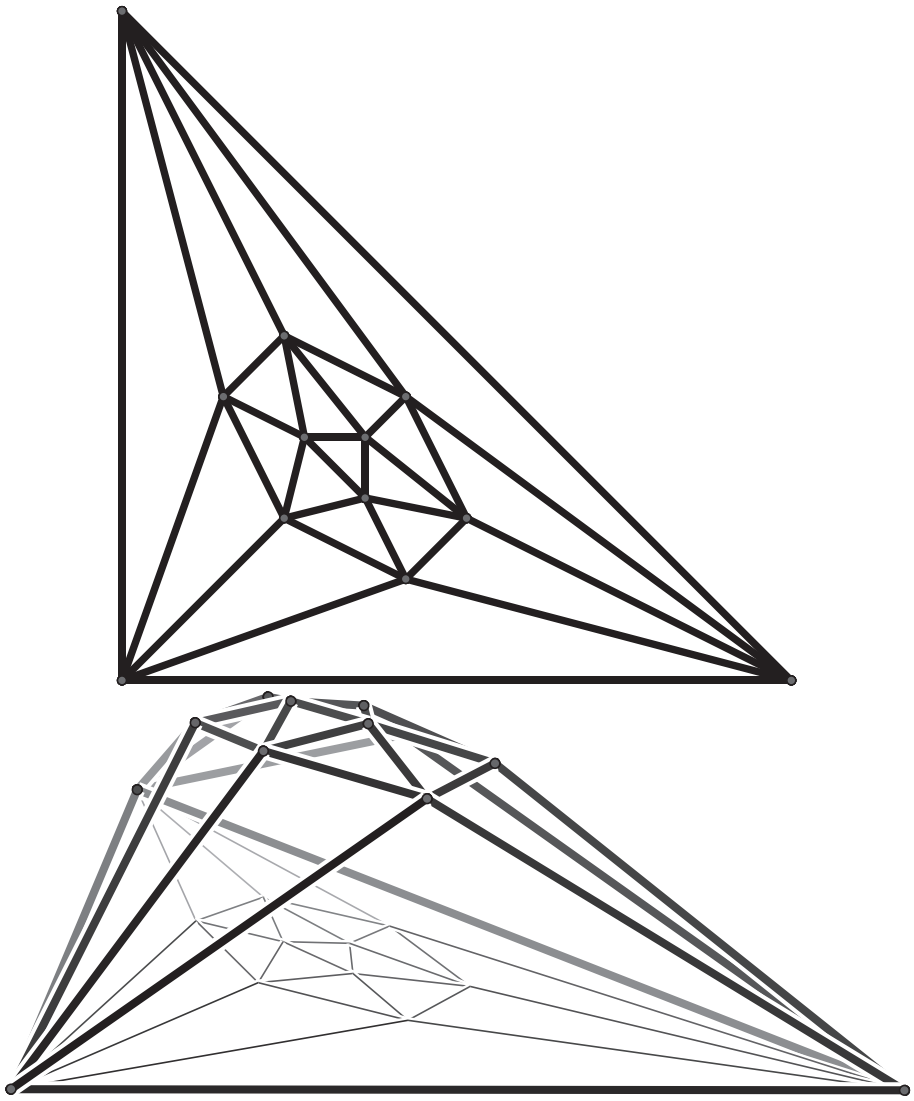


Figure 1.7. A Tutte drawing of the icosahedron graph, and the corresponding Maxwell–Cremona lifting

Koebe–Thurston type proofs. Geometric realizations of 3-polytopes *with all edges tangent to the sphere* may be derived from planar circle packings. Moreover, such a representation is essentially unique.

This seems to be essentially due to Bill Thurston [**76**] — who traces it back to Paul Koebe's [**46**] work on complex functions, and to work by E. M. Andreev [**5**] from the sixties on hyperbolic polyhedra. Thurston's insight was followed up, explained, extended and generalized by a number of authors. Pach & Agarwal [**53**, Chap. 8] describe the "standard" proof, based on a (non-constructive) fixed point argument. However, Mohar [**50**] described an effective construction algorithm, and Colin de Verdière [**20**] was the first to prove that the circle packings in question can

be derived from a variational principle (that is, an energy functional). In this line of work, Bobenko & Springborn [**17**] have quite recently discovered an explicit, elegant and quite general variational principle for the construction of circle patterns with prescribed intersection angles. In the following, we prove the Steinitz theorem based on their functional — taking advantage of all the simplifications that occur in their proof and formulas if one wants to "just" get the orthogonal circle patterns needed for the Steinitz theorem. (See also Springborn [**68**] for an additional discussion of uniqueness.)

1.3. Steinitz' Theorem via Circle Packings

Theorem 1.3 (The Koebe–Andreev–Thurston theorem)**.** *Each* 3*-connected planar graph can be realized by a* 3*-polytope which has all edges tangent to the unit sphere.*

Moreover, this realization is unique up to Möbius transformations (projective transformations that fix the sphere). The edge-tangent realization for which the barycenter of the tangency points is the center of the sphere is unique up to orthogonal transformations.

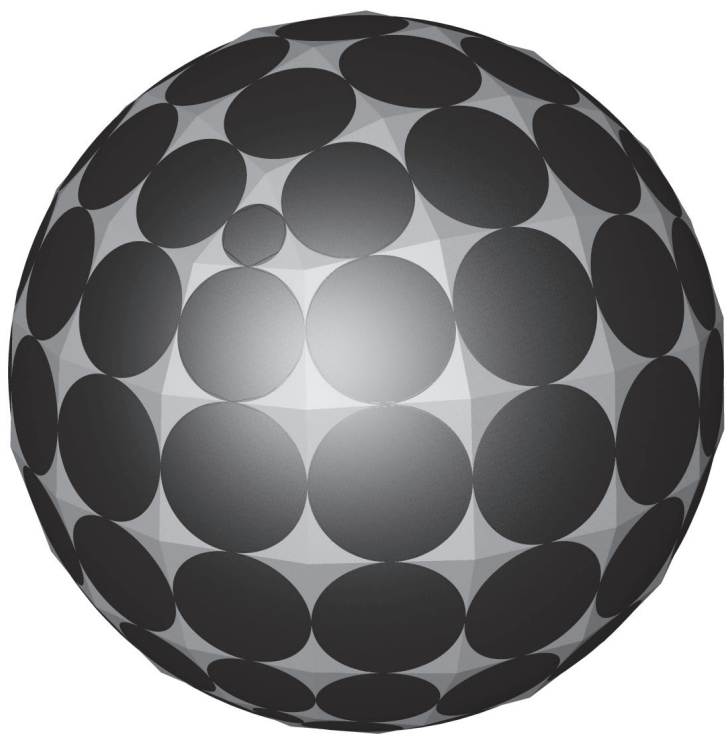
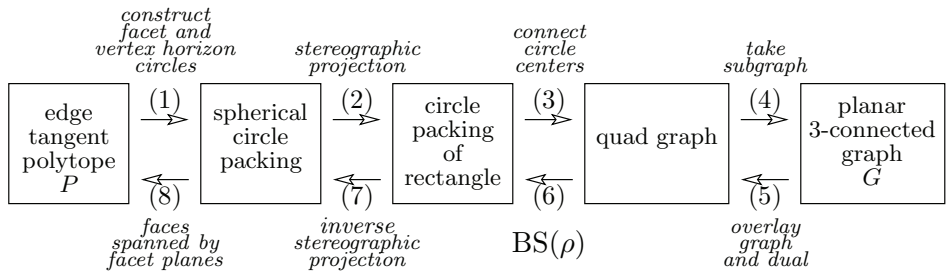


Figure 1.8. Edge-tangent representation of a polyhedron, according to the Koebe–Andreev–Thurston theorem [Graphics by Boris Springborn, Matheon]

In our presentation of the proof, we first explain how any edge-tangent representation of a polytope P induces a circle pattern on the sphere, which in turn yields

a planar circle pattern, and the combinatorics of the planar circle pattern yields a quad graph (a planar graph whose faces are quadrilaterals), which has $G(P)$ as a subdivided subgraph. This yields steps (1) to (4) in the following scheme:



Our plan is to then reverse this four-step process, in order to construct an edge-tangent polytope from the given graph G. In step (5), the quad graph is derived directly from the graph $G = G(P)$, by superposing the graph with its dual. Then, in step (6), we construct the rectangular circle pattern with the combinatorics of the quad graph, and then proceed to construct P from it.

The steps (5), (7), and (8) are quite straightforward: The key, non-trivial step is (6), the construction of the (unique) rectangular circle pattern, which we achieve via the "euclidean Bobenko–Springborn functional."

Proof. We start with a detailed description of the four-step process from edge-tangent polytopes to planar 3-connected graphs, via circle packings and quad graphs.

(1). Assume that $P \subset \mathbb{R}^3$ is a 3-polytope whose edges are tangent to the unit sphere $S^2 \subset \mathbb{R}^3$. Then the facet planes of P intersect the unit sphere S^2 in circles that we call the *facet circles*: We get one circle for each facet, and the circles are disjoint, but they touch exactly if the corresponding facets are adjacent. We also get a second set of circles which we call the *vertex horizon circles*: Each such circle is the boundary of the spherical cap consisting of all the points on the sphere that are "visible" from the respective vertex. We get one vertex horizon circle for each vertex, and the circles are disjoint, but they touch exactly if the corresponding vertices are adjacent.

Moreover, at each edge tangency point, the two touching facet circles and the two touching vertex horizon circles intersect orthogonally; see Figure 1.9 for an example. (The vertex horizon circles of P are the facet circles of the dual polytope P^*, whose edges have the same tangency points as the edges of P; the facet circles for P are also the vertex horizon circles for P^*; corresponding edges $e \subset P$ and $e^* \subset P^*$ intersect orthogonally at the respective tangency point.)

(2). We perform a stereographic projection to the plane, using one of the edge tangency points p_0 as the projection center, and mapping all the facet and vertex horizon circles to the equator plane corresponding to the projection point. In the resulting planar figure, the two facet circles through p_0 yield two parallel lines (and after a rotation we may assume that these are horizontal); the two vertex horizon circles through p_0 also yield two parallel lines, orthogonal to the first two (and thus

vertical). So we get a planar pattern that consists of four lines bounding an axis-parallel rectangle, and circles that touch resp. intersect orthogonally in the plane. This is the *rectangular circle pattern*.

If the faces adjacent to the edge f through p_0 are an h_1-gon and an h_2-gon, then we get $h_1 - 2$ resp. $h_2 - 2$ circles along the horizontal edges of the rectangle. Similarly, if the end vertices of f have degrees v_1 and v_2, then we get $v_1 - 2$ resp. $v_2 - 2$ circles along the vertical edges of the rectangle. The example that one obtains from the cube (Figure 1.9) is displayed in Figure 1.10.

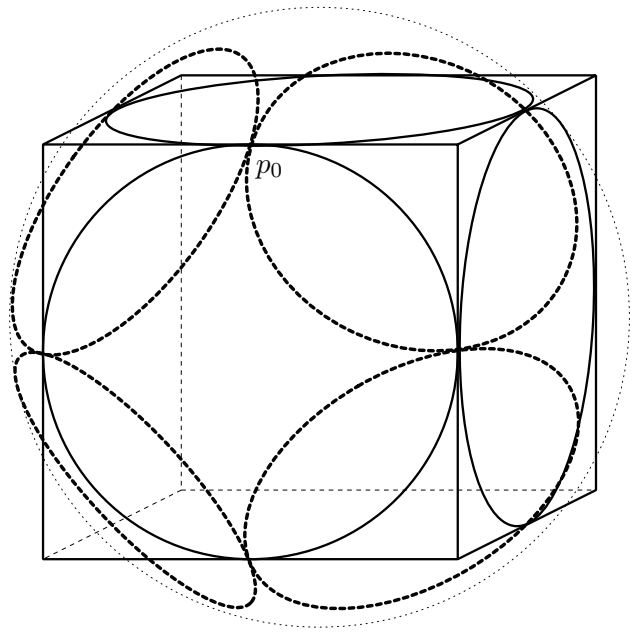


Figure 1.9. The facet circles and the vertex horizon circles (dashed) for an edge-tangent representation of a regular cube.

(3). Any rectangular circle pattern yields a quad graph drawing as follows: The vertex set consists of the centers of all the circles, with four additional vertices "far out" representing the four lines that bound the rectangles (as in Figure 1.11). We obtain drawings of both G and G^* by connecting the centers of touching facet circles resp. vertex horizon circles. This includes one horizontal edge f of G "going through infinity," while dual graph G^* has the corresponding edge f^* going through infinity vertically.

From the rectangular circle pattern, we obtain a decomposition of a rectangle into quadrilaterals by connecting the centers of adjacent facet circles, and the centers of adjacent vertex horizon circles. See the example of Figure 1.11, where the rectangle is shaded. The graph of this rectangle decomposition is the *quad graph*: Its vertices correspond to (the centers of) the facet circles that don't contain p_0, the vertex horizon circles that don't contain p_0, and intersection points of edges e and e^* of G and G^*, other than the edges f, f^* that contain p_0.

(4). In particular, the graph G may be derived from the quad graph, by "deleting the dashed edges."

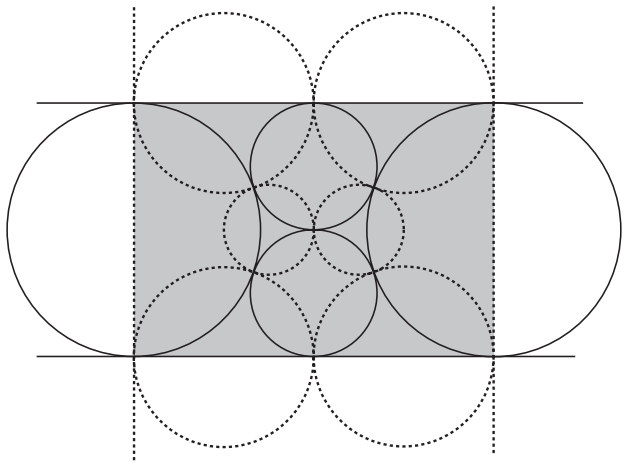


Figure 1.10. The rectangular circle pattern derived from an edge-tangent 3-cube (with $h_1 = h_2 = 4$, $v_1 = v_2 = 3$)

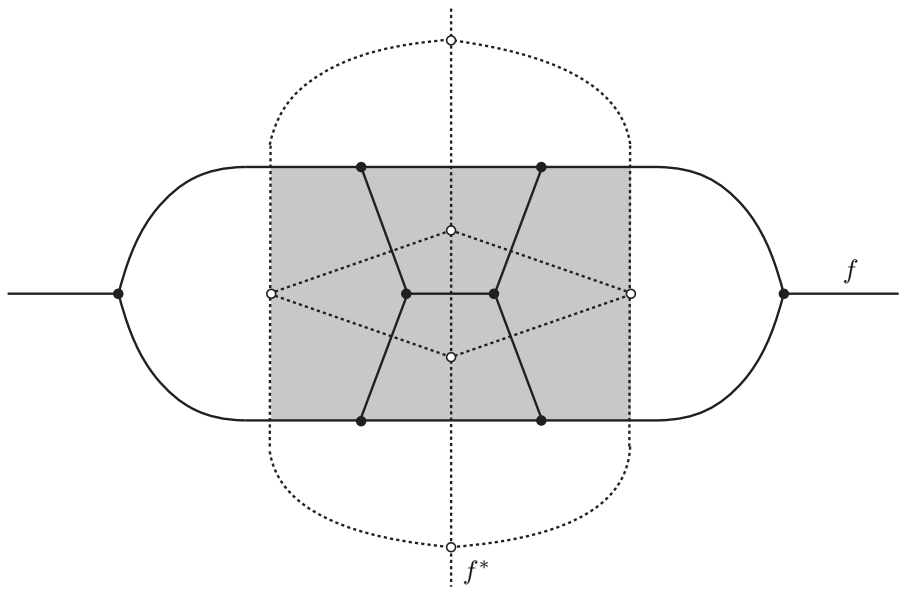


Figure 1.11. The quad graph for the cube, generated from Figure 1.10: The white vertices are given by the facet circle centers, while the black vertices correspond to the vertex horizon circles; the dashed edges connect the centers of adjacent facet circles, and the straight edges correspond to adjacent vertex horizon circles.

This ends the description of the passage from an edge tangent polytope to the planar graph drawing. Now we start the way back: Another four-step process leads us from graphs via quad graphs and circle patterns to edge-tangent 3-polytopes.

(5). The quad graph may be derived from knowledge of the graph G alone, plainly by overlaying G and G^*. For our cube example, the result may look like the drawing given in Figure 1.12.

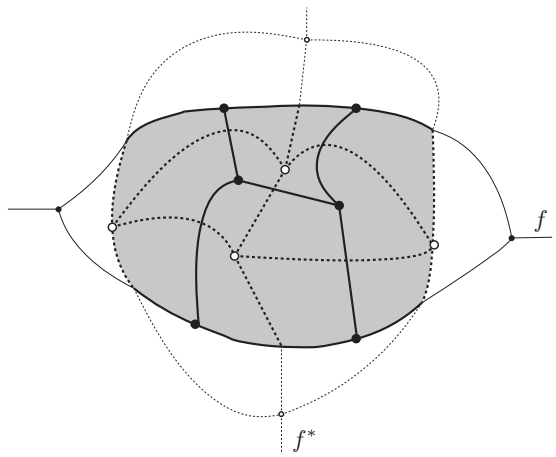


Figure 1.12. The quad graph for the cube, generated from an overlay of the cube graph (black edges) layed out with the edge f "at infinity" and the dual graph (dashed edges), with the dual edge f^* "at infinity." The shaded part defines the restricted quad graph.

The input for the next step will be the *restricted quad graph*: It is obtained from the full quad graph by deleting everything that is adjacent to the original edges f and f^*. Its bounded faces are quadrilaterals (*quads* for short), with two black and two dashed edges each. Each quad has

- a *black* vertex and a *white* vertex
  (the black vertex, where the two black edges meet, corresponds to the center of a face circle; the white one, where the two dashed edges meet, corresponds to the center of a horizon circle),
- and two more vertices where a black and a dashed edge meet
  (they correspond to edge tangency points).

For the following, we use I_0 as an indexing set for the black and white vertices in the restricted quad graph. It is in bijection with the vertices of G and of G^*, except for the vertices of the edges f and f^*, which yield lines rather than circles. That is, we have

$$I_0 \;:=\; V(G-f) \cup V(G^*-f^*).$$

The following step, which takes us from combinatorics (a graph drawing) to geometry (a circle pattern), is the crucial one.

(6). In the "correct" realization of the restricted quad graph, which would yield a circle packing, each quad is drawn as a *kite* in which

- the two black edges have the same length
  (radius r_i of the corresponding vertex horizon circle),
- the two dashed edges have the same length
  (radius r_j of the corresponding facet circle),
- and there are two right angles between black and dashed edges
  (where facet and vertex horizon circles intersect).

The kites have to look like the one in Figure 1.13.

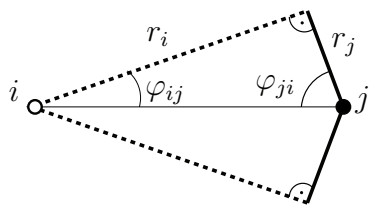


Figure 1.13. A kite, with radii $r_i = e^{\rho_i}$, $r_j = e^{\rho_j}$, and angles φ_{ij} and φ_{ji}

Hence, we have to solve the following construction problem:

> Given a quad graph decomposition of a rectangle, derived from the overlay of a 3-connected planar graph G and its dual G^*, construct a geometric drawing, with straight edges, as a kite decomposition of a rectangle.

The kites are completely determined if we know their edge lengths: If the edge lengths in a kite are $r_i, r_j > 0$, then the angles are given by

$$\varphi_{ij} \;=\; \arctan\left(\frac{r_j}{r_i}\right) \quad \text{and} \quad \varphi_{ij} \;=\; \arctan\left(\frac{r_i}{r_j}\right),$$

with $\varphi_{ij} + \varphi_{ji} = \frac{\pi}{2}$ (see Figure 1.13). Thus all we have to do is to determine radii r_i corresponding to the black and white vertices of the quad graph, such that the following system of equations is satisfied:

$$\boxed{\sum_{j\,:\,i \diamond\!\!\!\!\bullet\!\!\!\!\circ\, j} 2\arctan\left(\frac{r_j}{r_i}\right) \;=\; \Phi_i \qquad \text{for all vertices } i \in I_0,} \tag{1.1}$$

where the right-hand-sides are given by

$$\Phi_i \;:=\; \begin{cases} \pi & \text{if } i \text{ is on the boundary,} \\ 2\pi & \text{if } i \text{ is in the interior.} \end{cases}$$

In the equation whose right hand side is Φ_i, the sum on the left hand side is taken over all vertices $j \in I_0$ that are opposite to i in one of the kites. (If i is a white vertex, then j will be black, and vice versa.)

Indeed, if (1.1) is satisfied, then we can easily construct the kites and piece them together to get a flat rectangle and the circle packing. Badly enough, (1.1) is a non-linear system of equations, which we have to solve in positive variables $r_i > 0$. We want to know that this has a solution, which is unique up to multiplying all the r_is with the same factor, and which can be computed efficiently. Luckily, we can do this, since the system is solved by minimizing an explicit and easy-to-write-down "energy" functional which will turn out to be convex, with a unique minimum. For this, we first do a change of variables,

$$\rho_i \;:=\; \log r_i.$$

Then we normalize by the condition $\prod_i r_i = 1$, that is,

$$\sum_i \rho_i \;=\; 0.$$

Furthermore, we define

$$f(x) \;:=\; \arctan(e^x).$$

This auxiliary function is graphed in Figure 1.14. Note that $f(-x) = \frac{\pi}{2} - f(x)$.

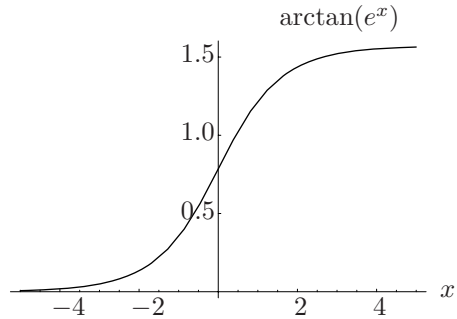


Figure 1.14. $f(x) = \arctan e^x$

We differentiate f,

$$f'(x) \;=\; \frac{1}{1+e^{2x}}e^x \;=\; \frac{1}{2\cosh x},$$

which yields $f'(-x) = f'(x) > 0$ for all $x \in \mathbb{R}$. We also integrate f, and define

$$F(x) \;:=\; \int_{-\infty}^{x} f(t)\mathrm{d}t.$$

This function satisfies $F(x) \geq 0$ for all x, but also $F(x) \geq \frac{\pi}{2}x$. Thus we get that

$$F(x) + F(-x) \;\geq\; \tfrac{\pi}{2}|x|. \tag{1.2}$$

The system (1.1) we have to solve may be rewritten in terms of $f(x)$ as

$$\sum_{i\,:\,i \diamond j} 2f(\rho_j - \rho_i) \;=\; \Phi_i \qquad \text{for all black or white vertices, } i \in I_0. \tag{1.3}$$

To solve this, Bobenko & Springborn [**17**] present the functional

$$\boxed{\mathrm{BS}(\rho) \;:=\; \sum_{i \diamond j} \Big\{F(\rho_j - \rho_i) + F(\rho_i - \rho_j) - \tfrac{\pi}{2}(\rho_i + \rho_j)\Big\} \;+\; \sum_{i \,\in\, I_0} \Phi_i \rho_i,} \tag{1.4}$$

where the first sum is over all *unordered* pairs $\{i, j\}$ of vertices $i, j \in I_0$ that are opposite in one of the kites. The claim is now that

(A) the critical points of $\mathrm{BS}(\rho)$ are exactly the solutions to our system (1.3),
(B) the functional is convex: Restricted to $\sum_i \rho_i = 0$ it is strictly positive definite, so the critical point is unique if it exists, and
(C) the functional gets large if any of the differences $\rho_i - \rho_j$ gets large: Thus the functional must have a critical point (a minimum) — the solution we are looking for.

For (A), a simple computation yields the gradient of $\mathrm{BS}(\rho)$:

$$\frac{\partial \mathrm{BS}(\rho)}{\partial \rho_i} \;=\; \Phi_i - \sum_{i \diamond j} 2f(\rho_j - \rho_i).$$

Thus the critical points of $\mathrm{BS}(\rho)$ are exactly the solutions to (1.3).

For (B), we compute the Hessian (the matrix of second derivatives) for $\mathrm{BS}(\rho)$, and find that

$$x^T \mathrm{BS}(\rho)'' x \;=\; 2 \sum_{i \diamond j} f'(\rho_j - \rho_i)\,(x_j - x_i)^2.$$

We know that $f'(\rho_j - \rho_i) > 0$, so this quadratic form can vanish only if all the differences $x_j - x_i$ vanish for "adjacent" $i, j \in I_0$ (that is, for black/white vertices that share a kite). But the graph we consider is connected, so this implies that all variables x_i are equal. Restricted to $\sum_i x_i = 0$ this yields that all x_i vanish, so the Hessian is positive definite on the restriction hyperplane, and the solution we are striving for is unique if it exists.

To prove the existence claim (C), we have to find that $\mathrm{BS}(\rho)$ grows large if any difference of variables $\rho_k - \rho_i$ gets large. With the same argument we just used this implies that some difference of "adjacent" variables will become large. Then also $F(\rho_j - \rho_i) + F(\rho_i - \rho_j) \geq \frac{\pi}{2}|\rho_j - \rho_i|$ gets large, but it will grow only linearly in $|\rho_j - \rho_i|$, and it is not obvious that the growing positive terms in (1.4) will "outrun" the negative terms. This will require a careful "matching" between positive and negative terms.

To achieve this, we use the existence of a *coherent angle system*, that is, an assignment of angles $\varphi_{ij}, \varphi_{ji} > 0$ to the kites that satisfies the conditions

$$\varphi_{ij} + \varphi_{ji} \;=\; \tfrac{\pi}{2} \qquad \text{and} \qquad \sum_{j\,:\,i \diamond j} 2\varphi_{ij} \;=\; \Phi_i. \tag{1.5}$$

Any solution to (1.1) would give us a coherent angle system, but the existence of such a coherent angle system is much weaker, far from solving the system (1.1): If we have a coherent angle system, then we could construct kites from this — whose angles would fit together at the black and white vertices, but whose side lengths might not. (Compare Figure 1.15.)

For any coherent angle system, $\varepsilon_0 := \min_{k,\ell} \varphi_{k\ell}$ is a positive number.

If there is a coherent angle system, then the minimum exists. Let's assume for now that a coherent angle system exists (this will be proved below). Then

$$\begin{aligned}
\mathrm{BS}(\rho) \;&=\; \sum_{i \diamond j} \Big\{ F(\rho_j - \rho_i) + F(\rho_i - \rho_j) - \tfrac{\pi}{2}(\rho_i + \rho_j) \Big\} \;+\; \sum_i \Phi_i \rho_i \\
&\overset{\text{(i)}}{>} \sum_{i \diamond j} \Big\{ \tfrac{\pi}{2}\big|\rho_i - \rho_j\big| - \tfrac{\pi}{2}(\rho_i + \rho_j) \Big\} \qquad\qquad + \sum_i \Phi_i \rho_i \\
&\overset{\text{(ii)}}{=} \sum_{i \diamond j} \Big\{ \tfrac{\pi}{2}\big|\rho_i - \rho_j\big| - \tfrac{\pi}{2}(\rho_i + \rho_j) \Big\} \;+\; \sum_{i \diamond j} 2(\varphi_{ij}\rho_i + \varphi_{ji}\rho_j) \\
&\overset{\text{(iii)}}{=} \sum_{i \diamond j} -\pi \min\{\rho_i, \rho_j\} \qquad\qquad + \sum_{i \diamond j} 2(\varphi_{ij}\rho_i + \varphi_{ji}\rho_j) \\
&\overset{\text{(iv)}}{\geq} \sum_{i \diamond j} -\pi \min\{\rho_i, \rho_j\} \;+\; \sum_{i \diamond j} \pi \min\{\rho_i, \rho_j\} + 2\min\{\varphi_{ji}, \varphi_{ij}\}|\rho_i - \rho_j| \\
&=\; \sum_{i \diamond j} 2\min\{\varphi_{ji}, \varphi_{ij}\}|\rho_i - \rho_j| \;\geq\; 2\varepsilon_0 \sum_{i \diamond j} |\rho_i - \rho_j|.
\end{aligned}$$

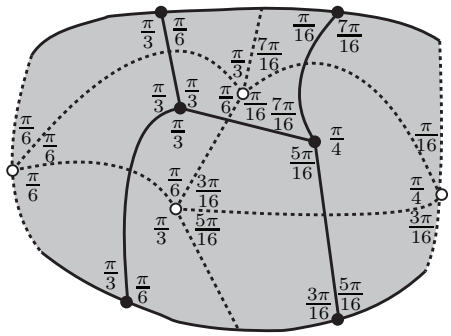


Figure 1.15. The assignment in this figure is a coherent angle system – but not one that corresponds to a correct circle pattern.
(Note that the construction of the coherent angle system proceeds from the plane graph without use of a straight edge drawing. In the figures further down we draw the graphs with straight edges for simplicity, but this structure is not used in the proof. Rather, it is produced by the proof.)

Here

- the estimate for (i) uses $F(x) + F(-x) \geq \frac{\pi}{2}|x|$, which is (1.2).
- (ii) is obtained by substituting (1.5). We need the second term in the second sum in (ii) since the sums over "$i \diamond j$" are sums over unordered pairs; there is no extra summand for "$j \diamond i$."
- (iii) follows from $|x - y| - (x + y) = -2\min\{x, y\}$,
- For (iv), in the case $\rho_j \geq \rho_i$ we compute

$$\begin{aligned} 2(\varphi_{ij}\rho_i + \varphi_{ji}\rho_j) &= \pi\rho_i - 2\varphi_{ji}\rho_i + 2\varphi_{ji}\rho_j \\ &= \pi\min\{\rho_i, \rho_j\} + 2\varphi_{ji}|\rho_i - \rho_j| \\ &\geq \pi\min\{\rho_i, \rho_j\} + 2\min\{\varphi_{ji}, \varphi_{ij}\}|\rho_i - \rho_j|, \end{aligned}$$

  and analogously for $\rho_i \geq \rho_j$.

We are dealing with a connected quad graph. Thus if the norm of the vector ρ gets large, while the sum of the ρ_i is zero, then also for two $i, j \in I_0$ in the same quadrilateral the difference $|\rho_i - \rho_j|$ gets large. Thus by the computation above, $\mathrm{BS}(\rho) > 2\varepsilon_0|\rho_i - \rho_j|$ gets large. This is sufficient to prove that the strictly convex function $\mathrm{BS}(\rho)$ does have a (unique) minimum — the solution to our problem.

A coherent angle system exists. Finally, we have to verify the existence of a coherent angle system. We will see here that via some simple network flow theory, this follows from an expansion property in the "diagonal graph" $D(G \cup G^*)$. After that, we will prove the expansion property.

Let G be a 3-connected planar graph, G^* its dual, both of them again drawn into the plane with dual edges f, f^* intersecting "at infinity." Then the *diagonal graph* $D = D(G \cup G^*)$ has the same vertex set as $G \cup G^*$. Its edges correspond to the diagonals in the quad graph given by $G \cup G^*$.

Equivalently, the diagonal graph D has black vertices corresponding to the vertices of G, and white vertices corresponding to the faces of G. The edges of D correspond to the vertex–face incidences of G. See Figure 1.16 for an example.

The *reduced diagonal graph* $D' = (V', E')$ is obtained from the diagonal graph $D = (V, E)$ by removing the two vertices of f, the two vertices of f^*, and the four

edges that connect them, but none of the others. So indeed, D' does have pending edges (half-edges) which have lost one of their end-vertices.* See Figure 1.17 for an example.

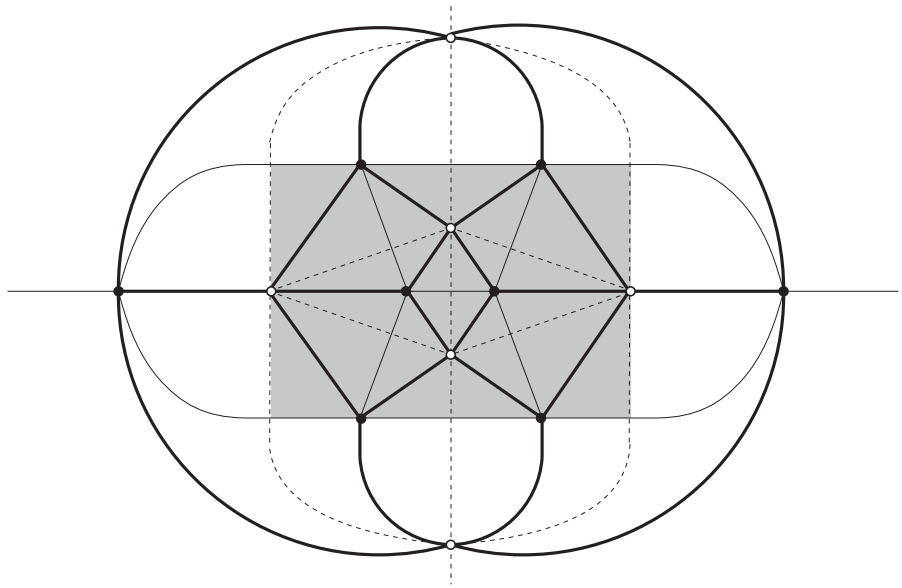


Figure 1.16. The diagonal graph $D = D(G \cup G^*)$, given by the fat edges, where G is the graph of the cube

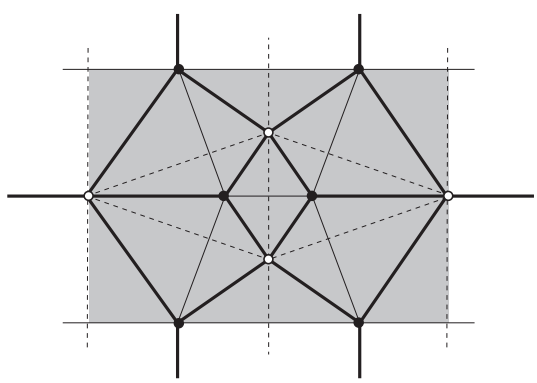


Figure 1.17. The fat edges in this figure display the reduced diagonal graph $D' = D'(G \cup G^*)$ in the case where G is the graph of the cube, derived from Figure 1.16. Note that the fat edges leaving the rectangle are included in D', their vertices at the other end are not. So in this example D' has 10 vertices and 20 edges, including 6 half-edges with only one end-vertex.

*I am sure you won't be troubled too much by the fact that this is not a graph in the usual technical sense, since it does have half-edges with only one end-point.

The diagonal graph $D = D(V, E)$ is a quad graph: All its faces, including the "unbounded" face (if we draw it in the plane) are quadrilaterals. From this, we get by double counting that $2|F| = |E|$ and thus $|V| = 2|E| - 4$ by Euler's relation. The reduced quad graph $D' = (V', E')$ has $|V'| = |V| - 4$ vertices and $|E'| = |E| - 4$ edges. Hence we get $|E'| = 2|V'|$: The reduced quad graph has exactly double as many edges as vertices.

The concept of a coherent angle system has a very nice interpretation in terms of the restricted diagonal graph: Each vertex v_i gets a weight of 2π, and this has to be distributed to the edges e incident to v_i such that

- each edge e incident to v_i gets a positive part of the weight 2π of v,
- all of the weight 2π of v_i is distributed to its incident edges, and
- the weights assigned to each edge sum to π.

Indeed, in such an assignment any half-edge clearly gets a weight of π from its only end-vertex, which corresponds to a boundary vertex of the restricted quad graph; thus the boundary vertex v_i distributes a weight of exactly $\Phi_i = \pi$ to its other incident edges, that is, to the (diagonals of the) kites it is incident to. The vertices of D' without an incident half-edge correspond to interior vertices v_j of the restricted quad graph, so they have a weight/angle of $\Phi_j = 2\pi$ to distribute to the incident edges/kites.

The "weight distribution problem" for the reduced diagonal graph $D' = (V', E')$ may also be interpreted as a flow problem (cf. [**1**]): We have to find a maximal flow, of weight $2\pi|V'| = \pi|E'|$, in a two-layer network as depicted in Figure 1.18. It consists of a source node s, then a layer of nodes formed by the vertex set V' of D', then a layer of nodes in bijection to the the edge set E', and then the sink node t. There are three groups of arcs: The arcs (s, v') emanating from the source all have an upper bound of 2π; the arcs of type (v', e'), where the edge e' is incident to v', get an upper bound of ∞, while the arcs at the sink, (e', t), have an upper bound of π.

We need a *positive* flow in this network; to get this, we put a small lower bound of $\varepsilon > 0$ on each edge of type (v', e'), and 0 on all other edges. There is a *feasible flow* in this network with upper and lower bounds on each edge: For this, let the flow value be ε on each (v', e')-arc, and a suitable multiple of ε on the other arcs.

We need a positive flow of value $2\pi|V'| = \pi|E'|$ in this network. There is a feasible flow, and no flow with a larger value than $2\pi|V'|$ can exist due to the cuts that separate s or t from the rest of the network. Thus we can apply the following generalization of the Max-Flow Min-Cut Theorem on network flows. (You should prove this yourself: See Exercise 1.9.)

Theorem 1.4 (Generalized Max-Flow Min-Cut Theorem; cf. [**1**, Sect. 6.7])**.** *If an (s, t)-network with lower and upper bounds has a feasible flow, then the value of a maximal (s, t)-flow is the capacity of a minimal (s, t)-cut.*

The *capacity* of an (s, t)-cut in a network with upper and lower bounds is the sum on the upper bounds of the forward arcs, minus the sum of the lower bounds on the backward arcs across the cut. So in our example the cuts $[\{s\}, V' \cup E' \cup \{t\}]$ and $[\{s\} \cup V' \cup E', \{t\}]$ have capacity $2\pi|V'| = \pi|E'|$. Could there be a cut of smaller capacity? Any (s, t)-cut is of the form

$$[\{s\} \cup V_1' \cup E_1',\, V_2' \cup E_2' \cup \{t\}]$$

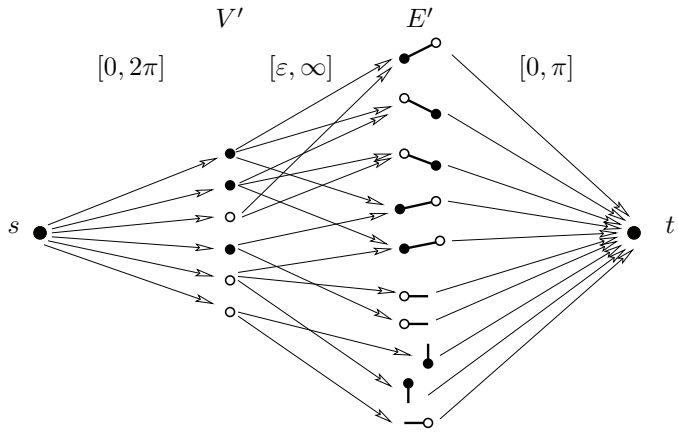


Figure 1.18. Construction of a coherent angle system from a network flow problem with lower and upper bounds, which are indicated by intervals like $[0, \pi]$.

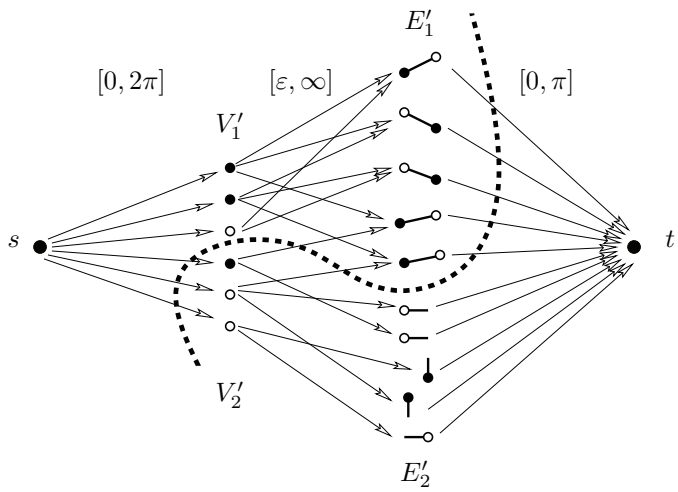


Figure 1.19. The dashed line indicates the cut $[\{s\} \cup V'_1 \cup E'_1,\ V'_2 \cup E'_2 \cup \{t\}]$ in our network

for partitions $V' = V'_1 \uplus V'_2$ and $E' = E'_1 \uplus E'_2$. Such a cut has finite capacity if there are no arcs (v', e') from V'_1 to E'_2; compare Figure 1.19. That is, we should take E'_1 to include all the edges that are incident to a vertex in V'_1.

The capacity of the cut $[\{s\} \cup V'_1 \cup E'_1,\ V'_2 \cup E'_2 \cup \{t\}]$ is

$$2\pi|V'_2| + \pi|E'_1| - \varepsilon|A(V'_2, E'_1)| \quad = \quad 2\pi|V'| - 2\pi|V'_1| + \pi|E'_1| - \varepsilon|A(V'_2, E'_1)|,$$

where $|A(V'_2, E'_1)|$ denotes the number of arcs from V'_2 to E'_1. For small enough ε, say $\varepsilon = 1/|A(V', E')|$, we have $\varepsilon|A(V'_2, E'_1)| < 1$. Thus the following "expansion property" for the diagonal graph implies that all cuts $[\{s\} \cup V'_1 \cup E'_1,\ V'_2 \cup E'_2 \cup \{t\}]$ have capacity larger than $2\pi|V'|$, except in the two trivial cases given as examples above, where the capacity is exactly $2\pi|V'|$. Thus the maximal flow, of value $2\pi|V'|$, exists; it is positive, and yields the coherent angle system.

Expansion in the diagonal graph. It remains to verify the following: Let $V_1' \subseteq V'$ be a set of vertices in the reduced diagonal graph $D'(G \cup G^*) = (V', E')$, and assume that $E_1' \subseteq E'$ includes all edges of D' that are incident to a vertex in V_1'. Then

$$|E_1'| \quad \geq \quad 2|V_1'|, \tag{1.6}$$

with equality only in the trivial cases $V_1' = \emptyset$ and $V_1' = V'$.

For this we may assume that the subgraph induced by V_1' is connected, because we can consider its components separately. We may also assume that $|V_1| \geq 2$, so V_1' contains both a black and a white vertex.

Now let U be an open subset of the plane (or of S^2) whose boundary curves separate V_1' from the $h+1$ components of the graph $D \setminus V_1'$, as illustrated in Figure 1.20. Topologically, U is an open disk with $h \geq 0$ holes.

The diagonal graph yields a cell decomposition of U, consisting of $f_0 = |V_1'|$ vertices, f_1^{int} interior edges, f_1^{bdy} other (half-)edges, q quadrilateral faces, and $b_1 + b_2 + b_3$ boundary faces, where b_i counts the faces with i vertices in I'. In particular the total number of edges is $f_1 = f_1^{\text{int}} + f_1^{\text{bdy}} = |E_1'|$,

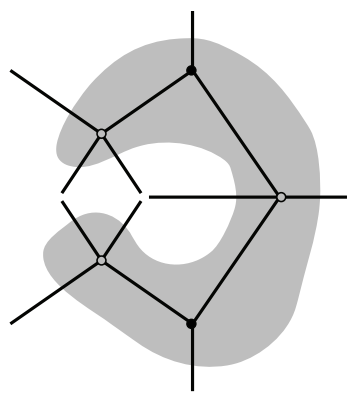


Figure 1.20. An example of five vertices in the reduced diagonal graph of Figure 1.17. The neighborhood U is shaded.
$f_0 = |V_1'| = 5$, $f_1 = |E_1'| = 14$, $f_1^{\text{int}} = 4$, $f_1^{\text{bdy}} = 10$, $h = 0$, $q = 0$, $b_1 = 4$, $b_2 = 4$, $b_3 = 2$.

Double counting the edge-face incidences yields

$$2f_1^{\text{int}} \;=\; 4q + b_2 + 2b_3 \qquad \text{and} \qquad 2f_1^{\text{bdy}} \;=\; 2b_1 + 2b_2 + 2b_3. \tag{1.7}$$

The Euler characteristic of U is

$$1 - h \;=\; f_0 - f_1 + q + b_1 + b_2 + b_3 \;=\; f_0 - f_1^{\text{int}} + q. \tag{1.8}$$

With this we get

$$\begin{aligned} |E_1'| - 2|V_1'| \;=\; f_1 - 2f_0 \;&\overset{(1.8)}{=}\; (f_1^{\text{int}} + f_1^{\text{bdy}}) \;-\; 2(f_1^{\text{int}} - q + 1 - h) \\ &=\; f_1^{\text{bdy}} - f_1^{\text{int}} + 2q + 2h - 2 \\ &\overset{(1.7)}{=}\; (b_1 + b_2 + b_3) - (2q + \tfrac{1}{2}b_2 + b_3) \;+\; 2q + 2h - 2 \\ &=\; \tfrac{1}{2}(2b_1 + b_2 - 4) \;+\; 2h. \end{aligned}$$

To conclude that $|E_1'| - 2|V_1'| \geq 0$, with equality only if $V_1' = V'$, we use $h \geq 0$, and need to verify that $2b_1 + b_2 \geq 4$ holds, with equality only in the trivial case $V_1' = V'$.

For this we count the vertices v of $D \setminus V_1'$ which are *adjacent to* V_1', that is, such that some quad in the full quad-graph D contains both v and a vertex from V_1'. Walking along the boundary curves of U, and exploring the quads that we traverse that way, we see that there are not more than $2b_1 + b_2$ such vertices v: We find at most two new vertices in any quad that contains a boundary cell with 1 vertex in V_1', and at most one new vertex in the quad of a boundary cell with 2 vertices in V_1'. The vertices found during the walk need not be all distinct, and some may not even lie outside V_1' (compare Figure 1.21). Thus we get only an inequality,

$$2b_1 + b_2 \quad \geq \quad \#\{\text{vertices of } D \setminus V_1' \text{ adjacent to } V_1'\}.$$

In the boundary of each "hole" of U we will discover at least one vertex of $D' \setminus V_1'$. In the outer face during our walk we even discover a cycle of D (see Figure 1.21). Since D is bipartite, this cycle has even length. In the trivial case of $V_1' = V'$ this is exactly the 4-cycle C' given by $D \setminus D'$. If $V_1' \neq V'$, then the vertices we discover either yield the cycle C' plus additional vertices, or we find a different cycle. But any cycle other than C' must have at least 6 vertices: Indeed, it is an even cycle, on which black and white vertices alternate. The black vertices on the cycle either include both the vertices of f, or with respect to the original graph G they separate a black vertex in V_1' from a vertex of f; from the 3-connectivity of G we thus get that the cycle contains at least three black vertices, that is, at least 6 vertices in total. The same holds for the white vertices, the dual graph G^*, which is also 3-connected, and the vertices of f^*. Thus

$$\#\{\text{vertices of } D \text{ in the boundary of } U\} \quad \geq \quad 4,$$

with equality only if $V_1' = V'$. This completes the proof for the expansion property, and thus for the existence of a coherent angle system, and of the circle packing.

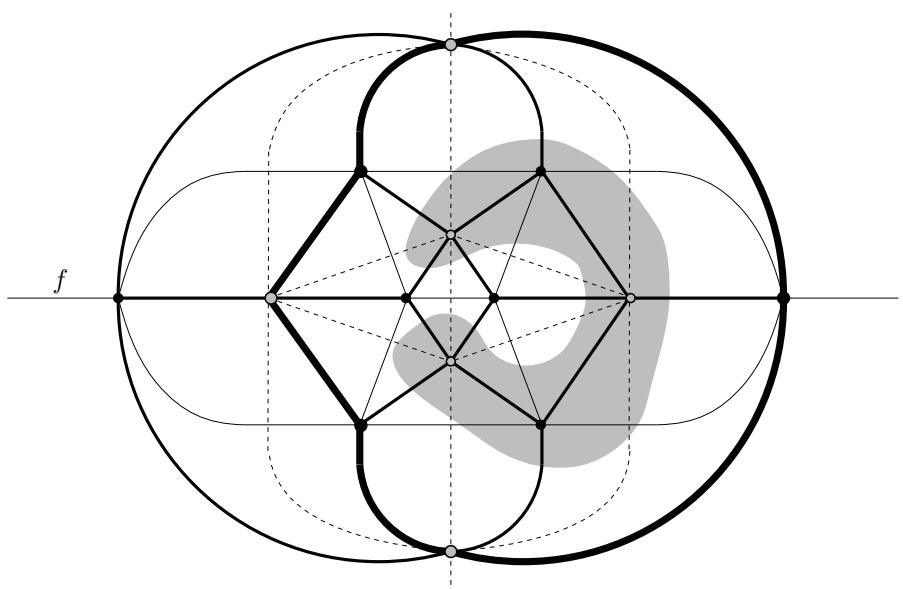


Figure 1.21. The cycle in the outer face to be discovered during the walk along the boundary curve of U is drawn with fat edges; it is a 6-cycle. With respect to G, which is drawn in thin black lines, the three vertices of the 6-cycle separate a vertex of f from the two black vertices in V_1'.

(7), (8). Given a correct rectangular circle pattern, it is easy to reconstruct the spherical circle pattern (via an inverse stereographic projection). From this, we obtain the edge-tangent polytope: Its face planes are given by the facet circles (and its vertices are given by the cone points for which the vertex horizon circles do indeed appear on the horizon). Thus construction steps (7) and (8) are easy — the hard part was (6). □

Is this the perfect proof? I think it is really nice, but still one could dream of a proof that avoids the stereographic projection, and produces the circle packing directly from some functional on the sphere ....

Exercises

1.1. Show that each 3-polytope has a triangle face, or a simple vertex (a vertex of degree 3), or both. Even stronger, show that the number of triangle faces plus the number of simple vertices is at least eight, so there are at least four triangle faces, or at least four simple vertices.
Hint: Use the Euler equation.

1.2. Prove that each 3-polytope has two faces with the same number of vertices.
Hint: Do not use the Euler equation.

1.3. Prove the Steinitz Lemma 1.1:
- Prove the "upper bound theorem" for dimension 3, that is, that $f_2 \leq 2f_0 - 4$ (you may use Euler's equation), and derive $f_0 \leq 2f_2 - 4$ by duality.
- Compute the f-vectors of the pyramids over n-gons.
- How does (f_0, f_2) change if you stack a pyramid onto a triangle 2-face, or if you truncate a simple vertex?

1.4. If a 3-dimensional polytope has $f_1 = 23$ edges, how many vertices/faces can it have? Construct an example for each possible pair (f_0, f_2).

1.5. Alternative homogeneous coordinates for the cone of f-vectors are given by the "imbalance" $\sigma := \frac{f_2 - f_0}{f_1 - 6}$, where the self-dual term $f_1 - 6$ measures the "size." Show that $-\frac{1}{3} \leq \sigma \leq +\frac{1}{3}$, where $\sigma = \pm\frac{1}{3}$ characterizes simple resp. simplicial polytopes.

1.6. Characterize the possible (f_0, f_2)-pairs for *cubical* 3-polytopes, that is, for all polytopes with quadrilateral 2-faces only. Where are the (f_0, f_2)-pairs of cubical 3-polytopes in Figure 1.5?
How about 3-polytopes with pentagon faces only? Hexagon faces only?

1.7. Construct quad graphs and the planar circle patterns for
(a) a square pyramid,
(b) a cube/octahedron,
(c) a cube with vertex cut off,
(d) a dodecahedron.
Which of the circle patterns do you get with rational coordinates?

1.8. Show that every 3-dimensional cyclic polytope $C_3(n)$ is a stacked polytope. (However, $C_d(n)$ is not stacked, for $d \geq 4$ and $n \geq d + 2$.)

1.9. Describe a computational procedure to construct a coherent angle system: For this use a scheme to augment flows along undirected paths in the network with lower and upper bounds (increasing the value along forward arcs, decreasing the values on backward arcs). Your procedure should also imply a proof for the Generalized Max-Flow Min-Cut Theorem [**1**, Thm. 6.10, p. 193].

LECTURE 2
Shapes of f-Vectors

Let's look at the f-vectors of d-dimensional convex polytopes P, where the dimension d is really large. Any such f-vector

$$\begin{aligned} f(P) &= (f_0, f_1, f_2, \quad \dots \quad \dots \quad \dots \quad , f_{d-3}, f_{d-2}, f_{d-1}) \\ &= (\#\text{vertices}, \#\text{edges}, \#\text{2-faces}, \dots, \#\text{subridges}, \#\text{ridges}, \#\text{facets}) \end{aligned}$$

is a long sequence of large numbers, which we may graph just like a continuous function, and ask for its "shape." Indeed, we might look at a shape function $\varphi : [0,1] \to \mathbb{R}$ that is defined by $\varphi(x) := f_{x(d-1)}$; this is defined for any $x = \frac{k}{d-1}$ that is a multiple of $\frac{1}{d-1}$, and these values are rather dense if d is large. We might interpolate if we want. But what types of f-vector shape functions φ do we get that way?

Figure 2.1 shows two "naive" views, of the shape of an f-vector, and — equivalently — of the shape of a typical face lattice (displayed as a *Hasse diagram*, so the sizes of rank levels are the f_i-values).

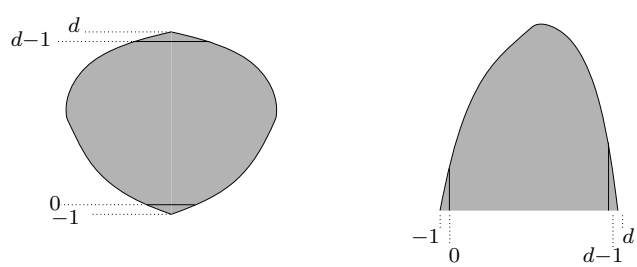


Figure 2.1. A rough, "naive" picture of the shape of the face lattice, and the f-vector, for a high-dimensional polytope

A very simple observation is that each vertex of a d-polytope has degree at least d, so double counting yields $f_1 \geq \frac{d}{2} f_0 > f_0$; dually, we have $f_{d-2} \geq \frac{d}{2} f_{d-1} > f_{d-1}$. So in the first step, the f-sequence increases, in the last step it decreases. Does this mean that the f-vector "first goes up, then comes down," that it is *unimodal*, with no "dip" in the middle?

2.1. Unimodality Conjectures

Unimodality conjectures and theorems abound in combinatorics [**71**] [**19**]: for binomial coefficients, Stirling numbers and their generalizations, matroids and geometric lattices, etc. . . . The basic unimodality conjecture for convex polytopes was posed at least twice, by Theodore Motzkin in the late fifties, and by Dominic Welsh in 1972 (see [**13**]). Apparently it was disproved dramatically by Ludwig Danzer, already in the early sixties (presented in a lecture in Graz in 1964, according to Jürgen Eckhoff), but this is "lost mathematics," no published account exits.

Conjecture 2.1. *The f-vectors of convex polytopes are unimodal, that is, for each d-polytope P there is an $\ell = \ell(P)$ such that*

$$f_0 \leq f_1 \leq \ \cdots \ \leq f_\ell \geq \ \cdots \ \geq f_{d-2} \geq f_{d-1}.$$

The main point of this lecture will be to see that this is dead wrong, even for simplicial polytopes. Moreover, we want to see this "asymptotically," without substantial amounts of computation, without having to list explicit f-vectors.

This asymptotic view is also motivated by the fact that the conjecture fails only in high dimensions. For example, for simplicial polytopes, it is true up to $d = 19$, and fails beyond this dimension. For general polytopes, we will see a counterexample for $d = 8$, but none are known for a smaller dimension. The conjecture holds in full for $d \leq 4$ (Exercise 2.2), and also for $d = 5$, according to Werner [**78**].

Since the conjecture is so badly wrong, it might pay off to explicitly state what remains from it:

Conjecture 2.2 (Björner [**13**] [**15**])**.** *The f-vectors of convex polytopes increase on the first quarter, and they decrease on the last quarter:*

$$f_0 < f_1 < \ \cdots \ < f_{\lceil \frac{d-1}{4} \rceil}, \qquad\qquad f_{\lfloor \frac{3(d-1)}{4} \rfloor} > \ \cdots \ > f_{d-2} > f_{d-1}.$$

This is trivially true for $d \leq 5$. It also is true for *simplicial* d-polytopes (the f-vectors of simplicial polytopes indeed increase up to the middle, and they decrease in the last quarter), but the available proof for this depends on the necessity part of the g-theorem, so it is quite non-trivial; see [**15**].

To demonstrate our ignorance on such basic f-vector shape matters, here is a suspiciously innocuous conjecture. Apparently no one has an idea for a proof, up to now.

Conjecture 2.3 (Bárány)**.** *For any d-polytope, $f_k \geq \min\{f_0, f_{d-1}\}$.*

Bárány's conjecture holds for $d \leq 6$ [**78**]. However, not even

$$f_k \quad \geq \quad \tfrac{1}{10000} \min\{f_0, f_{d-1}\}$$

is proven for large dimensions d! We know so little . . .

2.2. Basic Examples

Let's compute the f-vector shapes for the most basic high-dimensional polytopes that we can come up with. For rough estimates, we use a very crude version of Stirling's formula,

$$n! \ \sim \ \left(\frac{n}{e}\right)^n.$$

Example 2.4 (The simplex)**.** For the $(d-1)$-simplex Δ_{d-1} we have

$$f_{k-1}(\Delta_{d-1}) \;=\; \binom{d}{k}.$$

With logarithms taken with base 2, $x := \frac{k}{d}$, and $\varphi(x) = f_{xd-1}$, we get

$$\log\varphi(x) \;=\; \log\binom{d}{xd} \;\sim\; -x\log x - (1-x)\log(1-x).$$

A little bit of analysis shows from this that the f-vector is symmetric, with a sharp peak in the middle (at $x = \frac{1}{2}$), of width $\sim \frac{1}{\sqrt{d}}$. Figure 2.2 displays a realistic example.

Of course, this is a well-known property of binomial coefficients, and the strong limit theorems of probability theory depend on it. (In this context $\varphi(x)$ is known as the "entropy function.")

Example 2.5 (Cross polytopes)**.** For the d-dimensional cross polytope $C_d^* = \operatorname{conv}\{\pm e_1, \ldots, \pm e_d\}$ we have

$$f_k(C_d^*) \;=\; \binom{d}{k+1}2^{k+1}.$$

Again, approximating crudely and taking logarithms base 2, we get

$$\log\varphi(x) \;\sim\; -x\log x - (1-x)\log(1-x) + x.$$

The derivative

$$\frac{\mathrm{d}}{\mathrm{d}x}\varphi(x) \;\sim\; -\log x - \frac{1}{\ln 2} + \log(1-x) + \frac{1}{\ln 2} + 1$$

vanishes at $x = \frac{2}{3}$: That's where $\log\varphi(x)$ has its maximum, and where $\varphi(x)$ has a sharp peak (compare Figure 2.2).

Thus the f-vector of a d-dimensional cross polytope, for large d, has a sharp peak at $k = \frac{2}{3}d$. By duality, this means that the f-vector of the d-cube peaks at $k = \frac{1}{3}d$, for large d.

Example 2.6 (Cyclic polytopes)**.** Let's look at cyclic polytopes $C_d(n)$ with many vertices, $n \gg d$. For simplicity, we assume that the dimension d is even.

A curve in $\mathbb{R}^d$ has degree d if no $d+1$ points on the curve lie on a hyperplane. The convex hull of any $n > d$ points on such a curve is a cyclic polytope $C_d(n)$. *Gale's evenness criterion* [**30**] gives a combinatorial description for the facets, which is easy to visualize (see Figure 2.3): Any d points on a degree d curve span a hyperplane H. If the d points are supposed to span a facet of the polytope, then all the other $n-d$ points must lie on the same side of H. Since the curve crosses H *only* in these d points, this means that the d points split into $\frac{d}{2}$ adjacent pairs. So, if we number the points $1, 2, \ldots, n$ along the curve, then the facets of their convex hull (the cyclic polytope) are given by $\frac{d}{2}$ pairs $i, i+1 \bmod n$. The $(k-1)$-faces are given by the k-subsets of such a d-set: For $k \le \frac{d}{2}$ *any* such subset will do (the cyclic polytopes are *neighborly*), while for $k > \frac{d}{2}$ the faces consist of $k - \frac{d}{2}$ pairs, and $d-k$ singletons. Thus the $(k-1)$-faces may be obtained by choosing $\frac{d}{2}$ vertices

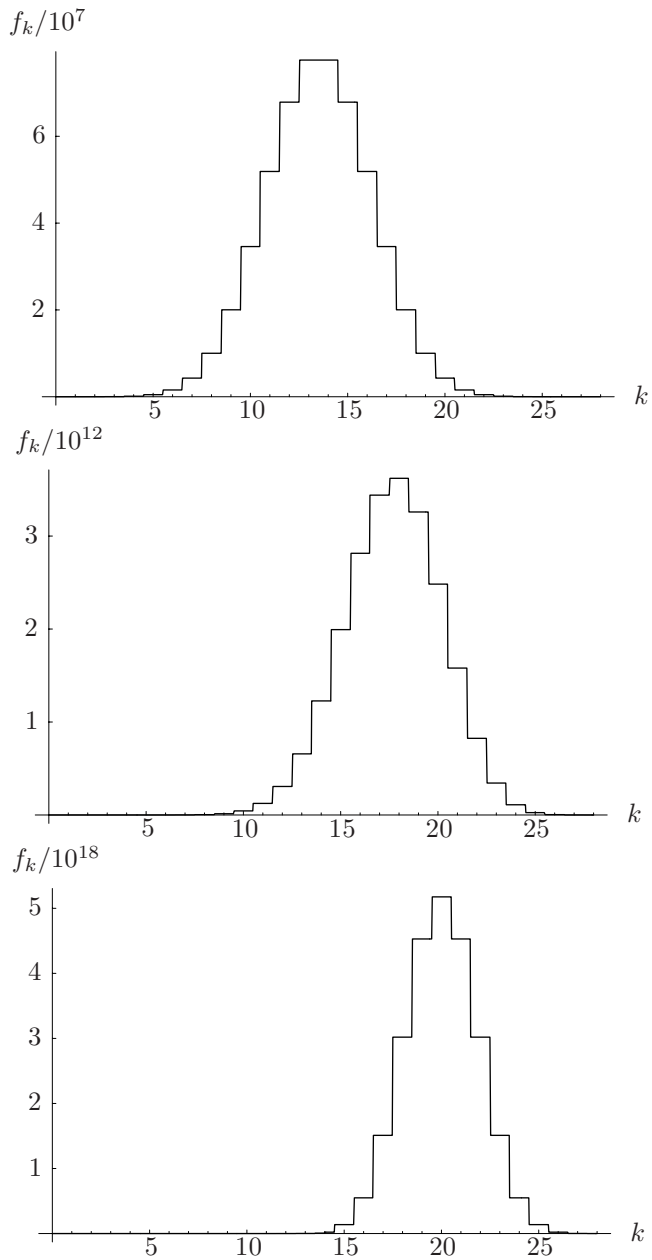


Figure 2.2. The f-vector shapes of the 28-dimensional simplex Δ_{28}, the cross polytope C^*_{28}, and a cyclic polytope with 80 vertices $C_{28}(80)$

i_j arbitrarily, and also taking $i_j + 1$ for $k - \frac{d}{2}$ of these (see Figure 2.4). Thus, with a bit of an over-count, we get

$$f_{k-1}(C_d(n)) \begin{cases} = \binom{n}{k} & \text{for } k \leq \frac{d}{2}, \\ \sim \binom{n}{\frac{d}{2}}\binom{\frac{d}{2}}{k-\frac{d}{2}} & \text{for } k > \frac{d}{2}. \end{cases}$$

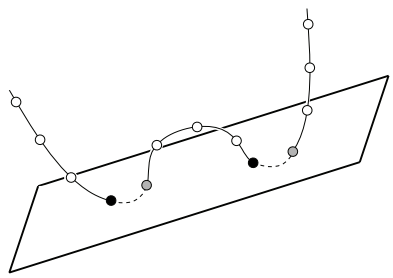


Figure 2.3. Sketch for Gale's evenness criterion.

Figure 2.4. An estimate for the number of facets of $C_d(n)$, for $n \gg d$, with d even: There are $\binom{n}{d/2}$ choices for the black points; with high probability, they are non-adjacent; the $\frac{d}{2}$ pairs can be completed by taking the gray points.

Clearly this peaks at $x = \frac{3}{4}$: We get the larger entries in the case $k > \frac{d}{2}$, and then the maximum is achieved when $\binom{d/2}{k-d/2}$ is maximal, that is, for $k = \frac{3}{4}d$. Figure 2.2 gives a realistic impression of the f-vector shape of a cyclic polytope.

An explicit, exact formula for $f_{k-1}(C_d(n))$ is available (Exercise 2.3), but this doesn't answer all the questions. In particular, is it really true that the f-vector is unimodal? As far as I know, the Unimodality Conjecture 2.1 has not been established in full for the cyclic polytopes. It does hold for small $n > d$, and certainly also if $n \gg d$ is sufficiently large compared to d (with the f-vector peak at $k = \lfloor \frac{3(d-1)}{4} \rfloor$), but in an intermediate range for n a challenge remains ...

2.3. Global Constructions

We have seen classes of simplicial d-polytopes whose normalized f-vector functions $\varphi(x) = f_{x(d-1)}$ peak at $x = \frac{1}{2}$, at $x = \frac{2}{3}$, or at $x = \frac{3}{4}$. By dualization we get simple polytopes with peaks at $x = \frac{1}{3}$, and at $x = \frac{1}{4}$. The "global constructions" of products and joins now yield examples with peaks in the whole range between $x = \frac{1}{4}$ and $x = \frac{3}{4}$. (The product construction is elementary, well-known, and well-understood, but a review perhaps can't harm, also in view of our needs for Lecture 5. Joins are similarly elementary and well-understood, but perhaps not that well-known.)

Example 2.7 (Products)**.** Let P and Q be polytopes of dimensions d and e. Then the product

$$P \times Q \;:=\; \{(x,y) : x \in P,\, y \in Q\}$$

is a polytope of dimension $\dim(P \times Q) = \dim P + \dim Q = d + e$.

The nonempty faces of $P \times Q$ are the products of nonempty faces of P and nonempty faces of Q: In particular, the vertices of $P \times Q$ are of the form "vertex times vertex," the edges are of the form "edge times vertex" or "vertex times edge," and the facets are "P times facet of Q" or "facet of P times Q." With the convention $f_d(P) = f_e(Q) = 1$ this yields the formula

$$f_m(P \times Q) \;=\; \sum_{\substack{k+\ell=m \\ k,\ell \geq 0}} f_k(P)\, f_\ell(Q) \tag{2.1}$$

for $m \geq 0$.

The product construction is dual to the "free sum" construction, $P \oplus Q$: For this let $x_0 \in P \subset \mathbb{R}^d$ and $y_0 \in Q \subset \mathbb{R}^e$ be interior points, and take the convex hull

$$P \oplus Q \;:=\; \operatorname{conv}\big(P \times \{y_0\} \;\cup\; \{x_0\} \times Q\big).$$

The *proper* faces of $P \oplus Q$ (that is, faces other than the polytope itself) arise as joins of proper faces of P and of Q.

The product and the free sum construction are illustrated in Figure 2.5.

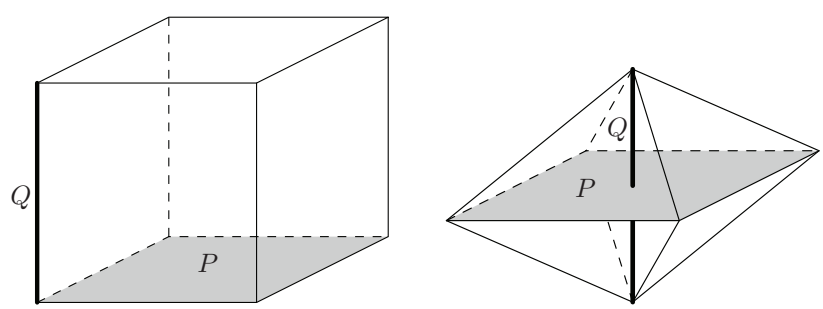


Figure 2.5. Product and free sum, for $P = I^2$ and $Q = I$

Since joins come up as faces of free sums, let's briefly talk about joins.

Example 2.8 (Joins)**.** Let again P and Q be polytopes of dimensions d and e. Then the *join* $P * Q$ is obtained by positioning P and Q into skew affine subspaces, and taking the convex hull. Thus the join is a polytope of dimension $\dim(P * Q) = \dim P + \dim Q + 1 = d + e + 1$.

The faces of $P * Q$ are the joins of faces of P and faces of Q: This refers to *all* faces, including the empty face and the polytope itself. The corresponding formula, with $f_{-1}(P) = f_{-1}(Q) = 1$, is

$$f_m(P * Q) \;=\; \sum_{\substack{k+\ell=m-1 \\ k,\ell \geq -1}} f_k(P)\, f_\ell(Q), \tag{2.2}$$

valid for all m, that is, for $-1 \leq m \leq d + e + 1$.

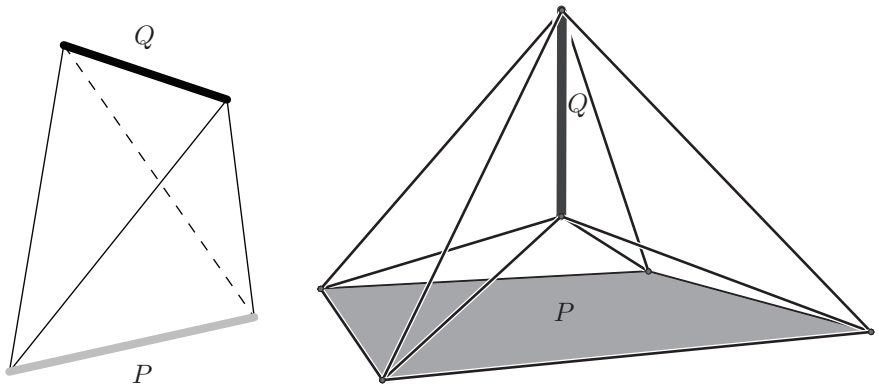


Figure 2.6. Joins $I * I$ and $I^2 * I$, of an edge with an edge, resp. of a square with an edge

Joins are illustrated in Figure 2.6. The dual construction to taking joins is the join construction again.

Product and join are two distinct constructions, and they do yield different polytopes, of different dimensions (by 1). However, in a birds' eye view, asymptotically, they do behave quite similarly, and indeed, their effects on f-vector shapes are almost the same. Namely, the formulas (2.2) and (2.1) describe finite convolutions, and the only difference is whether the entry $f_{-1} = 1$ is counted. For large dimensions, and large f-vectors, this does not make much of a difference, and in both cases we get a convolution of f-vector shapes. Thus, in particular, if the f-vectors of P and of Q have sharp peaks, then the product or join will have a peak as well:

$$(\text{peak at } x) * (\text{peak at } y) \;\longrightarrow\; (\text{peak at } \tfrac{d}{d+e}x + \tfrac{e}{d+e}y).$$

In particular, for $d = e$ this yields

$$(\text{peak at } x) * (\text{peak at } y) \;\longrightarrow\; (\text{peak at } \tfrac{x+y}{2}).$$

To see this, just compute that if the peak (or, just the largest f-vector entry) for P_1 is at $x = \frac{k}{d}$ and for P_2 at $y = \frac{\ell}{e}$, then the peak for $P_1 * P_2$ will be at

$$\tfrac{k+\ell}{d+e} \;=\; \tfrac{k}{d}\tfrac{d}{d+e} + \tfrac{\ell}{e}\tfrac{e}{d+e} \;=\; x\tfrac{d}{d+e} + y\tfrac{e}{d+e}.$$

This also yields a convolution formula for the f-vector shape of $P_1 \times P_2$ or $P_1 * P_2$, for large dimensions:

$$\varphi(x) \;=\; \int_0^1 \varphi_1\big(t\tfrac{d}{d+e}\big)\,\varphi_2\big((1-t)\tfrac{e}{d+e}\big)\,\mathrm{d}t$$

Thus, by just taking products of sums of suitable cyclic polytopes and their duals, we do get polytopes with f-vector peaks in the whole range between $\frac{1}{4}$ and $\frac{3}{4}$.

2.4. Local Constructions

Perhaps the simplest local operation that can be applied to a polytope is to "stack a pyramid onto a simplicial facet." To perform such a *stacking* operation geometrically, the new vertex of course has to be chosen carefully (*beyond* the simplicial facet, and *beneath* all other facets, in Grünbaum's terminology [**39**, Sect. 5.2]), but

the combinatorial description is easy enough. In particular, we get the following f-vector equation:

$$f_k(\text{stack}\, P) \;=\; f_k(P) + f_k(\Delta_d) - f_k(\Delta_{d-1}).$$

This is valid for $k < d-1$, the rest is "boundary effects" that we may safely ignore. Furthermore, the usual binomial recursion yields $f_k(\Delta_d) - f_k(\Delta_{d-1}) = f_{k-1}(\Delta_{d-1})$, and we get

$$f_k(\text{stack}\, P) \;=\; f_k(P) \;+\; f_{k-1}(\Delta_{d-1}).$$

So, the effect of stacking on the f-vector is to add a bump at $x = \frac{1}{2}$. The effect may be negligible if the f-vector of P is large, and has large slopes. However, any stacking operation destroys a simplicial facet and creates d new ones, so it can be repeated. We write $\text{stack}^N P$ for a polytope that is obtained from P by N subsequent stacking operations. Thus we get

$$f_k(\text{stack}^N P) \;=\; f_k(P) \;+\; N\, f_{k-1}(\Delta_{d-1}),$$

where we may choose $N \geq 0$ freely. Thus we are adding a function with peak at $\frac{1}{2}$ to a function whose peak may be, for example, at $\frac{2}{3}$.

Corollary 2.9 (Danzer 1964)**.** *For large enough d and suitable N, the Unimodality Conjecture 2.1 fails for "N-fold stacked crosspolytopes"* $\text{stack}^N C_d^*$.

Indeed, Danzer apparently also derived that the f-vector of a simplicial d-polytope may have not only one dip (between two peaks), but arbitrarily many dips and peaks!

Also, dualization yields that a suitable number of vertex truncations applied to a high-dimensional cube leads to a simple polytope with a non-unimodal f-vector.

However, cross polytopes are not the most effective starting points for non-unimodal examples: If we use cyclic polytopes, then the peak (at $\frac{3}{4}$) is further away from the peak for a simplex that we can "add" by stacking (at $\frac{1}{2}$). Moreover, in cyclic polytopes we can control the number of vertices in fixed dimension as well, and thus make the peak at $\frac{3}{4}$ as sharp as we want.

Theorem 2.10 (Björner [**13**] [**15**], Lee [**47**] [**12**], Eckhoff [**24**])**.** *The Unimodality Conjecture 2.1 holds for simplicial d-polytopes of dimensions $d \leq 19$, but it fails for $d \geq 20$.*

Specifically: Stacking $N = 259 \cdot 10^{11}$ times onto the cyclic polytope $C_{20}(200)$, one obtains a polytope with a dip $f_{11} > f_{12} < f_{13}$ in the f-vector,

$$\begin{array}{rcl} f_{11} & = & 5049794068451336750 \\ \vee & & \\ f_{12} & = & 5043828885028647000 \\ \wedge & & \\ f_{13} & = & 5045792044986529500. \end{array}$$

The proof of the first part of Theorem 2.10 utilizes the g-theorem (see Stanley [**69**] and Björner [**14**]), which explicitly describes the f-vectors of the simplicial polytopes, plus a substantial amount of "binomial coefficient combinatorics." See [**15**] for $d \leq 16$; the extension to $d \leq 19$, due to Eckhoff, unfortunately is still not published.

If we leave the realm of simplicial polytopes, then it becomes even easier to construct polytopes with a non-unimodal f-vector. Then we can try to add the

f-vectors of two polytopes with peaks at $\frac{1}{4}$ and at $\frac{3}{4}$, say a cyclic polytope and its dual. And indeed, just as we can glue a pyramid onto a simplicial facet, we can glue any polytope with a simplicial facet onto another one — after a projective transformation, if needed [**79**, p. 274]. The f-vector effect of such a glueing is essentially

$$f(P\#P') \;=\; f(P) + f(P') - f(\Delta_{d-1});$$

if the f-vector components of P and of P' are large, then the simplex may be neglected, and we are essentially just "adding the f-vectors."

We can even do this with cyclic polytopes: For example, $C_d(n)$ is simplicial; its dual, $C_d(n)^*$ is simple (without simplicial facets), but if we cut off ("truncate") one of the simple vertices, then a simplicial facet results. Write $C_d(n)'$ for the "dual with a vertex cut off."

Corollary 2.11 (Eckhoff [**24**]). *The Unimodality Conjecture 2.1 fails for d-polytopes of dimensions $d \geq 8$. In particular,*

$$f(C_8(25)\#C_8(25)') \;=\; (7149, 28800, 46800, 46400, 46400, 46800, 28800, 7149).$$

This f-vector has a nice "1% dip" in the middle! We don't know whether the Unimodality Conjecture 2.1 is true for dimensions $d = 6$ or 7.

Exercises

2.1. For $d = 3, 4, 5, \ldots$ construct a d-polytope with 12 vertices and 13 facets. How far do you get?
2.2. Show that f-vectors of 4-polytopes are unimodal.
2.3. Derive an *exact* formula for $f_{d-1}(C_d(n))$, and for $f_k(C_d(n))$, for even n.
2.4. Compute $f_i(C_8(25))$. How bad is the approximation given in Example 2.6?
2.5. Count and describe the 2-faces of a product of a pentagon and a heptagon, $P_5 \times P_7$.
2.6. Compute $f((C_{10})^{10})$, for the product of ten 10-gons. Where is the peak?
2.7. Estimate/compute d and N such that the "N-fold truncated d-cube" has a non-unimodal f-vector.
2.8. If you stack "too often" onto $C_{20}(200)$, then unimodality is restored. How often?

LECTURE 3
2-Simple 2-Simplicial 4-Polytopes

The boundary complex of a 4-polytope is a 3-dimensional geometric structure. So, in contrast to the high-dimensional polytopes discussed in the previous lecture, we can hope to approach 4-polytopes via explicit visualization and geometric constructions. *Schlegel diagrams* are a key tool for this.* Another one, which we will also depend on in a key moment of this lecture, is *dimensional analogy*: To describe a construction of 4-polytopes, we phrase a key step as a statement that it is valid "for all $d \geq 3$," where the visualization is done for the special case $d = 3$, while the most interesting results are obtained for $d = 4$.

The geometry and combinatorics of polytopes in dimension 4 is much more interesting, rich, and difficult than in 3 dimensions, because 4-polytopes aren't constrained between only two extremes, simple and simplicial. Some of the most fascinating examples around, such as Schläfli's 24-cell, are neither simple nor simplicial, but 2-*simple* 2-*simplicial*. This property was thought to be rare until recently: Only a few years ago, exactly 8 such polytopes were known. (Unfortunately, a claim by Shephard from 1967 did not work out: In [**39**, p. 82] it had been claimed that Shephard could produce infinite families, and that each 4-dimensional convex body could be approximated by 2-simple 2-simplicial 4-polytopes, which would have established a conjecture by David Walkup. Compare [**39**, p. 96b])

The main goal for this lecture is to describe a simple, explicit, geometric construction that produces rich infinite families of 2-simple 2-simplicial 4-polytopes. The first infinite families, obtained by Eppstein, Kuperberg & Ziegler in 2001 [**26**], relied on rather subtle constructions, via Koebe–Thurston type edge-tangent realizations of 4-polytopes (which exist only in rare cases), and hyperbolic angle measurements. In contrast to this, the *deep vertex truncation* construction to be described here is remarkably simple; it appears in Paffenholz & Ziegler [**57**], while special instances (for semi-regular polytopes) can be traced back to Coxeter's classic [**22**, Chap. VIII], who refers to Cesro (1887) for the construction of the 24-cell by what we here call a "deep vertex truncation" of the regular 4-cube.

*These were apparently introduced by Dr. Victor Schlegel, a highschool (Gymnasium) teacher from Waren an der Müritz, in his paper [**67**] from 1883. The plates for the paper include a Schlegel diagram ("Zellgewebe") of a 4-cube, as well as two quite insufficient drawings representing the 24-cell. Classical, beautiful drawing may be found in Hilbert & Cohn-Vossen [**41**, p. 135].

3.1. Examples

Let's start with examples of well-known 4-polytopes — and for each of those let's look at a Schlegel diagram, and record the f-vector

$$(f_0, f_1, f_2, f_3) \;=\; (\,\#\text{ vertices}, \#\text{ edges}, \#\text{ 2-faces } (=\text{ridges}), \#\text{ facets}\,).$$

A *Schlegel diagram* is a way to visualize a 4-polytope in terms of a 3-dimensional complex. We can't develop the theory of Schlegel diagrams here (see [**39**, Sect. 3.3] and [**79**, Lect. 5]), but we can offer two interpretations, both in terms of dimensional analogy.

- Assume that one face of a 3-polytope is transparent (a "window"), press your nose to the window, and look inside: Then you will see all the other faces of the polytope through the window. If you now close one eye (and thus lose the spatial impression, or depth view), then you will see how the other faces tile the window; you can see how they fit together, and thus the whole combinatorial structure of the 3-polytope is projected into a 2-dimensional window. This is the Schlegel diagram of a 3-polytope.
- Any 3-polytope can be projectively deformed in such a way that looking at it from a suitable point, you *see* all faces except for one single face, which is on the back. What you see is a polytopal complex which has the same shape as the back face, but this is broken into all the many faces that you see on the front side. What you see is the 2-dimensional Schlegel diagram of a 3-polytope.

The Schlegel diagram of a 4-polytope, analogously, is a 3-dimensional complex that represents all the faces of the polytope, except for one facet (the window resp. back facet). The whole combinatorial structure of the polytope may be read from such a visualization. Thus, for example, one can tell whether the polytope is simple, or simplicial, or cubical, etc.

The pictures of Schlegel diagrams as presented in the following are generated automatically in the `polymake` system by Gawilow & Joswig [**31**], with the `javaview` back-end by Polthier et al. [**58**]. They have three limitations: They show only a 2-dimensional projection of an object that you should see rotating, 3-dimensionally, on a screen; they depict only the edges, so in some examples it is hard to tell/imagine where the faces and facet-boundaries go; and we don't have color available here. Nevertheless, I think they are impressive, and you should be able to "see" in them what the (boundary complexes of) some 4-polytopes look like.

Example 3.1 (Simplex, cube, and cross polytope)**.** Schlegel diagrams of the 4-simplex, the 4-cube and the 4-dimensional cross polytope appear in Figure 3.1. You should read off the f-vectors from this figure: $f(\Delta_4) = (5, 10, 10, 5)$, $f(C_4) = (16, 32, 24, 8)$, and $f(C_4^*) = (8, 24, 32, 16)$.

The simplex and cube are simple, so $f_1 = 2f_0$, while the simplex and cross polytope are simplical, so $f_2 = 2f_3$.

Example 3.2 (A cubical 4-polytope with the graph of a 5-cube [**43**])**.** The construction

$$P \;:=\; \operatorname{conv}((2Q \times Q) \cup (Q \times 2Q)),$$

for a square such as $Q = [-1, 1]^2$, yields a 4-polytope whose Schlegel diagram is displayed in Figure 3.2. This polytope is *cubical*: All its facets are combinatorially equivalent to the 3-cube $[-1, 1]^3$.

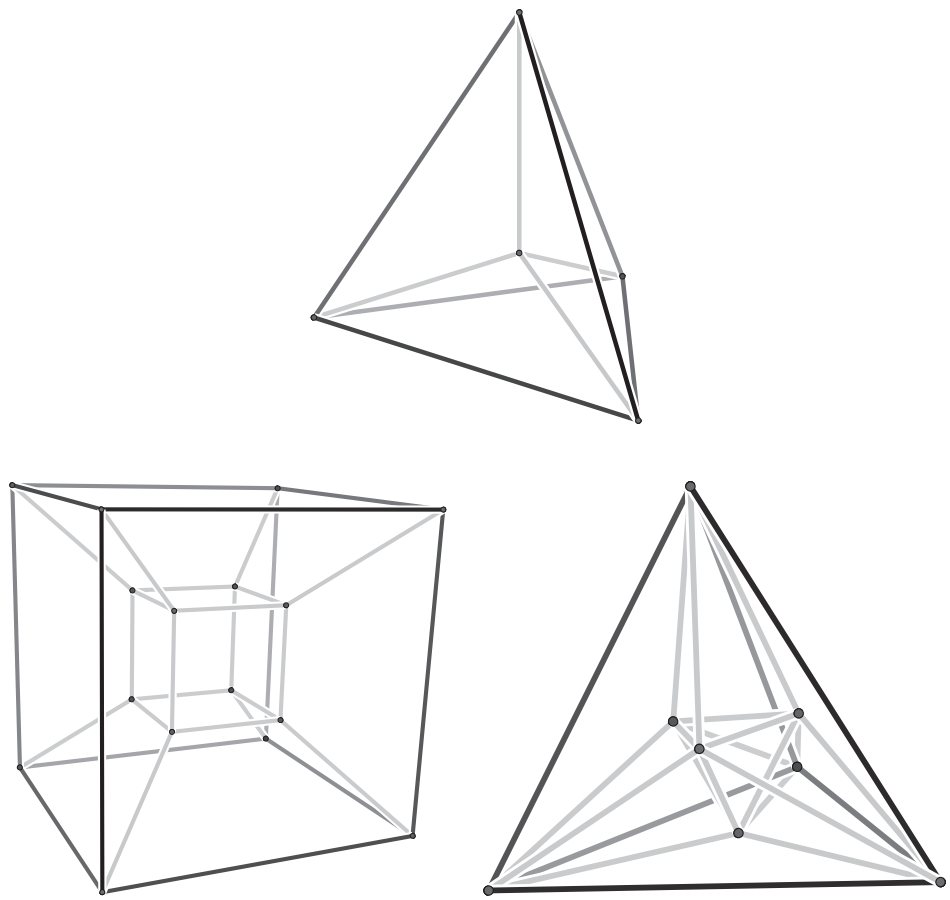


Figure 3.1. Schlegel diagrams for the 4-dimensional simplex, cube, and cross polytope

The f-vector $(32, 80, 72, 24)$ may be derived from the figure, but indeed it may also be deduced just from the information that this is a cubical 4-polytope with the graph of a 5-cube. (The latter yields f_0 and f_1, the "cubical" property implies $2f_2 = 6f_3$ by double counting, and then there is the Euler–Poincaré equation [**79**, Sect. 8.2], which for 4-polytopes reads $f_0 - f_1 + f_2 - f_3 = 0$. See also Exercise 3.2.)

Example 3.3 (The hypersimplex)**.** The hypersimplexes form a 2-parameter family $\Delta_{d-1}(k)$ of remarkable polytopes; as Robert MacPherson said in his PCMI lectures, they have by far not received the attention, study, and popularity that they deserve. They do appear, for example, as K_k^d in [**39**, p.65], as $\Delta^{k,\ell}$ in [**29**, Sect. 1.6] (where apparently the name "hypersimplex" appeared first), in [**35**], in [**34**, p. 207], and in [**23**]; but also elsewhere they appear under disguise, for example, as the cycle polytopes of uniform matroids (see e.g. [**38**]).

The hypersimplex $\Delta_{d-1}(k)$ may be defined as the convex hull of all the 0/1-vectors of length d that consist of k ones and $d-k$ zeroes. This is a $(d-1)$-dimensional polytope with $\binom{d}{k}$ vertices. In the special case $k = 1$ and $k = d-1$ we obtain simplices.

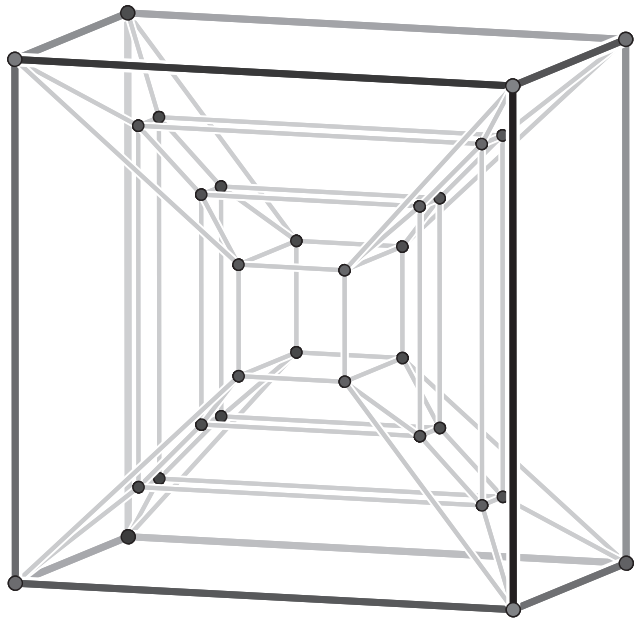


Figure 3.2. A cubical 4-polytope with the graph of the 5-cube

What we call *the hypersimplex* is a 4-dimensional polytope $\Delta_4(2)$ that appears in this family. It may be defined, lying on a hyperplane in $\mathbb{R}^5$, as

$$\left\{x \in [0,1]^5 : \sum_{i=1}^{5} x_i = 2\right\} \;=\; \operatorname{conv}\{e_i + e_j : 1 \le i < j \le 5\},$$

or equivalently, after projection to $\mathbb{R}^4$ by "deleting the last coordinate," as

$$\left\{x \in [0,1]^4 : 1 \le \sum_{i=1}^{4} x_i \le 2\right\} \;=\; \operatorname{conv}\big(\{e_i : 1 \le i \le 4\} \cup \{e_i + e_j : 1 \le i < j \le 4\}\big).$$

The first representation is more symmetric: It yields "by inspection" that all $\binom{5}{2} = 10$ vertices of this polytope are equivalent (under symmetries that permute the coordinates), but that there are two types of facets, five simplices and five octahedra, which appear in vertex-disjoint pairs, "opposite to each other," in parallel hyperplanes. In particular, all the facets are simplicial, that is, all the 2-faces are triangles, so the polytope is 2-*simplicial*.

The second representation has the advantage of being full-dimensional, and it supplies us with a Schlegel diagram (using an octahedron facet as a "window"), as displayed in Figure 3.3. In the figure we may see that the (ten, equivalent) vertex figures are triangular prisms, so they are simple; thus in this 4-polytope, each edge is in exactly three facets, so the polytope is 2-*simple*. So we have seen our first example (other than the 4-simplex) of a 2-simple, 2-simplicial 4-polytope.

From the data given it is easy to compute the f-vector of the hypersimplex: It is $f = (10, 30, 30, 10)$.

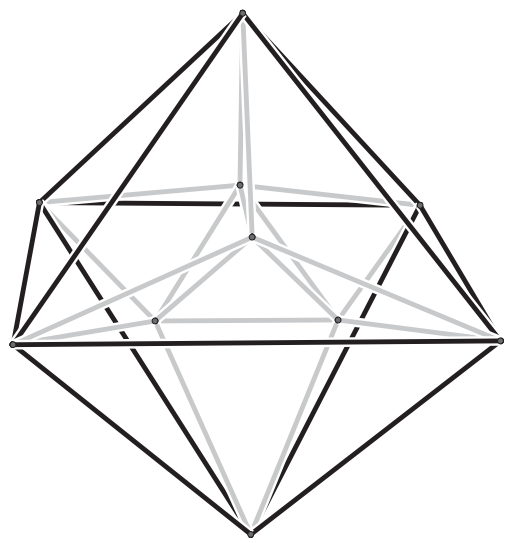


Figure 3.3. A Schlegel diagram of the hypersimplex

3.2. 2-simple 2-simplicial 4-polytopes

Definition 3.4. A 4-polytope $P \subseteq \mathbb{R}^4$ is 2-*simple* 2-*simplicial* ("2*s*2*s*" for short) if all 2-faces of P, and of P^*, are triangles.

The definition given here has the nice feature of being self-dual: Clearly, P is 2s2s if and only if its dual P^* is 2s2s. A more explicit version is that a 4-polytope is 2s2s if and only if

- every 2-face has the minimal number 3 of vertices, and if
- every 1-face (edge) lies in the minimal number 3 of facets.

Still equivalently, this is if and only if

- for every 2-face G the lower interval $[\emptyset, G]$ in the face lattice of P is boolean, and if
- for every 1-face e the upper interval $[e, P]$ in the face lattice of P is boolean.

Thus the 2s2s property may be pictured in analogy with the properties of being simple, or being simplicial. For this we note that, for example, P is simplicial if

- for every 3-face F (facet) the lower interval $[\emptyset, F]$ in the face lattice of P is boolean, and if
- for every 2-face R (ridge) the upper interval $[R, P]$ in the face lattice of P is boolean.

(The first property just says that the facets should be simplices; the second property is automatically satisfied: Every ridge lies in two facets.) And similarly for simple 4-polytopes — see Figure 3.4.

Of course all this suggests generalizations, to ask for *h-simple k-simplicial d-polytopes*, apparently introduced by Grünbaum [**39**, Sect. 4.5]. For $h + k > d$ these don't exist (other than the d-simplex), but also for small h and k they are hard to construct. Indeed, are there any 5-simple 5-simplicial d-polytopes that are not simplexes? Not a single example is known. Compare [**57**] for more information. Here we will restrict ourselves to the 4-dimensional case of 2s2s polytopes. Let's

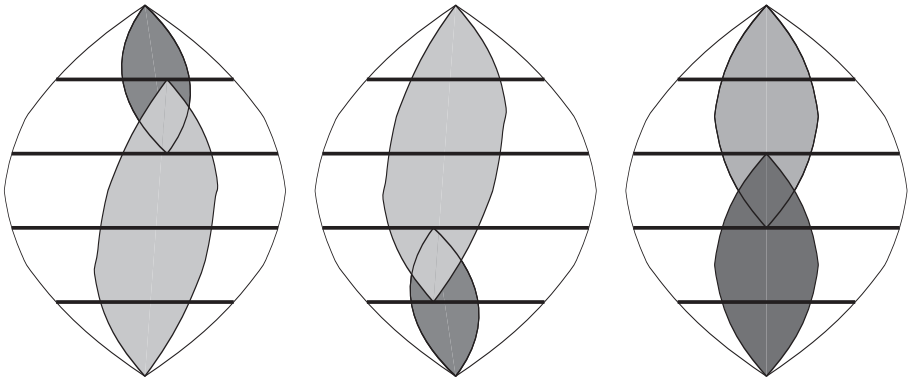


Figure 3.4. Simplicial, simple, and 2s2s 4-polytopes in terms of their face lattices: The shaded intervals, and all the other intervals between the same rank levels, must be boolean.

note one interesting property that is specific for the 4-dimensional case, and which also confirms the impression that 2s2s 4-polytopes form a "diagonal" case.

Lemma 3.5. *Every* 2*s*2*s* 4*-polytope has a symmetric* f*-vector:* $f_0 = f_3$*,* $f_1 = f_2$*.*

Proof. If P is 2-simplicial, then each 2-face has three edges. Thus the number of incidences between 2-faces and edges, denoted f_{12}, is $f_{12} = 3f_2$. If it is 2-simple, then each edge lies in three 2-faces, that is, the number of indicences is $f_{12} = 3f_1$. Combination of the two conditions forces $f_1 = f_2$. With this, Euler's equation yields $f_0 = f_3$. □

This proof may be rephrased in terms of the face lattice: For 4-polytopes the 2s2s conditions force the two middle rank levels of the face lattice to form a bipartite cubic graph — which as any other regular bipartite graph has to have the same number of vertices on each shore. You should identify this bipartite cubical graph in the face lattice of the hypersimplex, as displayed in Figure 3.5, and thus verify the 2s2s property for this face lattice. The symmetry of the f-vector $(10, 30, 30, 10)$ is explained by Lemma 3.5; nevertheless, the hypersimplex and its face lattice are not self-dual: There are two types of facets, but only one symmetry class of vertices.

The fact that the dual of any 2s2s 4-polytope is again 2s2s (by definition), and the symmetry property for the f-vector, might suggest that 2s2s polytopes live in some sense "between" simple and simplicial. This is not true, as we will see in the next lecture, when we locate their f-vectors in the cone of all f-vectors of 4-polytopes. Indeed, the 2s2s polytopes are so interesting because they form a class of extremal polytopes in terms of the flag vector: A 4-polytope is 2s2s if and only if the valid inequality

$$2f_{03} \;\geq\; (f_1 + f_2) + 2(f_0 + f_3)$$

holds with equality. (Compare Exercise 3.7.)

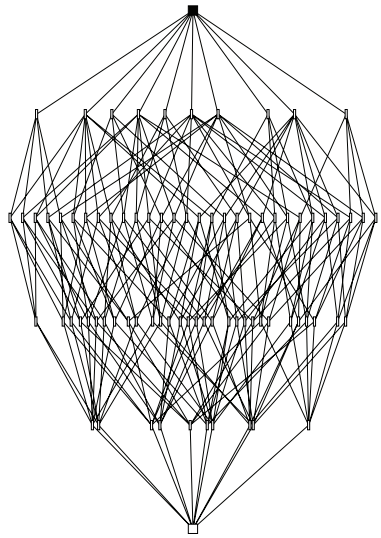


Figure 3.5. The face lattice of the hypersimplex

3.3. Deep Vertex Truncation

The idea for "deep vertex truncation" is very easy: Cut off all vertices of a polytope — but don't just truncate the vertices, but cut them off by "deep cuts," that is, so deeply that exactly one point remains from each edge.

All that is said and done about "deep vertex truncation" in the following works and makes sense for $d \geq 3$. Nevertheless, the pictures will primarily represent the case $d = 3$, while the most interesting results appear for $d = 4$.

Definition 3.6 (Deep vertex truncation)**.** Let P be a d-polytope, $d \geq 2$.

A *deep vertex truncation*

$$\mathrm{DVT}(P) \;=\; P \;\cap \bigcap_{v \in V(P)} H_v^-$$

of P is obtained by cutting off all the vertices $v \in V(P)$ of P (by closed halfspaces H_v^-, one for each vertex v) in such a way that from each edge e of P, exactly one (relative interior) point p_e remains.

Equivalently, a deep vertex truncation is obtained as the convex hull

$$\mathrm{DVT}(P) \;=\; \mathrm{conv}\left\{p_e : e \in E(P)\right\}$$

of points p_e placed on the edges $e \in E(P)$ of P in such a way that for each vertex of P, the points p_e chosen on the edges adjacent to v lie on a hyperplane H_v.

It is quite obvious that a deep vertex truncation $\mathrm{DVT}(P)$ can be constructed for each simple polytope P, but we will be particularly interested in the case of simplicial polytopes: For these it is not so clear that the cutting can be performed so that all constraints are satisfied simultaneously.

Lemma 3.7. *Every* 3*-polytope has a realization for which deep vertex truncation can be performed.*

Proof. Take an edge-tangent Koebe–Andreev–Thurston representation (according to Lecture 1). Then p_e can be taken as the tangency points, and the cutting hyperplanes H_v are spanned by the vertex horizon circles. □

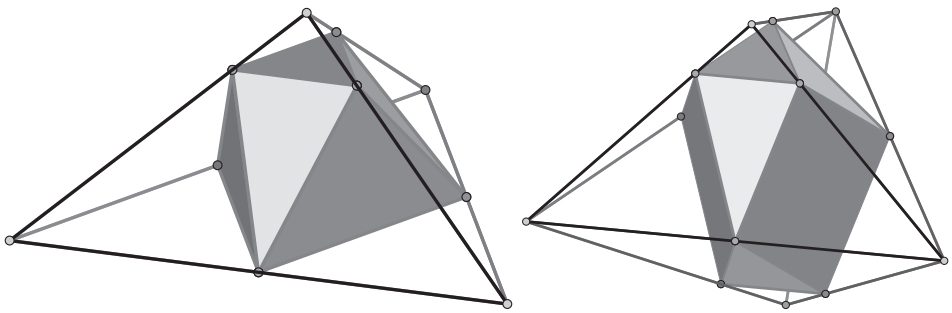


Figure 3.6. Deep vertex truncation of a simplex and of a bipyramid yields an octahedron, and a polytope that is "glued" from two octahedra. (Pictures from [**57**])

For $d \geq 3$, every deep vertex truncation polytope $\mathrm{DVT}(P)$ has two types of facets:

- deep vertex truncations $\mathrm{DVT}(F)$ of the facets F of P, and
- the vertex figures $P \cap H_v = \mathrm{conv}\{p_e : e \ni v\}$ of P.

Proposition 3.8 (Paffenholz & Ziegler [**57**])**.** *If P is a simplicial 4-polytope, then any deep vertex truncation* $\mathrm{DVT}(P)$ *is 2-simple and 2-simplicial.*

Proof. The two types of facets of $\mathrm{DVT}(P)$ are the octahedra $\mathrm{DVT}(F)$, for the tetrahedron facets F of P, and the vertex figures of P, which are simplicial. Thus $\mathrm{DVT}(P)$ 2-simplicial.

Since all edges of P are reduced to points by deep vertex truncation, all the edges of $\mathrm{DVT}(P)$ are "new," they arise by deep vertex truncation from the 2-faces (that is, the ridges) of P. Each such ridge lies in two facets F_1, F_2 of P, so the edge we are looking at lies in two facets $\mathrm{DVT}(F_1)$ and $\mathrm{DVT}(F_2)$ of the first type, and in one facet of the second type. Thus each edge of $\mathrm{DVT}(P)$ lies in exactly three facets, that is, $\mathrm{DVT}(P)$ is 2-simple. □

So we have that $\mathrm{DVT}(P)$ is 2s2s for any simplicial 4-polytope $P \ldots$ if it exists. And that's the problem: In general it is not at all guaranteed that deep vertex truncation can be performed. One would try to realize cyclic 4-polytopes in such a way that deep vertex truncations can be performed, but it seems that this is not possible. Similarly, if a sum $P_m \oplus P_n$ is realized "the obvious way," with regular polygons in orthogonal subspaces, then deep vertex truncation is not possible except for very special cases (such as $\frac{1}{m} + \frac{1}{n} \geq \frac{1}{2}$): It is quite surprising that the sums of polygons *do* have a realization such that deep vertex truncation is possible, as proved by Paffenholz [**55**]. On the other hand, there does not seem to be a single example of a simplicial polytope for which it has been *proved* that deep vertex truncation is impossible for all realizations.

However, in special cases deep vertex truncation can indeed be performed. In particular, any *regular* polytope admits a deep vertex truncation — just take the

edge midpoints for p_e. From this we get the following three examples of 2s2s 4-polytopes:

- Deep vertex truncation of a simplex, $\mathrm{DVT}(\Delta_4)$, yields the hypersimplex.
- Deep vertex truncation of the 4-dimensional cross polytope,
$$C_4^* \;=\; \mathrm{conv}\{\pm e_i : 1 \le i \le 4\} \;=\; \{x \in \mathbb{R}^4 : |x_1| + |x_2| + |x_3| + |x_4| \le 1\},$$
yields Schläfli's 24-cell (see Figure 3.7):
$$\begin{aligned}\mathrm{DVT}(C_4^*) &= \mathrm{conv}\{\pm\tfrac{1}{2}e_i \pm \tfrac{1}{2}e_j : 1 \le i < j \le 4\}\\ &= \{x \in \mathbb{R}^4 : |x_i| \le 1 \text{ for } 1 \le i \le 4,\ |x_1| + |x_2| + |x_3| + |x_4| \le 1\}.\end{aligned}$$
- Deep vertex truncation of the regular 600-cell (which has 600 regular tetrahedra as facets) yields a 2s2s 4-polytope with 720 vertices, whose vertex figures are prisms over regular pentagons; its facets are 600 octahedra, and 120 regular icosahedra. It seems that this remarkable polytope, with f-vector (720, 3600, 3600, 720), first occured in the literature in 1994, as the dual of the "dipyramidal 720-cell" constructed by Gévay [**36**]. See also Exercise 3.2.

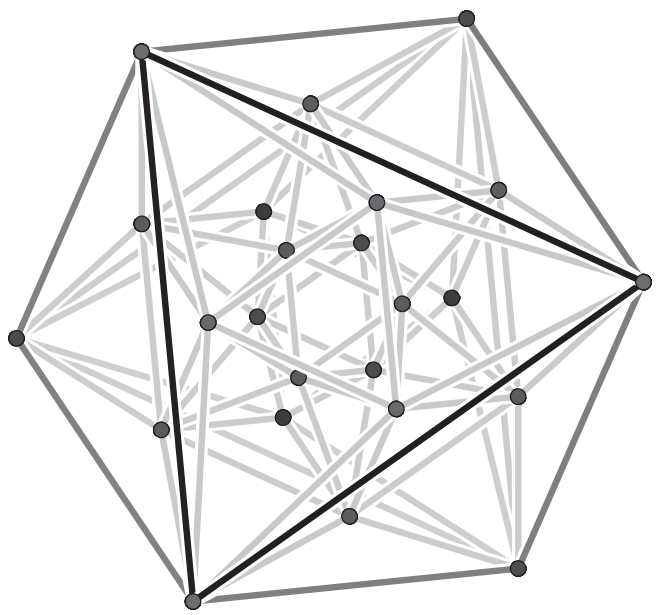


Figure 3.7. The 24-cell

3.4. Constructing DVT(Stack(n, 4))

The stacked polytopes form an infinite family of simplicial polytopes which can quite easily be realized in such a way that deep vertex truncation can be performed.

For this, we denote by $\mathrm{Stack}(n,d) := \mathrm{stack}^n(\Delta_d)$ any combinatorial type of a d-polytope, $d \ge 3$, which is obtained by n times stacking a pyramid onto a simplex facet, starting at a d-simplex. This is a simplicial d-polytope with $d+1+n$ vertices and $d+1+n(d-1)$ facets; see Exercise 3.3. Note that the notation "$\mathrm{Stack}(n,d)$" does not specify a combinatorial type; many different types may be obtained by stacking onto different sequences of facets (cf. Exercise 3.6).

Theorem 3.9 (Paffenholz & Ziegler [**57**]). *Any combinatorial type of a stacked d-polytope* $\mathrm{Stack}(n,d)$ *can be realized so that it admits a deep vertex truncation.*

Proof. We proceed by induction on n, starting at $n=0$, with a d-simplex, and a deep vertex truncation that takes the convex hull of the edge midpoints.

Assume now that $\mathrm{Stack}(n,d)$ has been realized as $P \subset \mathbb{R}^d$ such that $\mathrm{DVT}(P)$ can be obtained by a suitable choice of points p_e on the edges $e \subset P$. Assume that $\mathrm{Stack}(n+1,d)$ arises by stacking onto a facet of $\mathrm{Stack}(n,d)$ that is realized by the facet $F \subset P$ with vertex set $\{v_1,\ldots,v_d\}$. The "new" vertex w is now chosen "beyond" the facet $\mathrm{DVT}(F)$ of $\mathrm{DVT}(P)$, and "beneath" all other facets of $\mathrm{DVT}(P)$. That is, addition of w to $\mathrm{DVT}(P)$ would mean stacking a pyramid onto the facet $\mathrm{DVT}(F)$ of $\mathrm{DVT}(P)$. In particular, w lies "beyond" the facet F of P, and "beneath" all other facets of P, so $P' := \mathrm{conv}(\{w\} \cup P)$ is a stacked polytope realizing $\mathrm{Stack}(n+1,d)$, as required.

The facet hyperplanes H_{v_i} of $\mathrm{DVT}(P)$ cut the edges $[v_i,w]$ of P' in points p_i: This is since w is beneath H_{v_i}, while v_i is cut off by H_{v_i}. Thus we obtain points p_i on the new edges of P', and the hyperplane $H_w := \mathrm{aff}\{p_1,\ldots,p_d\}$ may be taken to cut off the new vertex w of P'. This new truncation plane is determined uniquely by the d intersection points, because the new vertex w of P' is simple. □

This theorem is valid for all $d \geq 3$; in particular, 3D-pictures work. (Figure 3.8 is a feeble attempt.) However, the construction produces by far the most interesting results for $d=4$.

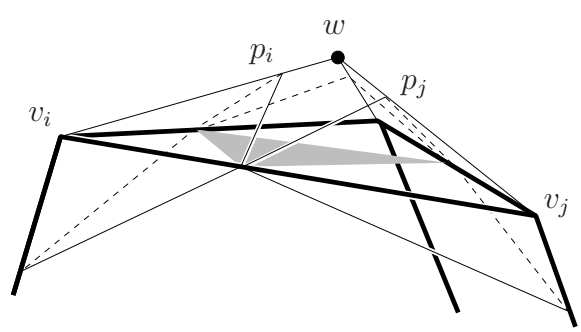


Figure 3.8. The induction step in Theorem 3.9, for $d=3$. $\mathrm{DVT}(F)$ is drawn shaded.

Corollary 3.10 ([**57**]). *For each $n \geq 0$, and for every type of stacked 4-polytope* $\mathrm{Stack}(n,4)$ *with f-vector* $(5+n, 10+4n, 10+6n, 5+3n)$*, there is a corresponding 2-simple 2-simplicial 4-polytope* $\mathrm{DVT}(\mathrm{Stack}(n,4))$*, with f-vector*

$$f(\mathrm{DVT}(\mathrm{Stack}(n,4))) \;=\; (10+4n, 30+18n, 30+18n, 10+4n).$$

In particular, this yields infinitely many combinatorial types of 2-simple 2-simplicial 4-polytopes. Moreover, with a bit of care the proof of Theorem 3.9 yields these polytopes with rational vertex coordinates. See [**54**] for explicit examples of such coordinates.

Corollary 3.11 ([**57**]). *The number of combinatorial types of 2-simple 2-simplicial 4-polytopes with $10+4n$ vertices grows exponentially in n.*

See Paffenholz & Werner [**56**] for further constructions of 2-simple 2-simplicial 4-polytopes with interesting f-vectors. In particular, they describe the "smallest" example of such a polytope (other than the simplex), which has only 9 vertices.

Exercises

3.1. Show that any simple or simplicial d-polytope with $f_0 = f_{d-1}$ must be a simplex, or 2-dimensional.

3.2. Compute the full f-vectors, as well as the number f_{03} of vertex-facet incidences, for the following 4-polytopes, based only on the information given here:
(a) The 24-cell: a 2s2s polytope whose facets are 24 octahedra;
(b) The 600-cell: a simple polytope whose facets are 120 dodecahedra;
(c) The 720-cell: a 2s2s 4-polytope whose facets are 720 bipyramids over pentagons;
(d) A neighborly cubical polytope NCP_4^n, a cubical polytope with the graph of the n-cube ($n \geq 4$).

3.3. Compute the full f-vectors of the stacked d-polytopes $\mathrm{Stack}(n, d)$.

3.4. Show that if a 4-polytope P is not simplicial, then $\mathrm{DVT}(P)$ cannot be 2-simplicial.

3.5. Find coordinates for $\mathrm{DVT}(\mathrm{Stack}(1, 4))$. Check them with `polymake`.
(This is Braden's "glued hypersimplex" [**18**].)

3.6. Show that there are exponentially many distinct combinatorial types of stacked d-polytopes with $d + 1 + n$ vertices, for any $d \geq 3$. Derive that there are exponentially many types of 2-simple 2-simplicial 4-polytopes with the same f-vector.

3.7. Show that $f_{13} = f_{03} + 2f_2 - 2f_3$, and dually $f_{02} = f_{03} + 2f_1 - 2f_0$, holds for the flag vector of each 4-polytope.
(Hint: Sum the Euler equations for the facets, which are 3-polytopes.)
Derive from this that the inequality $2f_{03} \geq (f_1 + f_2) + 2(f_0 + f_3)$ is valid for all 4-polytopes, and that it is tight exactly for the 2-simple 2-simplicial 4-polytopes.

3.8. Show that there is no f-vector inequality (not involving f_{03}) that characterizes the 2s2s 4-polytopes.

3.9. If P is a d-dimensional simplicial polytope, and if $\mathrm{DVT}(P)$ exists, is $\mathrm{DVT}(P)$ then 2-simple? 2-simplicial?

LECTURE 4
f-Vectors of 4-Polytopes

The f-vector of a 4-polytope is a quadruple of integers $f(P) = (f_0, f_1, f_2, f_3)$, but due to the Euler-Poincaré relation the set of all f-vectors of 4-polytopes is a 3-dimensional set: It lies on the "Euler-Poincaré hyperplane" in $\mathbb{R}^4$, given by

$$f_0 - f_1 + f_2 - f_3 \;=\; 0.$$

The task we are facing is to describe the set of all f-vectors,

$$\mathcal{F}_4 \;:=\; \{f(P) \;=\; (f_0, f_1, f_2, f_3) \in \mathbb{Z}^4 : P \text{ a convex 4-polytope}\}.$$

Here "describe" may mean a number of different things: Probably one should not hope for a complete description (as Steinitz got for the 3-dimensional case), since the set of f-vectors is way more complicated in the 4-dimensional case.

Indeed, $\mathcal{F}_4$ is *not* the set of all integral points in a polyhedral cone, or even in a convex set. This may be seen from the characterizations of the projections of $\mathcal{F}_4$ to the coordinate 2-planes in $\mathbb{R}^4$, by Grünbaum, Barnette, and Reay [**39**, Sect. 10.4] [**8**] [**10**], which show non-convexities and holes (see Figure 4.1). Or you just note that some of the rather basic, tight inequalities, such as the upper bound inequality $f_1 \le \binom{f_0}{2}$, are concave. For example,

$$\begin{aligned} f(C_4(5)) &= (5, 10, 10,\ 5), \\ f(C_4(7)) &= (7, 21, 28, 14), \\ f(C_4(9)) &= (9, 36, 54, 27). \end{aligned}$$

The midpoint of the segment between $f(C_4(5))$ and $f(C_4(9))$ is the integral point $(7, 23, 32, 16)$: It violates the upper bound inequality, and indeed a 4-polytope with 7 vertices cannot contain more than the $21 = \binom{7}{2} = f_1(C_4(7))$ edges. (See also Bayer [**11**], Höppner & Ziegler [**42**].)

In the following, we will head for a complete description of the f-vector cone for 4-polytopes, $\mathrm{cone}(\mathcal{F}_4)$. This seems to be a challenging but realistic goal. Once that is achieved (the "2006 project"), a logical next goal might be a description of the "large" f-vectors, that is, of

$$\{f(P) = (f_0, f_1, f_2, f_3) \in \mathbb{Z}^4 : P \text{ a convex 4-polytope},\ f_0 + f_3 \ge M\}$$

for some large M. But let's not get too ambitious too fast.

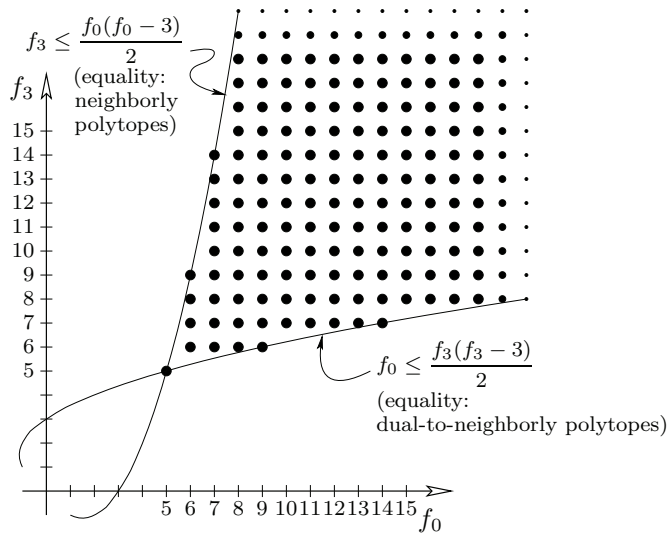


Figure 4.1. The (f_0, f_3)-pairs of convex 4-polytopes, according to Grünbaum [**39**, Sect. 10.4])

4.1. The f-Vector Cone

Definition 4.1 (f-vector cone)**.** The *f-vector cone* of 4-polytopes, $\mathrm{cone}(\mathcal{F}_4)$, is the topological closure of the convex cone with apex $f(\Delta_4) = (5, 10, 10, 5)$ that is spanned by the f-vectors of 4-polytopes,

$$\Big\{ f(\Delta_4) + \sum_{i=1}^{N} \lambda_i \big(f(P_i) - f(\Delta_4) \big) : P_1, \ldots, P_N \text{ 4-polytopes}, \ \lambda_1, \ldots, \lambda_N \ge 0 \Big\}.$$

Equivalently, $\mathrm{cone}(\mathcal{F}_4) \subset \mathbb{R}^4$ is the solution set to all the linear inequalities that are valid for all f-vectors for 4-polytopes, and that are tight at the f-vector of the simplex.

The equivalence between the two versions of the definition rests on basic facts about closed convex sets, which you should put together yourself (Exercise 4.1). You are also asked to verify that the cone generated by the f-vectors is not closed, so we do have to take the topological closure (Exercise 4.2.)

The closed convex cone we are looking at is 3-dimensional, so we may view it as the cone over a 2-dimensional convex figure, which might be just a pentagon or hexagon. Instead of looking at a 2-dimensional section (say intersecting by $f_1 + f_2 = 100$), we may equivalently introduce homogeneous ("projective") coordinates, which are rational linear functions, normalized to yield "$\frac{0}{0}$" at the f-vector of a simplex (compare Lecture 1). There is no unique best way to do this; we choose

$$\varphi_0 \ := \ \frac{f_0 - 5}{f_1 + f_2 - 20} \qquad \text{and} \qquad \varphi_3 \ := \ \frac{f_3 - 5}{f_1 + f_2 - 20}$$

as our homogeneous coordinates. (Figure 4.2 illustrates the geometry of such a rational function on a cone.) So we are trying to describe $\mathrm{proj}(\mathcal{F}_4) \subset \mathbb{R}^2$, the closure of

$$\mathrm{conv}\{(\varphi_0(P), \varphi_3(P)) \in \mathbb{R}^2 : P \text{ a convex 4-polytope}\}.$$

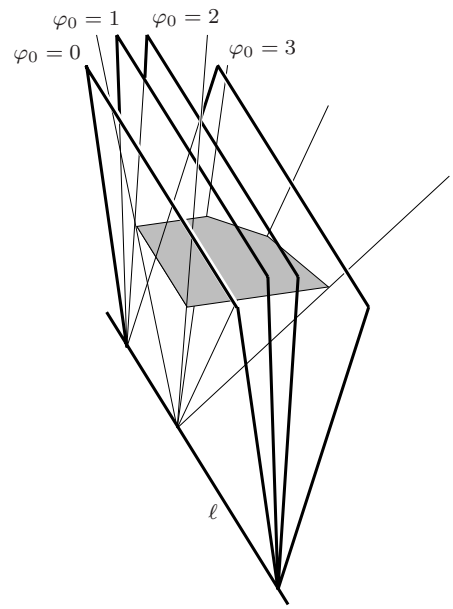


Figure 4.2. The function φ_0 is constant on certain planes that contain the apex of the cone. It is not defined on the line ℓ where all those planes intersect. (In terms of (f_0, f_1, f_2)-coordinates, ℓ is defined by $f_0 = 5$ and $f_1 + f_2 = 20$.)

Any 4-polytope yields a (rational) point in the (φ_0, φ_3)-plane. Any valid linear inequality, tight at the 4-simplex, translates into a linear inequality in φ_0 and φ_3. So let's look at some families of polytopes and of linear inequalities that we know, and let's see what they buy us.

Some 4-polytopes we know:

Stacked:	$(5+\ \ n, 10+4n, 10+6n, 5+3n)$	$\longrightarrow$	$(\frac{1}{10}, \frac{3}{10})$
Truncated:	$(5+3n, 10+6n, 10+4n, 5+\ \ n)$	$\longrightarrow$	$(\frac{3}{10}, \frac{1}{10})$
Cyclic:	$(n, \frac{n(n-1)}{2}, n(n-3), \frac{n(n-3)}{2})$	$\overset{n\to\infty}{\longrightarrow}$	$(0, \frac{1}{3})$
Dual-to-cyclic:	$(\frac{n(n-3)}{2}, n(n-3), \frac{n(n-1)}{2}, n)$	$\overset{n\to\infty}{\longrightarrow}$	$(\frac{1}{3}, 0)$.

The truncated polytopes are the duals of the stacked polytopes, so they are simple. Similarly, the duals of cyclic polytopes are simple. Thus we find the four points $(\frac{1}{10}, \frac{3}{10})$, $(\frac{3}{10}, \frac{1}{10})$, $(0, \frac{1}{3})$, and $(\frac{1}{3}, 0)$, which span a quadrilateral subset of $\mathrm{proj}(\mathcal{F}_4)$. This quadrilateral also represents the f-vectors of simple and of simplicial polytopes and "everything in between." (Note that duality interchanges the coordinates φ_0 and φ_3, and thus $\mathrm{proj}(\mathcal{F}_4)$ is symmetric with respect to the main diagonal.)

Five linear constraints we know:

"Few Vertices":	$f_0 \geq 5$	$\Longleftrightarrow$	$\varphi_0 \geq 0$,
"Few Facets":	$f_3 \geq 5$	$\Longleftrightarrow$	$\varphi_3 \geq 0$,
"Simple":	$f_1 \geq 2f_0$	$\Longleftrightarrow$	$3\varphi_0 + \varphi_3 \leq 1$,
"Simplicial":	$f_2 \geq 2f_3$	$\Longleftrightarrow$	$\varphi_0 + 3\varphi_3 \leq 1$,
"Lower bound":	$2f_1 + 2f_2 \geq 5f_0 + 5f_3 - 10$	$\Longleftrightarrow$	$\varphi_0 + \varphi_3 \leq \frac{2}{5}$.

The first four inequalities are quite trivial, and we have named them by the polytopes that satisfy them with equality, at least asymptotically. The translation into (φ_0, φ_3)-inequalities, using the Euler-Poincaré relation, poses no problem. There is no polytope with $\varphi_0 = 0$, but the condition is satisfied asymptotically by any family of 4-polytopes with far more vertices than facets. For example, the products of n-gons, with $f(P_n \times P_n) = (n^2, 2n^2, n^2 + 2n, 2n)$, yield $(\varphi_0, \varphi_3) = (\frac{n^2-5}{3n^2+2n-20}, \frac{2n-5}{3n^2+2n-20}) \in \operatorname{proj}(\mathcal{F}_4)$, which in the limit $n \to \infty$ yields $(\frac{1}{3}, 0)$.

The one non-trivial inequality in our table above is the last one, a "Lower Bound Theorem." It may be derived quite easily [**11**] from the inequality $f_{03} \geq 3f_0 + 3f_3 - 10$, which was first established by Stanley [**70**] in terms of the so-called toric g-vector (it is the inequality "$g_2^{\mathrm{tor}}(P) \geq 0$"); a proof via rigidity theory was later given by Kalai [**44**].

Figure 4.3 summarizes our discussion up to this point: We are interested in $\operatorname{proj}(\mathcal{F}_4)$, the closure of the set

$$\operatorname{conv}\{(\varphi_0(P), \varphi_3(P)) : P \text{ is a 4-polytope, not a simplex}\} \subset \mathbb{R}^2.$$

This set is contained in the pentagon cut out by the five linear inequalities discussed above, and it contains the shaded trapezoid, which represents "everything between simple and simplicial polytopes." Indeed, simple and simplicial polytopes satisfy the additional linear inequality $\varphi_0 + \varphi_3 \geq \frac{1}{3}$.
Thus we are left with the following "upper bound problem":

"Upper Bound Problem". Are there 4-polytopes with $\varphi_0 + \varphi_3 \to 0$?

The inequality $\varphi_0 + \varphi_3 \geq \frac{1}{3}$ is certainly not valid for all (possibly non-simple non-simplicial) 4-polytopes: Already for the hypersimplex we get $(\varphi_0, \varphi_3) = (\frac{1}{8}, \frac{1}{8})$.

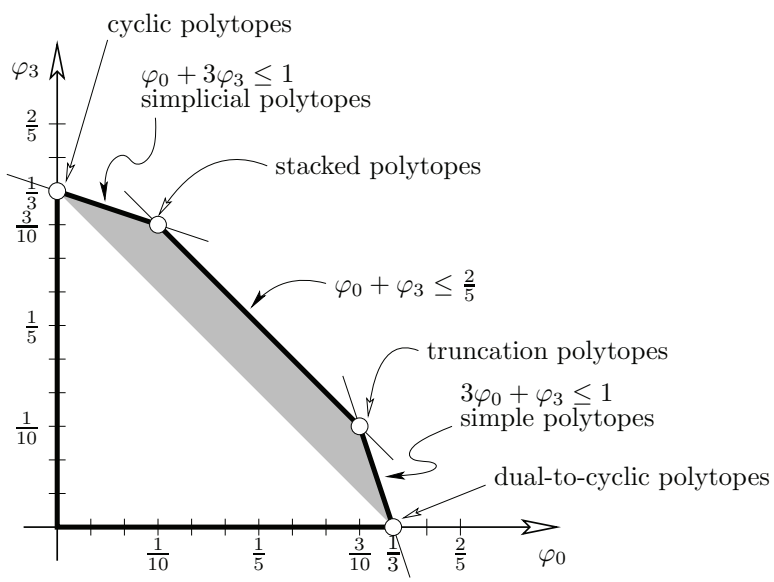


Figure 4.3. Projective representation of the f-vectors of 4-polytopes, in the (φ_0, φ_3)-plane. The convex set $\operatorname{proj}(\mathcal{F}_4)$ is contained in the bold pentagon; it contains the shaded trapezoid.

However, currently it is not clear how small $\varphi_0 + \varphi_3$ can be for convex polytopes. Thus the Upper Bound Problem is the key remaining problem in the description of the f-cone for 4-polytopes.

(!) If the answer is YES to the problem as posed above, then the five inequalities above constitute a complete linear description of cone($\mathcal{F}_4$).

(!) If the answer is NO, then this is also exciting, since it means that the answers for cellular spheres and for convex polytopes are distinct! Indeed, cellular 3-spheres with arbitrarily small $\varphi_0 + \varphi_3$ have been constructed by Eppstein, Kuperberg & Ziegler [**26**]; see our discussion in Section 4.3.

4.2. Fatness and the Upper Bound Problem

We prefer to rephrase the Upper Bound Problem in terms of a somewhat more graphic quantity, which we call "fatness."

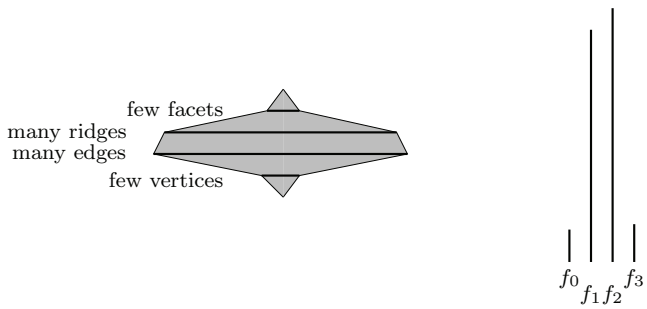


Figure 4.4. Fatness for a 4-polytope face lattice, and for an f-vector

Definition 4.2 (Fatness)**.** The *fatness* of a 4-polytope is the quotient

$$F(P) \;:=\; \frac{1}{\varphi_0 + \varphi_3} \;=\; \frac{f_1 + f_2 - 20}{f_0 + f_3 - 10}.$$

The fatness of a 4-polytope is large if both φ_0 and φ_3 are small. This happens if the polytope has relatively few vertices and facets, but many edges and 2-faces. Thus, graphically, the face lattice and the f-vector are "fat in the middle," whence the name (see Figure 4.4).

"Upper Bound Problem". Can the fatness of a 4-polytope be arbitrarily large?

Here are a few explicit values to start with: For stacked and truncated 4-polytopes we have $F(P) = \frac{5}{2}$ exactly. For cyclic polytopes we get $F(C_4(n)) \to 3$ for $n \to \infty$, and the same for the duals — fatness is a self-dual quantity, that is, any 4-polytope and its dual have the same fatness. Moreover, it is easy to compute (or to derive from Figure 4.3) that all simple and simplicial polytopes satisfy $\frac{5}{2} \leq F(P) < 3$.

But how large can fatness be? The attempts to answer this question have led to a multitude of interesting examples and constructions, and to a fast succession of record holders for "the fattest examples found so far." Many of them can be obtained by deep vertex truncation of simplicial polytopes, so they are 2-simple and 2-simplicial by Proposition 3.8:

– The hypersimplex, which is the dual of DVT(Δ_4), has fatness 4.
– Schäfli's 24-cell [**66**], DVT(cross polytope), has fatness 4.526.
– Gévay's 720-cell [**36**], the dual of DVT(120-cell), has 720 facets that are bipyramids over regular pentagons. It has fatness 5.020.
– Eppstein, Kuperberg & Ziegler [**26**] used hyperbolic geometry arguments to achieve a fatness of 5.048 by the "E-construction," which is dual to deep vertex truncation.
– Paffenholz [**55**] has very recently shown that there are realizations for any sum of an n-gon and m-gon such that the deep vertex truncation DVT($P_m \times P_n$) can be obtained. For $m = n \to \infty$ this yields fatness approaching 6.

However, we'll go a different route. In the next and final lecture we will present a construction that generalizes and extends the construction of "neighborly cubical" 4-polytopes of Joswig & Ziegler [**43**], to achieve fatness arbitrarily close to 9, the latest record (as far as I know at the time of writing). I would have been happy to have a "note added in proof" about this ...

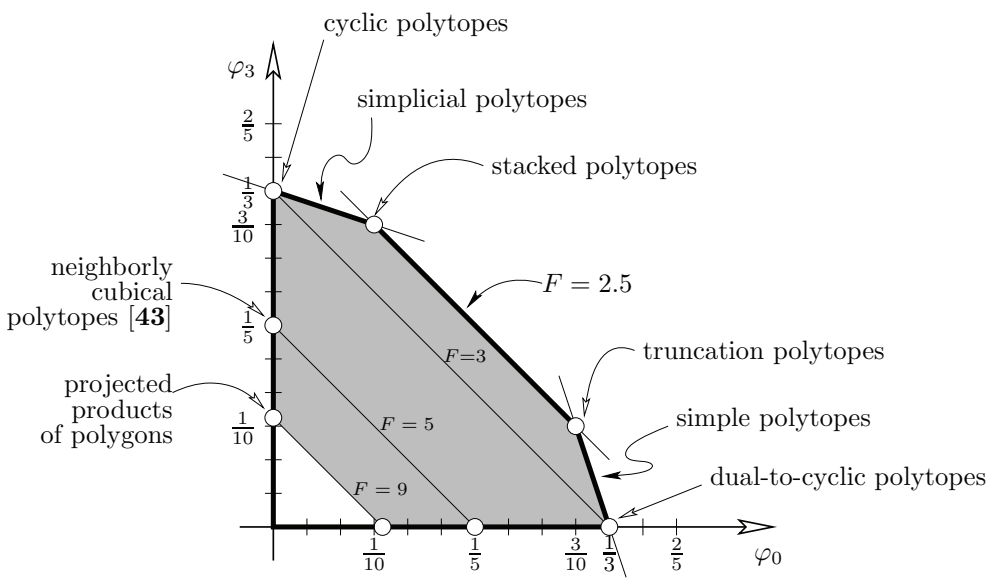


Figure 4.5. 4-polytopes in the (φ_0, φ_3)-plane. The shaded hexagon is spanned by the (φ_0, φ_3)-pairs of known 4-polytopes.

What do polytopes "of very high fatness" look like? You can verify (via Exercise 4.6) that they have two properties:
(1) The facets have many vertices (on average).
(2) The vertices are in many facets (on average).

Either of these properties are easy to satisfy — just look at the products $P_n \times P_n$ for the first property, and at their duals, the free sums $P_n \oplus P_n$, for the second one. The key question is whether they can *simultaneously* be satisfied.

Finally, here is a problem on 3-dimensional polytopal tilings that is "essentially" equivalent to the fatness problem: Consider face-to-face tilings of $\mathbb{R}^3$ (cf. [**65**]) that satisfy some regularity properties, e.g. one of the following (each implies the next):

– the tiling is triply periodic (that is, there are three linearly independent translational symmetries),
– there are only finitely many distinct congruence classes of tiles,
– in- and circumradius of the tiles are uniformly bounded.

For such tilings, we may define notions of "average" vertex degrees, face numbers, etc. The question is whether there is such a tiling where the tiles have lots of vertices on average, *and* the vertices are in many tiles on average. Again, either property is easy to achieve (look at tilings by Schlegel diagrams), but can they be *simultaneously* satisfied?

4.3. The Lower Bound Problem

The upper bound problem discussed here has a natural "lower bound" counterpart. It arises if we don't restrict ourselves to the geometric model of convex polytopes, but consider the larger class of cellular spheres that are "regular" in the sense that their cells have no identifications on the boundary, and that satisfy the "intersection property" that any two faces should intersect in a single cell (which may be empty). These are the regular CW spheres [**21**] whose face poset is a lattice (where the meet operation corresponds to intersection of faces).

"Lower Bound Problem". Does $\varphi_0 + \varphi_3 \leq \frac{2}{5}$ hold for the cellular spheres that satisfy the intersection property?

This problem seems crucial in terms of the separation of the "geometric" model of convex polytopes from the "topological" model of cellular spheres/balls.

(!) If the answer to the problem is NO, then this would establish such a separation, which would be quite remarkable.
(!) If the answer is YES, then this would imply a complete characterization of the f-vector cone for cellular 3-spheres, by the five linear inequalities given above; indeed, Eppstein, Kuperberg & Ziegler [**26**] have constructed cellular spheres for which fatness is arbitrarily large, that is, $\varphi_0 + \varphi_3$ is arbitrarily small.

We will not discuss this here further, but refer to [**26**] and [**80**].

Exercises

4.1. Show that the two definitions of the f-vector cone given in Definition 4.1 are indeed equivalent.
Hint: You need a separation lemma; see for example Matoušek [**48**, p. 6].

4.2. Show that the union of the line segments $[f(\Delta_4), f(C_4(n))] \subset \mathbb{R}^4$ has the whole ray $\{(5, 10+t, 10+2t, 5+t) : t \geq 0\}$ in its closure. Note that $f_0 \geq 5$ is a valid linear inequality, which is tight at $f(\Delta_4)$, but for no other f-vector.
Conclude that the cone with apex $f(\Delta_4)$ spanned by the f-vectors of 4-polytopes is not closed.

4.3. Compute the fatness and the (φ_0, φ_3)-pair for the hypersimplex, the 24-cell, and for DVT(600-cell).

4.4. Compute the fatness of the 2s2s polytopes DVT(Stack$(n, 4)$), and show that it lies in the interval [4, 4.5).
Show that for any simplicial 4-polytope P, the fatness of DVT(P) is smaller than 6.
Where would the f-vectors of the polytopes DVT(P) lie in proj$(\mathcal{F}_4)$, as graphed in Figure 4.5?

4.5. If C_4^n is a cubical 4-polytope with the graph of an n-cube (see Exercise 3.2), compute the fatness and the pair (φ_0, φ_3).

4.6. Define the *complexity* of a 4-polytope to be the quotient

$$C(P) \ := \ \frac{f_{03} - 20}{f_0 + f_3 - 10}.$$

(a) Show that $F(P) \leq 2C(P) - 2$, with equality if and only if P is 2-simple and 2-simplicial.

(b) Show that $C(P) \leq 2F(P) - 2$, with equality if and only if if all facets of P are simple, or equivalently, if all vertex figures are simple.

(c) Derive from this that fatness is high if and only if both the average number of vertices per facet, f_{03}/f_3, and the average number of facets per vertex, f_{03}/f_0, are large.

LECTURE 5
Projected Products of Polygons

In this lecture we present a construction of very recent vintage, "projected products of polytopes." We will not have the ambition to work through all the technical details for this; these appear in [**81**], see also [**63**]; rather, our main objective is here to identify the structural features of the construction which lead to fat polytopes, and to outline (possibly "for further use") some interesting components that go into the construction. In the following version of the main result, some concepts that will be explained below are highlighted by quotation marks.

Theorem 5.1 (Ziegler [**81**]). *For each $r \geq 2$, and even $n \geq 4$, there is a realization $P_n^r \subset \mathbb{R}^{2r}$ of a product of polygons $(P_n)^r$ (a "deformed product of polygons") such that the vertices and edges and all the "n-gon 2-faces" of P_n^r "survive" the projection $\pi : \mathbb{R}^{2r} \to \mathbb{R}^4$ to the last 4 coordinates.*

A number of nice tricks go into the construction that proves the theorem — see below. Before we look into these we want to derive the enumerative consequences: The f-vector of $\pi(P_n^r)$ can be derived purely from the information given in the theorem, not using details about the combinatorics of the resulting polytopes (which were worked out only recently [**62**] [**63**]).

5.1. Products and Deformed Products

We have discussed the construction and main properties of products of polytopes already in Example 2.7. A key observation is that the non-empty faces of a product are the products of non-empty faces of the "factors." Now we specialize to the case of products of (several) polygons: We consider products of r n-gons — and later we will be looking at polytopes that just have the combinatorics of such polytopes.

If P_n is an n-gon, then $(P_n)^r$ is a simple polytope of dimension $2r$. It has

- $f_0 = n^r$ vertices (of the form "vertex $\times$ vertex $\times$ ..... $\times$ vertex"), and
- $f_1 = rn^r$ edges (of the form "vertex $\times \ldots \times$ edge $\times \ldots \times$ vertex).

The products of polygons have two different types of 2-faces, "quadrilaterals" and "polygons," that we need to distinguish:

- $\binom{r}{2}n^r$ *quadrilateral* 2-faces (which arise as products of two edges, and vertices from the other factors), and
- rn^{r-1} *polygon* 2-faces (arising as a product of one n-gon factor with vertices from the other factors).

Thus we get $f_2 = \binom{r}{2}n^r + rn^{r-1}$.

Finally, let's note that $(P_n)^r$ has nr facets, which arise as a product of one edge (from one of the factors) with n-gons as the other factors.

The "deformed products" of polygons P_n^r considered below are combinatorially equivalent to the "orthogonal products" $(P_n)^r$. Thus, the f-vector count given here is valid for deformed products as well.

5.2. Computing the f-Vector

It is remarkable that for the proof of the following corollary to Theorem 5.1 we don't need detailed combinatorial information about $\pi(P_n^r)$; it is sufficient to know that it is a generic projection of a deformed product P_n^r, and that all the polygon 2-faces survive the projection. The facets of $\pi(P_n^r)$ are 3-faces of P_n^r that "have survived the projection," that is, they are combinatorial cubes and prisms over n-gons.

Corollary 5.2. *$\pi(P_n^r) \subset \mathbb{R}^4$ is a 4-polytope with f-vector*

$$\begin{aligned} & \left(n^r, rn^r, \tfrac{5}{4}rn^r - \tfrac{3}{4}n^r + rn^{r-1}, \tfrac{1}{4}rn^r - \tfrac{1}{2}n^r + rn^{r-1}\right) \\ = \; & \left(0 + \tfrac{4}{r}, 4, 5 - \tfrac{3}{r} + \tfrac{4}{n}, 1 - \tfrac{2}{r} + \tfrac{4}{n}\right) \cdot \tfrac{1}{4}rn^r. \end{aligned}$$

In particular, for $n, r \to \infty$ the fatness of the projected products of polygons $\pi(P_n^r)$ gets arbitrarily close to 9.

Proof. From the combinatorial information above we see that each vertex, and each edge, of a product of polygons $(P_n)^r$ is contained in a polygon 2-face: So if all the polygon 2-faces survive the projection, then in particular all vertices and edges do. Thus, for $(f_0, f_1, f_2, f_3) := f(\pi(P_n^r))$ we get $f_0 = n^r$, and $f_1 = rn^r$.

Furthermore, we know that $\pi(P_n^r)$ has $p := rn^{r-1}$ polygon 2-faces (and a yet unknown number of quadrilaterals). Moreover, the facets of $\pi(P_n^r)$ are prisms over polygons, and cubes. Each prism has two polygon faces (and n quadrilateral faces), while each cube has 6 quadrilateral 2-faces, but no polygon faces. So, each polygon (ridge!) lies in two prism facets, and each prism facet contains two polygons: Double counting yields that there are exactly $p = rn^{r-1}$ prism facets.

Denote by c the number of cube facets. Then we have

- $f_3 = c + p$ (all facets are cubes or prisms),
- $2f_2 = 6c + (n+2)p$ (double counting the facet-ridge incidences), and
- $f_2 - f_3 = (r-1)n^r$ (Euler's equation).

These three linear equations can now be solved for the three unknowns, f_2, f_3, and c. □

5.3. Deformed Products

A number of different polytope constructions have been studied in attempts to produce interesting polytopes. For example, *any* polytope may be obtained by projection of a simplex; on the other hand, any projection of an orthogonal cube is a zonotope, which is a polytope of very special structure (see [**79**, Lect. 7]). The projections of orthogonal products of centrally symmetric polygons are still zonotopes, and the projections of orthogonal products of arbitrary polygons are only a bit more general. However, it has been noted since the seventies (in the context of linear programming, trying to construct "bad examples" for the simplex

algorithm; cf. [**45**] [**3**]) that some projections of "deformed products" have very interesting extremal properties. For a very simple example, look at a 3-cube (which is a product of three 1-polytopes). Any orthogonal cube projected to the plane will produce a hexagon (at best), while a deformed cube can be projected to the plane to yield an octagon: All the vertices "survive the projection" (see Figure 5.1).

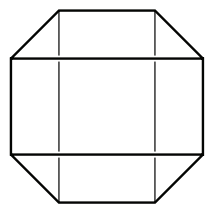


Figure 5.1. A combinatorial 3-cube, in a deformed realization such that all eight vertices "survive" the projection to the plane

And indeed, it is not so hard to show that one can realize an n-cube in such a way that all its vertices "survive the projection" to the plane — this was first proved by Murty [**51**] and Goldfarb [**37**]; in the context of linear programming it yields exponential examples for the "shadow vertex" pivot rule for linear programming.

Similarly, if you project an orthogonal product of polygons to $\mathbb{R}^4$, you cannot expect that all the edges survive the projection, but with deformed products, this is possible (although hard to visualize — proofs are mostly based on linear algebra criteria rather than on geometric intuition).

Here are linear algebra descriptions of the polytopes we'll be looking at:

Polygons: If V is any $(n \times 2)$-matrix whose rows are non-zero, (w.l.o.g.) ordered in cyclic order, and positively span $\mathbb{R}^2$, then for a *suitable* positive right-hand side vector $b \in \mathbb{R}^n$ the system $Vx \leq b$ describes a convex n-gon $P_n \subset \mathbb{R}^2$.

Product of polygons: Given such a V, we immediately get the system

$$\begin{pmatrix} V & & & & \\ & V & & 0 & \\ & & V & & \\ 0 & & & \ddots & \\ & & & & V \end{pmatrix} x \leq \begin{pmatrix} b \\ b \\ b \\ \vdots \\ b \end{pmatrix},$$

with a block-diagonal matrix of size $rn \times 2r$, which describes the orthogonal product of polygons $(P_n)^r \subset \mathbb{R}^n$.

The combinatorics of the product of polygons is reflected in the facet-vertex incidences, as follows: Each inequality defines a facet; each of the n^r vertices is the unique solution of a linear system of equations that is obtained by requiring that from each block, two cyclically adjacent inequalities are tight.

Deformed product of polygons: Our Ansatz is as follows:

$$\left(\begin{array}{cccccc} V^\varepsilon & & & & & \\ W & V^\varepsilon & & 0 & & \\ U & W & V^\varepsilon & & & \\ & U & W & \ddots & & \\ & & \ddots & & V^\varepsilon & \\ & & & \ddots & & \\ 0 & & & U & W & V^\varepsilon \\ & & & & U & W \quad V^\varepsilon \end{array}\right) x \leq \left(\begin{array}{c} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_{r-1} \\ b_r \end{array}\right). \tag{5.1}$$

If the diagonal V^ε-blocks $V^\varepsilon \in \mathbb{R}^{n\times 2}$ satisfy the conditions above, then the blocks $U, W \in \mathbb{R}^{n\times 2}$ below the diagonal blocks can be arbitrary — the right-hand sides can be adjusted in such a way that we get a *deformed product*, which is combinatorially equivalent to a product $(P_n)^r$.

For this, we only have to verify that we get the correct combinatorics: If for each of the r blocks we choose two cyclically adjacent inequalities and require them to be tight, then this should yield a linear system of equations with a unique solution, for which all other inequalities are satisfied, but not tight. It is now easy to prove (by induction on r) that our system satisfies this — if we just choose the right-hand sides suitably; in particular, $b_i := M^{i-1}b \in \mathbb{R}^n$ works, for large enough M, if $V^\varepsilon x \leq b$ (as above) defines an n-gon.

Example 5.3. To illustrate this in a simple case, let's consider deformed products in the low-dimensional case of $I \times I$, a square, where I denotes an interval (a 1-dimensional polytope) such as $I = [0, 1]$.

In this case I can be written as

$$I \;=\; \{x \in \mathbb{R} : \begin{pmatrix} -1 \\ 1 \end{pmatrix} x \;\leq\; \begin{pmatrix} 0 \\ 1 \end{pmatrix}\}.$$

Consequently, the product $I \times I$ may be represented by

$$I \times I \;=\; \left\{x \in \mathbb{R}^2 : \begin{pmatrix} -1 & 0 \\ 1 & 0 \\ 0 & -1 \\ 0 & 1 \end{pmatrix} x \;\leq\; \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}\right\}.$$

Now changing the matrix into

$$\begin{pmatrix} -1 & 0 \\ 1 & 0 \\ a & -1 \\ b & 1 \end{pmatrix} x \;\leq\; \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}$$

leaves the first two inequalities (and thus the first I factor) intact, but it changes the slopes of the other two inequalities — and if you are unlucky (that is, for $a + b > 1$) the resulting polytope will not be equivalent to $I \times I$ any more. This situation is depicted in the middle part of Figure 5.2. However, it can be remedied by increasing right-hand sides: For any given a and b, a suitably large M, namely $M > a + b$, in

$$\begin{pmatrix} -1 & 0 \\ 1 & 0 \\ a & -1 \\ b & 1 \end{pmatrix} x \;\leq\; \begin{pmatrix} 0 \\ 1 \\ 0 \\ M \end{pmatrix}$$

will result in a product again (as in the right part of Figure 5.2).

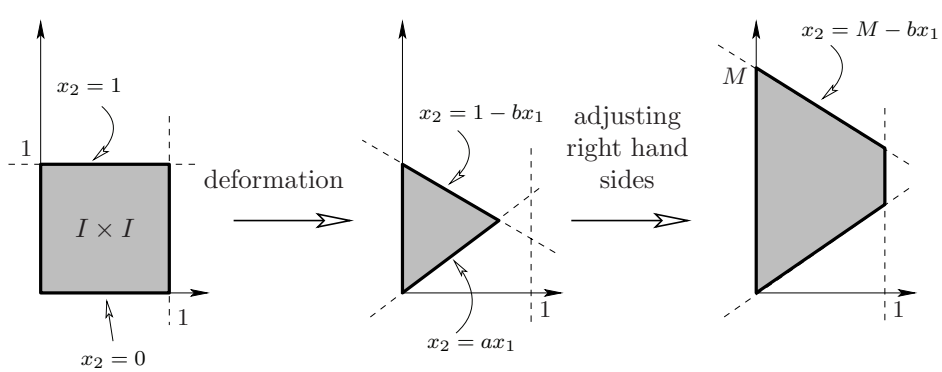


Figure 5.2. Construction of a deformation of $I \times I$

5.4. Surviving a Generic Projection

We are looking at *generic projections* of *simple polytopes* — that is, the direction of projection is in general position with respect to all the edges of the polytope to be projected, and thus a small perturbation of the inequalities of the polytope, and of the direction of projection, will not change the combinatorics of the projected polytope.

It should thus seem plausible that in this setting the following holds (compare Figure 5.3) for a polytope projection $\pi : P \to \pi(P)$:

- The normal vectors to a face $G \subset P$ are the positive linear combinations of the (defining) normal vectors n_F to all the facets $F \supset G$ that contain G.
- The proper faces of $\pi(P)$ are isomorphic copies of the faces of P that *survive* the projection.
- A face $G \subset P$ survives the projection exactly if it has a normal vector that is orthogonal to the direction of projection.

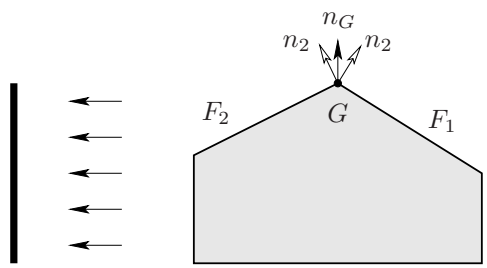


Figure 5.3. A vertex G surviving a projection, a normal vector n_G orthogonal to the projection, and the normal vectors n_1 and n_2 of facets F_1 and F_2 it can be combined from

For Theorem 5.1 we have to specify a deformed product realization for $(P_n)^r$ of the type given in Ansatz (5.1), such that all n-gon 2-faces are strictly preserved by the projection. That is,

- if we choose two cyclically adjacent rows from each block except for one, and
- truncate these rows to the first $2r - 4$ coordinates,

then the resulting $2r - 2$ vectors must be positively dependent and span.

5.5. Construction

Now we want to specify the lower-triangular block matrix in our Ansatz (5.1) so that it satisfies the following two main properties:

(1) the diagonal blocks have "rows in cyclic order," and
(2) any "choice of two cyclically adjacent rows" from all but one of the blocks, truncated to the first $2r - 4$ components, yields a positively-spanning set of vectors.

Five observations (you may call them "tricks") help us to achieve this:

(i) Condition (2) is stable under perturbation. So, we first construct a matrix that satisfies (2), then perturb it in order to achieve (1).
(The diagonal blocks of the matrix that we construct to satisfy (2) are denoted V; after perturbation, they will be V^ε.)

(ii) The submatrices V, W, and U of size $n \times 2$ are constructed to have alternating rows: So if you choose two cyclically adjacent rows from a block, you know what you get!

Specifically, we will let matrix V have rows that alternate between $(1,0)$ and $(0,0)$, matrix W gets rows $(0,1)$ and (a,b), and matrix U gets rows (c,d) and (e,f), with the six parameters a,b,c,d,e,f to be determined.

(iii) To make sure that you get the positive linear dependence for (2), we specify a positive coefficient sequence and compute the matrix entries to satisfy them.

(iv) Rather than admitting that from one of the blocks no row is chosen, we will prescribe coefficient sequences that could have zeroes on any one of the blocks, which yields linear dependencies for which the vectors from one block are "not used."

Specifically, we take coefficient sequences of the form $\alpha_k := (2^{k-t}-1)^2$ and $\beta_k := (2^{k-t}-1)(2^{k-t}-\frac{3}{2})$ for the odd resp. even-index rows. These coefficients are clearly positive for integral k, except they vanish at $k=t$.

Moreover, they are linear combinations of the three exponential functions 2^{k-t}, 4^{k-t}, and 1. If we write out the condition that "the rows chosen should be dependent, with coefficients α_k (for the even-index row chosen from the k-th block) and β_k (for the odd-index row chosen from the k-th block), then this leads to a system of six linear equations, in six unknowns a,b,c,d,e,f — solve it!

(v) The properties "alternating rows" and "rows in cyclic order" are, of course, incompatible — but a matrix with rows in cyclic order can be obtained as a perturbation V^ε of the matrix V that has rows $(1,0)$ and $(0,0)$ in alternation. Figure 5.4 suggests a way to do this.

This completes our sketch of "how to do it" — it should be sufficient to let you construct the polytopes P_n^r and thus prove Theorem 5.1 (but [**81**] provides these details, too.)

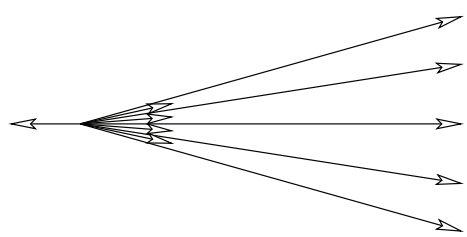


Figure 5.4. A configuration of 10 vectors, which positively span the plane, and which (in cyclic order) alternate between vectors close to $(1,0)$, and close to $(0,0)$

If you think about this construction, can you perhaps manage to simplify some of the details, or even to improve the construction? After all, on the way to the characterization of the f-vector cone for 4-polytopes, this construction of polytopes of fatness up to 9 should be taken just as an intermediate step. Can you get further? There is a long way to go from 9 to infinity ...

Exercises

5.1. Show that every polytope arises as a projection of a simplex.

5.2. Show that if $\pi : \mathbb{R}^n \to \mathbb{R}^d$, $P \to \pi(P)$ is a polytope projection, for an n-polytope P, and $n > d$, then the strictly preserved faces have dimension at most $d-1$. In particular, no facet of P is strictly preserved by the projection.

5.3. Give examples of polytope projections where no faces are strictly preserved.

5.4. Realize the prism over an n-gon ($n \geq 3$) in such a way that the projection $\mathbb{R}^3 \to \mathbb{R}^2$ strictly preserves all $2n$ vertices.

5.5. Show that the product $\Delta_2 \times \Delta_2$ of two triangles cannot be realized in such a way that all 9 vertices are strictly preserved in a projection to $\mathbb{R}^2$.

5.6. How large do we have to choose r and n in order to obtain 4-polytopes of fatness $F(\pi(P_n^r)) > 8$?

5.7. Any projection of a (non-deformed) product of centrally symmetric polygons is a *zonotope*. For these, one knows that $f_1 < 3f_0$ (for such inequalities, for the dual polytope, see [**16**, pp. 198/199]). Deduce from this information that the fatness is smaller than 5.

APPENDIX
A Short Introduction to polymake

by Thilo Schröder and Nikolaus Witte

The software project polymake [**31**] has been developed since 1997 in the Discrete Geometry group at TU Berlin by Ewgenij Gawrilow and Michael Joswig, with contributions by several others. It was initially designed to work with convex polytopes. Due to its open design the polymake framework can also be used on other types of objects; the current release includes a second application, topaz, which treats simplicial complexes.

polymake is designed to run on any Linux or Unix system, including Mac OS X. It runs in a shell using command line input. This introduction is for polymake versions 2.0 and 2.1. polymake is free software and you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation.

This introduction aims at getting you to work on the computer rather than explaining the details about the machinery of polymake. Therefore, there will be only a short description of the software design, in Section A.2. For further reading we refer to Gawrilow & Joswig [**32**] [**33**]. On the polymake website

http://www.math.tu-berlin.de/polymake

you will find extensive *online documentation* as well as an introductory *tutorial*.

A.1. Getting Started

For polymake, every polytope is treated as an *object* (the file storing the data) with certain *properties* such as its f-vector, its Hasse diagram, etc. The $\mathcal{V}$- and $\mathcal{H}$-representation are also regarded as properties, and any one of them may be used to define the polytope in the first place. If you are only interested in the combinatorial structure you may also input the vertex-facet incidences. For more details concerning the polymake file format see Section A.1.3. Once a polytope is defined in terms of some property, you may ask polymake to compute further properties.

polymake also provides standard constructions for polytopes. You can either construct polytopes from scratch (e.g. the d-cube) or by applying constructions to an existing polytope (e.g. the pyramid). In both cases polymake will produce a new file defining the new polytope.

This section explains the command line syntax of `polymake` for constructing, analyzing and visualizing polytopes. The `polymake` file format is briefly described at the end of this section.

A.1.1. Constructions of Polytopes

If you want to use one of `polymake`'s constructions, you have to call a *client program* to create the polytope. The clients have appropriate names, e.g. the client producing the d-cube is called `cube`. All clients are documented at [**31**]. If in doubt about the exact syntax of a client program just type the client's name, press return and you will get a *usage* message.

Clients producing polytopes from scratch
The basic syntax to create a polytope from scratch is the command line

```
<client> <file> [ <options> ]
```

For example, to produce a 4-cube type

```
cube c4.poly 4
```

Some other clients which produce polytopes you might have heard of are
`simplex`, `cyclic`, `cross`, `rand_sphere`, `associahedron`, and `permutahedron`.

Clients producing polytopes from others
To construct a polytope from existing one(s), use

```
<client> <out_file> <in_file> [ <options> ]
```

For example, to produce the pyramid over our 4-cube `c4.poly` type

```
pyramid c4.pyr.poly c4.poly
```

Some other constructions to get interesting polytopes are obtained via the clients
`bipyramid`, `prism`, `minkowski_sum`, `vertex_figure`, `center`, `truncation`, and `polarize`.

A.1.2. Computing Properties and Visualizing

To compute a property of a polytope, just ask `polymake` using the following syntax

```
polymake <file> <PROPERTY_1> <PROPERTY_2> ...
```

The properties are written in capital letters. To compute the numbers of facets and vertices of the pyramid over the 4-cube constructed above, type

```
polymake c4.pyr.poly N_FACETS N_VERTICES
```

Some other useful properties are

`GRAPH`, `DUAL_GRAPH`, `HASSE_DIAGRAM`, `CENTERED`, `VERTICES_IN_FACETS`, `F_VECTOR`, `H_VECTOR`, `SIMPLE`, `SIMPLICIAL`, `CUBICAL`, `SIMPLICIALITY`, and `SIMPLICITY`.

The visualizations use the same syntax. For example, to take a look at the graph of a polytope, just use

```
polymake my.poly VISUAL_GRAPH
```

Here are some more visualizations

`VISUAL`, `SCHLEGEL`, `VISUAL_FACE_LATTICE`, and `VISUAL_DUAL_GRAPH`.

A.1.3. File Format

Let's have a brief look at the `polymake` file format. If you have a look at the file `c4.pyr.poly`, you will find a paragraph for each property. The paragraphs are headed by the property's name in capital letters, followed by the data. The `VERTICES` and `POINTS` are represented in *homogeneous* coordinates, where the first coordinate is used for homogenization. The inequality $a_0 + a_1x_1 + \ldots + a_dx_d \geq 0$ is encoded as the vector $(a_0\ a_1\ \ldots\ a_d)$.

To define a polytope with your own $\mathcal{V}$- or $\mathcal{H}$-description you should create a file that contains a `POINTS` or `INEQUALITIES` section.

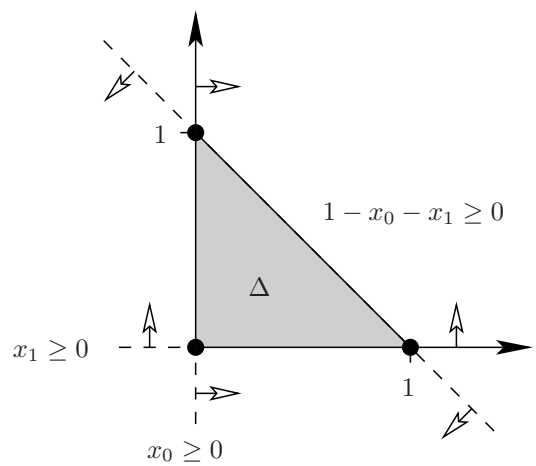


Figure A.1. The triangle Δ and its $\mathcal{V}$- and $\mathcal{H}$-description.

Example. If you want to construct the triangle Δ (cf. Figure A.1) that is given as

$$\Delta \quad := \quad \text{conv}\{(0,0),(0,1),(1,0)\}$$

your input file should look as follows (homogeneous coordinates):

```
POINTS
1 0 0
1 0 1
1 1 0
```

If you prefer to enter the same example by the inequality description

$$\Delta \quad := \quad \{x \in \mathbb{R}^2 : 1 - x_0 - x_1 \geq 0, x_0 \geq 0, x_1 \geq 0\},$$

this should be your `polymake` file:

```
INEQUALITIES
1 -1 -1
0 1 0
0 0 1
```

The difference between `POINTS` and `VERTICES` is that the former may contain redundancies. So you should enter `POINTS`, but ask for `VERTICES`. Similarly, one should enter `INEQUALITIES` and ask for `FACETS`.

A.2. The `polymake` System

The world seen through the eyes of the `polymake` system consists of *objects* with *properties*. For a given object you may ask for one of its properties and `polymake` will compute it. Yet `polymake` acts only as a framework for the computation, knowing little about the math involved. The actual computations are delegated to client programs. The open design, keeping the system and the math independent, allows for the use of external software as clients. It also makes `polymake` extremely flexible with respect to the kind of objects you want to examine. For example the applications `polytope` and `topaz` examine different classes of geometric objects.

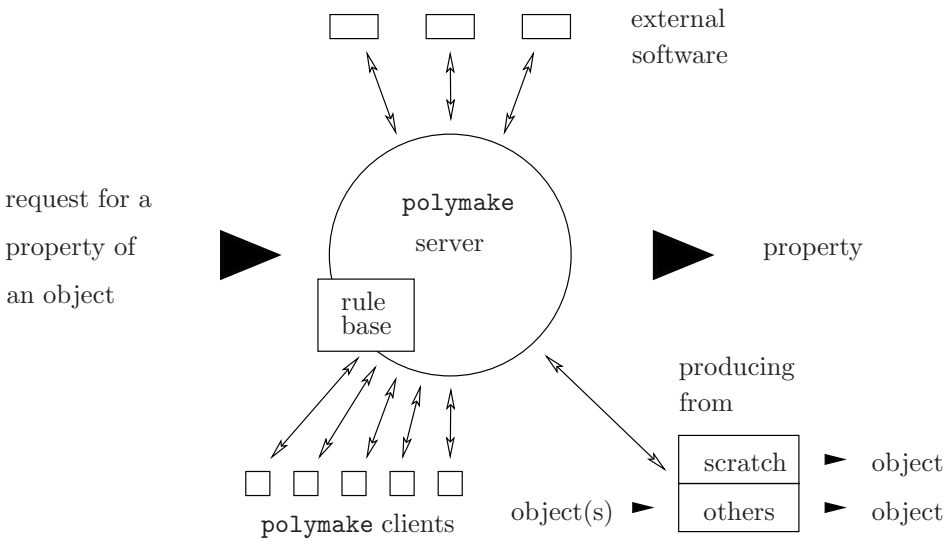


Figure A.2. Main components of the `polymake` system.

The `polymake` system consists of the following three components as illustrated in Figure A.2:

- The `polymake` *server* together with the *rule base*.
- `polymake` *client* programs computing properties and new objects.
- External software, such as the `cdd` [**28**] convex hull algorithm or the `JavaView` [**58**] visualization package.

The rule base is a collection of rules, each rule containing a set of input and output properties and an algorithm (that is a client or external software) which computes the output properties *directly* from the input properties. If you request a property of an object, the server has to determine how to compute the requested property from the ones which are already known. There might not be an algorithm computing the requested property directly, so other properties might have to be computed first. Therefore, the server has to compose a sequence of rules (from the rule base) to be executed in order to compute the requested property.

Examples and Exercises

A.1. Construct the d-simplex and the d-cube for $d = 3, 4, 5$. Visualize the face lattice of the polytope and its Schlegel diagram (see Section 3.1).

A.2. Construct the cyclic polytopes $C_3(7)$ and $C_4(8)$. Visualize and check Gale's evenness criterion (see Example 2.6). Alternatively, try constructing $C_4(8)$ using the client `cyclic_caratheodory`. What is the difference?

A.3. Produce the dual polytopes of the polytopes above. Watch out, they have to be `CENTERED`. (Check the documentation [**31**] for the property `CENTERED`)

A.4. Construct an octahedron using as many different ways as possible.

A.5. Build the bipyramid over a square, truncate both apexes and polarize. What does the resulting polytope look like?

A.6. Construct the product of a 5-gon and a 7-gon and visualize it.

A.7. Construct a 4-polytope with the `pcmi` logo as its Schlegel diagram.

A.8. Use the client `rand_sphere` to create a random polytope by uniformly distributing 1000 points on the unit 2-sphere, visualize it and compare it to Figure 1.4. Take a look at its `VERTEX_DEGREES` and cut off a vertex of maximal degree.

A.9. Take a 3-polytope and truncate all its vertices. Is the resulting polytope always simple?

A.10. Truncate the vertices of the 4-dimensional cross polytope, and let `polymake` compute the f-vector, and whether the resulting polytope is simple.
Can you justify the results theoretically?

A.11. For a given planar 3-connected graph G containing a triangular face, produce a 3-polytope with G as its graph (see Section 1.2). Use the client `tutte_lifting`.

A.12. Visualize the Schlegel diagrams of the `dwarfed_cube` that you get by using different projection facets.

A.13. Visualize the effect of standard constructions (such as truncation, stellar subdivision) on the Schlegel diagrams of 4-polytopes.

BIBLIOGRAPHY

1. R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network flows*, Prentice Hall Inc., Englewood Cliffs, NJ, 1993
2. M. Aigner and G. M. Ziegler, *Proofs from THE BOOK*, Springer-Verlag, Heidelberg Berlin, third ed., 2004
3. N. Amenta and G. M. Ziegler, *Shadows and slices of polytopes*, in Proceedings of the 12th Annual ACM Symposium on Computational Geometry, May 1996, pp. 10–19
4. ——, *Deformed products and maximal shadows*, in Advances in Discrete and Computational Geometry (South Hadley, MA, 1996), B. Chazelle, J. E. Goodman, and R. Pollack, eds., vol. 223 of Contemporary Mathematics, Providence RI, 1998, Amer. Math. Soc., pp. 57–90
5. E. M. Andreev, *On convex polyhedra in Lobačevskiĭ spaces*, Math. of the USSR — Sbornik, 10 (1970), pp. 445–478. Translation from Math. Sbornik (N.S.) 81 (123) (1970), pp. 445–478
6. M. L. Balinski, *On the graph structure of convex polyhedra in n-space*, Pacific J. Math., 11 (1961), pp. 431–434
7. D. W. Barnette, *Projections of 3-polytopes*, Israel J. Math., 8 (1970), pp. 304–308
8. ——, *The projection of the f-vectors of 4-polytopes onto the (E, S)-plane*, Discrete Math., 10 (1974), pp. 201–216
9. D. W. Barnette and B. Grünbaum, *Preassigning the shape of a face*, Pacific J. Math., 32 (1970), pp. 299–302
10. D. W. Barnette and J. R. Reay, *Projections of f-vectors of four-polytopes*, J. Combinatorial Theory, Ser. A, 15 (1973), pp. 200–209
11. M. M. Bayer, *The extended f-vectors of 4-polytopes*, J. Combinatorial Theory, Ser. A, 44 (1987), pp. 141–151
12. L. J. Billera and C. W. Lee, *A proof of the sufficiency of McMullen's conditions for f-vectors of simplicial polytopes*, J. Combinatorial Theory, Ser. A, 31 (1981), pp. 237–255
13. A. Björner, *The unimodality conjecture for convex polytopes*, Bulletin Amer. Math. Soc., 4 (1981), pp. 187–188
14. ——, *Face numbers of complexes and polytopes*, in Proceedings of the International Congress of Mathematicians (Berkeley CA, 1986), 1986, pp. 1408–1418

15. ———, *Partial unimodality for f-vectors of simplicial polytopes and spheres*, in Jerusalem Combinatorics '93, H. Barcelo and G. Kalai, eds., vol. 178 of Contemporary Math., Providence RI, 1994, Amer. Math. Soc., pp. 45–54
16. A. Björner, M. Las Vergnas, B. Sturmfels, N. White, and G. M. Ziegler, *Oriented Matroids*, vol. 46 of Encyclopedia of Mathematics, Cambridge University Press, Cambridge, second (paperback) ed., 1999
17. A. I. Bobenko and B. A. Springborn, *Variational principles for circle patterns, and Koebe's theorem*, Transactions Amer. Math. Soc., 356 (2004), pp. 659–689
18. T. Braden, *A glued hypersimplex*. Personal communication, 1997
19. F. Brenti, *Log-concave and unimodal sequences in algebra, combinatorics, and geometry: an update*, in Jerusalem Combinatorics '93, vol. 178 of Contemp. Math., Amer. Math. Soc., Providence, RI, 1994, pp. 71–89
20. Y. Colin De Verdière, *Un principe variationnel pour les empilements de cercles*, Inventiones Math., 104 (1991), pp. 655–669
21. G. E. Cooke and R. L. Finney, *Homology of Cell Complexes*, Princeton University Press, Princeton NJ, 1967
22. H. S. M. Coxeter, *Regular Polytopes*, Macmillan, New York, second ed., 1963. Corrected reprint, Dover, New York 1973
23. J. De Loera, B. Sturmfels, and R. R. Thomas, *Gröbner bases and triangulations of the second hypersimplex*, Combinatorica, 15 (1995), pp. 409–424
24. J. Eckhoff, *Combinatorial properties of f-vectors of convex polytopes*. Unpublished manuscript, Dortmund 1985
25. D. Eppstein, *Nineteen proofs of Euler's formula: $V-E+F=2$*. The Geometry Junkyard, `http://www.ics.uci.edu/~eppstein/junkyard/euler/`, October 2005
26. D. Eppstein, G. Kuperberg, and G. M. Ziegler, *Fat 4-polytopes and fatter 3-spheres*, in Discrete Geometry: In honor of W. Kuperberg's 60th birthday, A. Bezdek, ed., vol. 253 of Pure and Applied Mathematics, Marcel Dekker Inc., New York, 2003, pp. 239–265, `http://www.arXiv.org/math.CO/0204007`
27. P. J. Federico, *Descartes on Polyhedra*. A study of the *De solidorum elementis*, vol. 4 of Sources in the History of Mathematics and Physical Sciences, Springer-Verlag, New York, 1982
28. K. Fukuda, *CDD and CDD+ — implementations of the double description method*, `http://www.ifor.math.ethz.ch/~fukuda/cdd_home/`
29. A. M. Gabriélov, I. M. Gel'fand, and M. V. Losik, *Combinatorial computation of characteristic classes*, Functional Analysis Appl., 9 (1975), pp. 103–115
30. D. Gale, *Neighborly and cyclic polytopes*, in Convexity, V. Klee, ed., vol. VII of Proc. Symposia in Pure Mathematics, Providence RI, 1963, Amer. Math. Soc., pp. 225–232
31. E. Gawrilow and M. Joswig, `Polymake`: *A software package for analyzing convex polytopes*, `http://www.math.tu-berlin.de/diskregeom/polymake/`
32. ———, `Polymake`: *A framework for analyzing convex polytopes*, in Polytopes — Combinatorics and Computation, G. Kalai and G. M. Ziegler, eds., vol. 29 of DMV Seminar, Birkhäuser-Verlag, Basel, 2000, pp. 43–73

33. ———, *Geometric reasoning with* `POLYMAKE`, Preprint, July 2005, 13 pages, `http://www.arxiv.org/math/0507273`
34. I. M. Gelfand, M. M. Kapranov, and A. V. Zelevinsky, *Discriminants, Resultants, and Multidimensional Determinants*, Birkhäuser, Boston, 1994
35. I. M. Gel'fand and R. D. MacPherson, *Geometry in Grassmannians and a generalization of the dilogarithm*, Advances in Math., 44 (1982), pp. 279–312
36. G. Gévay, *Kepler hypersolids*, in Intuitive geometry (Szeged, 1991), vol. 63 of Colloq. Math. Soc. János Bolyai, Amsterdam, 1994, North-Holland, pp. 119–129
37. D. Goldfarb, *On the complexity of the simplex algorithm*, in Advances in optimization and numerical analysis. Proc. 6th Workshop on Optimization and Numerical Analysis, Oaxaca, Mexico, January 1992, Dordrecht, 1994, Kluwer, pp. 25–38. Based on: *Worst case complexity of the shadow vertex simplex algorithm*, preprint, Columbia University 1983, 11 pages
38. M. Grötschel, *Cardinality homogeneous set systems, cycles in matroids, and associated polytopes*, in The Sharpest Cut: The Impact of Manfred Padberg and His Work, M. Grötschel, ed., MPS-SIAM, 2004, pp. 99–120
39. B. Grünbaum, *Convex Polytopes*, vol. 221 of Graduate Texts in Math., Springer-Verlag, New York, 2003. Second edition prepared by V. Kaibel, V. Klee and G. M. Ziegler (original edition: Interscience, London 1967)
40. M. Henk, J. Richter-Gebert, and G. M. Ziegler, *Basic properties of convex polytopes*, in Handbook of Discrete and Computational Geometry, J. E. Goodman and J. O'Rourke, eds., Chapman & Hall/CRC Press, Boca Raton, second ed., 2004, ch. 16, pp. 355–382
41. D. Hilbert and S. Cohn-Vossen, *Anschauliche Geometrie*, Springer-Verlag, Berlin Heidelberg, 1932. Second edition 1996. English translation: Geometry and the Imagination, Chelsea Publ., 1952
42. A. Höppner and G. M. Ziegler, *A census of flag-vectors of* 4*-polytopes*, in Polytopes — Combinatorics and Computation, G. Kalai and G. M. Ziegler, eds., vol. 29 of DMV Seminars, Birkhäuser-Verlag, Basel, 2000, pp. 105–110
43. M. Joswig and G. M. Ziegler, *Neighborly cubical polytopes*, Discrete & Computational Geometry (Grünbaum Festschrift: G. Kalai, V. Klee, eds.), (2-3)24 (2000), pp. 325–344, `http://www.arXiv.org/math.CO/9812033`
44. G. Kalai, *Rigidity and the lower bound theorem, I*, Inventiones Math., 88 (1987), pp. 125–151
45. V. Klee and G. J. Minty, *How good is the simplex algorithm?*, in Inequalitites, III, O. Shisha, ed., Academic Press, New York, 1972, pp. 159–175
46. P. Koebe, *Kontaktprobleme der konformen Abbildung*, Berichte Verh. Sächs. Akademie der Wissenschaften Leipzig, Math.-Phys. Klasse, 88 (1936), pp. 141–164
47. C. W. Lee, *Counting the faces of simplicial polytopes*, Ph.D. thesis, Cornell University, 1981, 171 pages
48. J. Matoušek, *Lectures on Discrete Geometry*, vol. 212 of Graduate Texts in Math., Springer-Verlag, New York, 2002
49. J. C. Maxwell, *On reciprocal figures and diagrams of forces*, Philosophical Magazine, Ser. 4, 27 (1864), pp. 250–261
50. B. Mohar, *A polynomial time circle packing algorithm*, Discrete Math., 117 (1993), pp. 257–263

51. K. G. Murty, *Computational complexity of parametric linear programming*, Math. Programming, 19 (1980), pp. 213–219
52. S. Onn and B. Sturmfels, *A quantitative Steinitz' theorem*, Beiträge zur Algebra und Geometrie/Contributions to Algebra and Geometry, 35 (1994), pp. 125–129
53. J. Pach and P. K. Agarwal, *Combinatorial Geometry*, J. Wiley and Sons, New York, 1995
54. A. Paffenholz, *The E-construction applied to products*, Webpage with `polymake` data files, TU Berlin 2004, `http://www.math.tu-berlin.de/~paffenho/2s2spages.shtml`
55. ———, *New polytopes from products*, Preprint, TU Berlin, 22 pages, November 2004, `arXiv:math.MG/0411092`
56. A. Paffenholz and A. Werner, *Constructions for 4-polytopes and the cone of flag vectors*, Preprint, TU Berlin, November 2005, 20 pages, `http://arXiv.org/math/0511751`
57. A. Paffenholz and G. M. Ziegler, *The E_t-construction for lattices, spheres and polytopes*, Discrete & Computational Geometry (Billera Festschrift: M. Bayer, C. Lee, B. Sturmfels, eds.), 32 (2004), pp. 601–624, `http://www.arXiv.org/math.MG/0304492`
58. K. Polthier, S. Khadem-Al-Charieh, E. Preuss, and U. Reitebuch, `JavaView` *visualization software, 1999-2004*, `http://www.javaview.de`
59. A. Ribó Mor, *Realization and Counting Problems for Planar Structures: Trees and Linkages, Polytopes and Polyominoes*, PhD thesis, FU Berlin, 2005, 23+167 pages
60. J. Richter-Gebert, *Realization Spaces of Polytopes*, vol. 1643 of Lecture Notes in Mathematics, Springer-Verlag, Berlin Heidelberg, 1996
61. G. Rote, *The number of spanning trees in a planar graph*, Oberwolfach Reports, 2 (2005), pp. 969–973
62. R. Sanyal, *On the combinatorics of projected deformed products*, Diplomarbeit, TU Berlin, 2005, 57 pages
63. R. Sanyal, T. Schröder, and G. M. Ziegler, *Polytopes and polyhedral surfaces via projection*, Preprint in preparation, TU Berlin 2005
64. G. Schaeffer, *Random sampling of large planar maps and convex polyhedra*, in Annual ACM Symposium on Theory of Computing (Atlanta, GA, 1999), ACM, New York, 1999, pp. 760–769
65. D. Schattschneider and M. Senechal, *Tilings*, in Handbook of Discrete and Computational Geometry, J. E. Goodman and J. O'Rourke, eds., Chapman & Hall/CRC Press, Boca Raton, second ed., 2004, ch. 3, pp. 53–72
66. L. Schläfli, *Theorie der vielfachen Kontinuität*, Denkschriften der Schweizerischen naturforschenden Gesellschaft, Vol. 38, pp. 1–237, Zürcher und Furrer, Zürich, 1901. Written 1850-1852. Reprinted in: Ludwig Schläfli, 1814-1895, Gesammelte Mathematische Abhandlungen Vol. I, Birkhäuser, Basel 1950, pp. 167–387
67. V. Schlegel, *Theorie der homogen zusammengesetzten Raumgebilde*, Nova Acta Leop. Carol. (Verhandlungen der Kaiserlichen Leopoldinisch-Carolinischen Deutschen Akademie der Naturforscher, Halle), 44 (1883), pp. 343–459, W. Engelmann, Leipzig

68. B. A. Springborn, *A unique representation theorem of polyhedral types: Centering via Möbius transformations*, Math. Zeitschrift, 249 (2005), pp. 513–517, http://www.arXiv.org/math.MG/0401005
69. R. P. Stanley, *The number of faces of simplicial polytopes and spheres*, in Discrete Geometry and Convexity (New York 1982), J. E. Goodman, E. Lutwak, J. Malkevitch, and R. Pollack, eds., vol. 440 of Annals of the New York Academy of Sciences, 1985, pp. 212–223
70. ——, *Generalized h-vectors, intersection cohomology of toric varieties, and related results*, in Commutative Algebra and Combinatorics, M. Nagata and H. Matsumura, eds., vol. 11 of Advanced Studies in Pure Mathematics, Kinokuniya, Tokyo, 1987, pp. 187–213
71. R. P. Stanley, *Log-concave and unimodal sequences in algebra, combinatorics, and geometry*, in Graph theory and its applications: East and West (Jinan, 1986), vol. 576 of Ann. New York Acad. Sci., New York Acad. Sci., New York, 1989, pp. 500–535
72. G. Stein, *Realisierung von* 3*-Polytopen*, Diplomarbeit, TU Berlin, 2000, in German, 100 pages
73. E. Steinitz, *Über die Eulerschen Polyederrelationen*, Archiv für Mathematik und Physik, 11 (1906), pp. 86–88
74. ——, *Polyeder und Raumeinteilungen*, in Encyklopädie der mathematischen Wissenschaften, Geometrie, III.1.2., Heft 9, Kapitel III A B 12, W. F. Meyer and H. Mohrmann, eds., B. G. Teubner, Leipzig, 1922, pp. 1–139
75. E. Steinitz and H. Rademacher, *Vorlesungen über die Theorie der Polyeder*, Springer-Verlag, Berlin, 1934. Reprint, Springer-Verlag 1976
76. W. P. Thurston, *Geometry and Topology of* 3*-Manifolds*, Lecture Notes, Princeton University, Princeton 1977–1978
77. W. T. Tutte, *How to draw a graph*, Proc. London Math. Soc. (3), 13 (1963), pp. 743–767
78. A. Werner, *Unimodality and convexity of f-vectors of polytopes*, Preprint, TU Berlin, December 2005, http://www.arXiv.org/math.CO/0512131
79. G. M. Ziegler, *Lectures on Polytopes*, vol. 152 of Graduate Texts in Mathematics, Springer-Verlag, New York, 1995. Revised edition, 1998; "Updates, corrections, and more" at http://www.math.tu-berlin.de/~ziegler
80. ——, *Face numbers of* 4*-polytopes and* 3*-spheres*, in Proceedings of the International Congress of Mathematicians (ICM 2002, Beijing), L. Tatsien, ed., vol. III, Beijing, China, 2002, Higher Education Press, pp. 625–634, http://www.arXiv.org/math.MG/0208073
81. ——, *Projected products of polygons*, Electronic Research Announcements AMS, 10 (2004), pp. 122–134, http://www.arXiv.org/math.MG/0407042