IRMA Lectures in Mathematics and Theoretical Physics 6

Edited by Vladimir G. Turaev

Institut de Recherche Mathématique Avancée
Université Louis Pasteur et CNRS
7 rue René Descartes
67084 Strasbourg Cedex
France

IRMA Lectures in Mathematics and Theoretical Physics
Edited by Vladimir G. Turaev

This series is devoted to the publication of research monographs, lecture notes, and other materials arising from programs of the Institut de Recherche Mathématique Avancée (Strasbourg, France). The goal is to promote recent advances in mathematics and theoretical physics and to make them accessible to wide circles of mathematicians, physicists, and students of these disciplines.

Previously published in this series:

1 Deformation Quantization, *Gilles Halbout* (Ed.)
2 Locally Compact Quantum Groups and Groupoids, *Leonid Vainerman* (Ed.)
3 From Combinatorics to Dynamical Systems, *Frédéric Fauvet and Claude Mitschi* (Eds.)
4 Three courses on Partial Differential Equations, *Eric Sonnendrücker* (Ed.)
5 Infinite Dimensional Groups and Manifolds, *Tilman Wurzbacher* (Ed.)

Volumes 1–5 are available from Walter de Gruyter (www.degruyter.de)

Athanase Papadopoulos

Metric Spaces, Convexity and Nonpositive Curvature

European Mathematical Society

Author:

Athanase Papadopoulos
CNRS and Université Louis Pasteur
Institut de Recherche Mathématiques Avancée
7 rue René Descartes
F-67084 Strasbourg Cedex
France

2000 Mathematics Subject Classification 26-01; 30F25; 30F45; 30F60; 32G15; 32Q45; 51-01; 51K05; 51K10; 51M09; 51M10; 51F99; 52-01; 52A07; 52A41; 53-01; 53C70; 54-01; 54E35.

Bibliographic information published by Die Deutsche Bibliothek

Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie;
detailed bibliographic data are available in the Internet at http://dnb.ddb.de.

ISBN 3-03719-010-8

Contact address:

European Mathematical Society Publishing House
Seminar for Applied Mathematics
ETH-Zentrum FLI C4
CH-8092 Zürich
Switzerland

Phone: +41 (0)1 632 34 36
Email: info@ems-ph.org
Homepage: www.ems-ph.org

Typeset using the author's TEX files: I. Zimmermann, Freiburg
Printed in Germany

9 8 7 6 5 4 3 2 1

Dedicated to the memory of my parents

Preface

This book is about metric spaces of nonpositive curvature in the sense of Busemann, that is, metric spaces whose distance function is convex. We have also included a systematic introduction to the theory of geodesics and related matters in metric spaces, as well as a detailed presentation of a few facets of convexity theory that are useful in the study of nonpositive curvature.

We have tried to start from first principles and to give full proofs, but there are two exceptions, which occur several times in examples that are spread through the book. These exceptions are the following:

- we give without proof some classical well-known facts about hyperbolic spaces $\mathbb{H}^n$;
- we refer to some elements of the theory of Teichmüller space.

The material that we present in connection with these spaces is used as an illustration for the general concepts that are developed. It can be asserted without exaggeration that Teichmüller space is the most beautiful space ever, and the metric theory of that space, in particular its convexity properties, is most interesting.

This book can be divided into three parts. Part I (Chapters 1 to 4) contains basic material on metric spaces, Part II (Chapters 5 to 7) concerns convexity in vector spaces and Part III (Chapters 8 to 12) concerns convexity in metric spaces and related matters.

Each chapter contains an introduction in which we describe its content. At the end of each chapter we have included a few notes, some of them historical and others indicating some further developments. In both cases, there is no pretention for completeness. The only reason for which a note is included is that the author of this book thinks that it is interesting information.

Acknowledgements. This text grew up from notes I wrote for graduate courses I gave at the Universities of Strasbourg and of Florence. The reader will easily realize that its content owes a lot to the work of Herbert Busemann. The course in Italy was supported by the Istituto Nazionale di Alta Matematica (Roma). I would like to thank Raffaella Lefkou and Vincenzo Ancona who invited me to give that course.

Contents

Introduction: Some historical markers

The aim of this introduction is threefold. First, we give a brief account of a paper by Jacques Hadamard which is at the genesis of many works on global properties of surfaces of nonpositive Gaussian curvature. Then, we make a brief account of several different notions of nonpositive curvature in metric spaces, each of which generalizes in its own manner the notion of nonpositive Gaussian curvature for surfaces or of nonpositive sectional curvature for Riemannian manifolds. In particular, we mention important works by Menger, Busemann and Alexandrov in this domain. Finally, we review some of the connections between convexity theory and the theory of nonpositive curvature, which constitute the main theme that we develop in this book.

The work of Hadamard

The theory of spaces of nonpositive curvature has a long and interesting history, and it is good to look at its sources. We start with a brief review of the pioneering paper by Hadamard, "Les surfaces à courbures opposées et leurs lignes géodésiques" [62]. This paper, which was written at the end of the nineteenth century, can be considered as the foundational paper for the study of global properties of nonpositively curved spaces.

Hadamard was one of the first mathematicians (and may have been the first) who strongly emphasized the importance of topological methods in the study of spaces of nonpositive curvature. In the introduction to the paper cited above, he writes : "The only theory which has to be studied profoundly, as a basis to the current work, is the *Analysis situs*, which, as one can expect after reading the work of Poincaré, plays an essential role in everything that will follow."

Let us look more closely at the content of that paper.

Hadamard considers a surface S equipped with a Riemannian metric of nonpositive curvature. The surface is smoothly embedded in $\mathbb{R}^3$, the Riemannian metric is induced from that inclusion and the set of points at which the curvature is zero is finite.

Hadamard starts by noting that this surface is necessarily unbounded since if one of the coordinates assumes a maximum or a minimum, then, in a neighborhood of that point, the surface would be situated at one side of its tangent plane and therefore the curvature would be positive. He then proves that such a surface can always be decomposed into the union of a compact region and of a collection of *infinite sheets* (we are translating the term "nappes infinies" used by Hadamard). Each infinite sheet is homeomorphic to a cylinder, and is connected to the compact region along a boundary curve. Hadamard constructs examples of surfaces with nonpositive curvature having an arbitrary number of infinite sheets. He classifies the infinite sheets into two types: the *flared* infinite sheets ("nappes infinies évasée"), which we call "funnels", and the *unflared* infinite sheets. An unflared infinite sheet is characterized by the fact that one can continuously push to infinity a homotopically non-trivial closed curve on such a sheet while keeping its length bounded from above. To each unflared infinite sheet, Hadamard associates an *asymptotic direction*. The simple closed curves that connect the funnels to the compact region can be taken to be closed geodesics that are pairwise disjoint.

Let us now suppose that the fundamental group of S is finitely generated (Hadamard says that S has "finite connectivity"). For such a surface, Hadamard describes a finite collection of homotopically non-trivial and pairwise non-homotopic simple closed curves $C_1, \ldots, C_n$ such that any non-trivial homotopy class of closed curves on the surface can be uniquely represented by a finite word written in the letters $C_1, \ldots, C_n, C_1^{-1}, \ldots, C_n^{-1}$. Hadamard pushes forward this analysis to make it include not only representations of closed curves, but also representations of long segments of open infinite curves. In fact, with these elements, Hadamard initiates the approximation of long segments of bi-infinite geodesic lines by closed geodesics, and therefore he also initiates the theory of symbolic representation of bi-infinite geodesics on the surface, that is, the theory of representation of elements of the geodesic flow associated to S by bi-infinite words in the finite alphabet $\{C_1, \ldots, C_n, C_1^{-1}, \ldots, C_n^{-1}\}$. We note in passing that the theory of symbolic dynamics for the geodesic flow associated to a nonpositively curved surface was thoroughly developed about twenty years later by Marston Morse, and that Morse, in his paper [113], refers to Hadamard as his source of inspiration. Hadamard calls a closed curve a "contour", and the curves $C_1, \ldots, C_n$ the "elementary contours". There are two sorts of closed curves in the collection $C_1, \ldots, C_n$ which Hadamard considers: the curves that are the boundaries of the infinite sheets, and those that correspond to the genus of the surface. (These curves come in pairs and they correspond to the "holes" of the compact part of the surface S, as Hadamard calls them.) He notices that there is (up to circular permutation) one and only one relation in the representation he obtains that allows us to consider any of these contours as a product of the elementary contours, and therefore, by eliminating one of the contours, he obtains a symbolic representation that is unique up to circular permutation.

In the same paper, Hadamard proves that in each homotopy class of paths on S with fixed endpoints, there is a unique geodesic.[1] He also obtains an analogous result for the free homotopy classes of closed curves on S that are neither homotopic to a point nor to the core curve of a flared infinite sheet.

Hadamard then makes the fundamental observation that the distance function from a fixed point in S to a point that moves along a global geodesic in this surface is convex. This observation is at the basis of the definition of nonpositive curvature in the sense

[1] Here, the word geodesic is used, as in Riemannian geometry, in the sense of local geodesic. We warn the reader that this is not the definition that we are adopting in the rest of these notes, where a geodesic is a distance-minimizing map between its endpoints (and therefore it is a *global* geodesic).

of Busemann, which is the main topic of this book.

Next, Hadamard introduces the notion of *asymptotic geodesic rays*, and he constructs such rays as follows. Given a geodesic ray, he considers a sequence of geodesic segments joining a fixed point on the surface to a sequence of points on that ray that tends to infinity. He proves that this sequence of geodesic segments converges to a geodesic ray, which he calls *asymptotic* to the initial ray. He studies the asymptoticity relation and introduces through an analogous construction the notion of (local) geodesics that are asymptotic to a given closed geodesic.

Using an argument that is based on the convexity of the distance function from a point in a funnel to the closed geodesic that connects this funnel to the compact region in S, Hadamard proves that a geodesic that penetrates a funnel cannot get out of it.[2]

Finally, in the case where all the infinite sheets of S are funnels, Hadamard investigates the distribution of geodesics that stay in the compact part of the surface. These geodesics are the flowlines of the geodesic flow associated with the compact nonpositively curved surface with geodesic boundary, obtained from the surface S by deleting the funnels. For a given point x in S, Hadamard considers the set of initial directions of geodesic rays that start at x and stay in that compact surface. He proves that this is a perfect set with empty interior, whereas the set of directions of geodesic rays that tend to infinity is open.

We already mentioned that the setting of the paper [62] by Hadamard is that of a differentiable surface S embedded in $\mathbb{R}^3$, such that at each point of S, the Gaussian curvature is negative except for a finite set of points where it is zero. A theory of *metric spaces with nonpositive curvature*, that is, a theory that does not make any differentiability assumption and whose methods use the distance function alone, without the local coordinates provided by an embedding in Euclidean space or by another Riemannian metric structure, was developed several decades after the paper [62]. It is good to remember, in this respect, that the theory of metric spaces itself was developed long after the theory of spaces equipped with differentiable structures. For instance, Gauss's treatise on the differential geometry of surfaces [51], in which he defines (Gaussian) curvature and proves that this curvature depends only on the intrinsic geometry of the surface and not on its embedding in $\mathbb{R}^3$, was published in 1827,[3] whereas the axioms for metric spaces were set down by Fréchet, some 90 years later.

Of course, the setting of surfaces embedded in $\mathbb{R}^3$ has the advantage of providing visual characteristics for the sign of the curvature. For instance, it is well-known that for such a surface, we have the following:

- if the Gaussian curvature at some point is > 0, then the surface, in the neighborhood of that point, is situated on one side of the tangent plane;
- if the Gaussian curvature at some point is < 0, then the surface, at that point, crosses its tangent plane.

[2] Hadamard already obtained a similar result in [60].

[3] We recall that the surfaces considered by Gauss were always embedded in $\mathbb{R}^3$. It is only 30 years after Gauss's paper was written that Riemann introduced the concept of spaces equipped with (Riemannian) metrics with no embedding in $\mathbb{R}^3$ involved in the definition.

We also recall that if the Gaussian curvature at some point is 0, then any one of the above configurations can occur.

Hadamard indicated in the note [61] how to extend some of the results he proved for surfaces to higher dimensions. The development of these ideas in the general setting of Riemannian manifolds of nonpositive curvature has been carried out by Elie Cartan, in particular in his famous "Leçons sur la géométrie des espaces de Riemann" [36].[4]

The works of Menger and Wald

A few years after the introduction by Fréchet of the axioms for metric spaces, Karl Menger initiated a theory of geodesics in these spaces. A *geodesic* in a metric space is a path whose length is equal to the distance between its endpoints. Menger generalized several results of classical geometry to this new setting and he introduced new methods that did not make any use of local coordinates or of differentials, but only of equalities involving the distance function, and of the triangle inequality. Menger wrote several important papers involving this new notion of geodesic, and he also introduced a notion of "discrete geodesic", which is based on his definition of *betweenness*: a point z in a metric space is said to lie between two distinct points x and y if z is distinct from x and from y and if we have

$$d(x, y) = d(x, z) + d(z, y).$$

In a complete metric space, the existence of a geodesic joining any two points is equivalent to the existence, for any distinct points, of a point that lies between them. We refer the reader to the commented edition of Menger's papers in the volume [108], published on the one hundredth anniversary of Menger's birth. We shall have several opportunities to mention Menger's work in the following chapters, but here, we mention an important notion that he introduced, which we shall not consider further in this book. This notion contains an idea that is at the basis of the various definitions of curvature that make sense in general metric spaces. The idea is to construct "comparison configurations" for sets of points (say of finite cardinality) in a given metric space X. The comparison configurations are built in a model space, which is generally one of the complete simply connected surfaces M_κ of constant Gaussian curvature κ, that is, either the Euclidean plane (of curvature $\kappa = 0$), or a sphere of curvature $\kappa > 0$, or a hyperbolic plane of curvature $\kappa < 0$. The comparison configuration for a given subset $F \subset X$ is a subset $F^\star$ of M_κ, equipped with a map from F to $F^\star$, which is generally taken to be distance-preserving and which is called a "comparison map". Then, one can define notions like curvature at a given point x in X by requiring the *ad hoc* property for the comparison configurations associated to certain classes of subsets contained in a neighborhood of x. Of course, a comparison

[4] Note that the later editions of Cartan's book [36] contains additional material, in particular the development of the work of Hadamard to higher dimensions.

configuration does not always exist, but in the case where F is of cardinality 3, one can always construct a comparison configuration $F^\star$ of F in any one of the surfaces M_κ with $\kappa \leq 0$. In fact, the axioms for a metric space X are equivalent to the fact that one can construct a comparison configuration in the Euclidean plane for any triple of points in X.[5] Let us now consider an example.

Given three pairwise distinct points a, b and c in the Euclidean plane M_0, either they lie on a unique circle (the circumscribed circle), or they lie on a Euclidean straight line. It is useful to consider here such a line as being a circle of radius ∞, in order to avoid taking subcases. Now for any triple of pairwise distinct points in a metric space X, Menger defined its "curvature" as being equal to $1/R$, where R is the radius of a circle in the Euclidean plane which is circumscribed to a comparison configuration associated to that triple. With this definition, the three points in X are aligned (that is, they satisfy a degenerate triangle inequality) if and only if their curvature is zero. Given an arc (or, say, a one-dimensional object) A contained in X, Menger says that the curvature of A at a point $a \in A$ is equal to κ if for any triple of pairwise distinct points in A that are sufficiently close to a, the curvature of this triple is close to κ.

With this definition, Menger introduced the notion of *curvature* for one-dimensional objects in an arbitrary metric space, and he posed the problem of the definition of curvature for higher-dimensional objects.

Now, we must mention the work of Abraham Wald, a student of Menger, who introduced a notion of two-dimensional curvature in an arbitrary metric space.[6] The definition again uses a limiting process, but now it involves quadruples of points in that space. The problem is that in general, a quadruple of points in a metric space does not necessarily possess a comparison configuration in the Euclidean plane. However, Wald starts by proving that if the metric space X is a differentiable surface, then, for every point x in X, there exists a real number $\kappa(x)$ satisfying the following property:

($\star$) for each $\epsilon > 0$, there exists a neighborhood $V(x)$ of x such that for every quadruple of points Q in $V(x)$, there is an associated real number $\kappa(Q)$ satisfying $|\kappa(x) - \kappa(Q)| < \epsilon$ such that the quadruple of points Q possesses a comparison configuration in the model surface $M_{\kappa(Q)}$.

Wald then proves that the quantity $\kappa(x)$ is equal to the Gaussian curvature of the surface at the given point x.

Now let X be a metric space that is "Menger convex", that is, a metric space X in which for every pair of distinct points x and y, there exists a point z that lies between them. Suppose furthermore that the Wald two-dimensional curvature exists at each point of X. In other words, suppose that one can associate to each point x in X a real number $\kappa(x)$ satisfying property ($\star$) that is stated above. Under these assumptions, Wald shows that the space X has the structure of a differentiable surface embedded in $\mathbb{R}^3$ that induces the same *length structure* as that of the original metric on X. In

[5] We also mention that an early version of the idea of a comparison map is already contained in the very definition of the Gauss map.

[6] Wald gave this 2-dimensional curvature the name "surface curvature", and Menger refers to it as "Wald curvature".

other words, the lengths of an arbitrary curve in X, on the one hand measured using the original metric and on the other hand measured using this differentiable surface structure, coincide. Furthermore, at each point a in X, the quantity $\kappa(a)$ is equal to the Gaussian curvature induced by the differentiable embedding of the surface in $\mathbb{R}^3$. With this result, Wald solved a problem that had been posed by Menger in [105], which asked for a metric characterization of Gauss surfaces among Menger convex metric spaces. We refer the reader to the Comptes Rendus Note [141] by Wald, presented by Elie Cartan, which describes this work.

We note in passing that for extra-mathematical reasons, Wald stopped working on this subject soon after he published the solution to Menger's problem, and his research interests switched to statistics and econometry. The story is interesting and it is told by Menger in [109]. It seems that there was no direct continuation to Wald's work.

Now, after the notion of curvature in metric spaces, we pass to the notion of nonpositive curvature.

The works of Busemann and Alexandrov

For the development of the theory of nonpositively curved metric spaces, we shall consider works that have been carried out in two different directions: the works of H. Busemann and the works of A. D. Alexandrov and his collaborators. Both Busemann and Alexandrov started their works in the 1940s, and the two approaches gave rise to rich and fruitful developments, with no real interaction between the two. The ramifications of these two theories continue to grow today, especially since the rekindling of interest that was given to nonpositive curvature by M. Gromov in the 1970s.

Let us briefly describe the basic underlying ideas of these works. First we need to recall a few definitions. Consider a metric space X in which each point x possesses a neighborhood U such that any two points in U can be joined by a geodesic path in U. A metric space X with such a property is said to be a *locally geodesic* space. We say that such a neighborhood U is a *geodesically convex* neighborhood of x. A *geodesic segment* $[a, b]$ in X is, by definition, the image of a geodesic path in X joining a and b.[7] A *triangle* in U is the union of three geodesic segments $[a, b]$, $[a, c]$ and $[b, c]$ contained in U. The segments $[a, b]$, $[a, c]$ and $[b, c]$ are called the *sides* of this triangle.

We start by presenting Busemann's definition of nonpositive curvature, which has the advantage of being the simplest to describe, and on which we shall focus in these notes.

We say that the space X has *nonpositive curvature in the sense of Busemann* if every point x in X possesses a geodesically convex neighborhood U such that for any

[7] From now on, we use the term "geodesic" in the sense of "global geodesic", that is, a path whose length is equal to the distance between its endpoints.

geodesic triangle with sides $[a, b]$, $[a, c]$ and $[b, c]$ contained in U, we have

$$\mathrm{dist}(m, m') \leq (1/2)\mathrm{dist}(b, c),$$

where m and m' are respectively the midpoints of $[a, b]$ and $[a, c]$. This property can be stated in terms of a convexity property of the distance function, defined on the product of any two geodesic segments $[a, b] \times [a, c]$ in X equipped with their natural (barycentric) coordinates. We shall develop this point of view in later chapters.

A nonpositively curved space in the sense of Busemann is sometimes referred to as a "locally convex metric space", or "local Busemann metric space". The terminology "nonpositively curved space", in this sense, is due to Busemann.[8]

A complete Riemannian manifold of nonpositive sectional curvature is an example of a metric space of nonpositive curvature in the sense of Busemann.

The most important writings of Busemann on this subject are certainly the paper [26] and the book [28].

Now we consider the point of view of A. D. Alexandrov.

Before giving the definition of nonpositive curvature in the sense of Alexandrov, we must recall the notion of angle which two geodesic segments (or more generally two paths) in a metric space, that start at a common point, make at that point.

The notion of angle in a metric space has been introduced by Alexandrov as a generalization of the notion of angle in a surface. The first papers by A. D. Alexandrov deal with the intrinsic geometry of surfaces, and the notion of angle is certainly the most important notion in that theory, after the notion of distance and that of length of a path.[9] In fact, Alexandrov introduced several notions of angle, and these notions coincide provided some reasonable hypotheses on the ambient metric spaces are satisfied. For our needs, it suffices to consider the notion of *upper angle* that two geodesic segments with a common initial point make at that point.

Let X be a locally geodesic metric space, let U be a geodesically convex open subset of X and let us consider a geodesic triangle Δ in U, with sides $[a, b]$, $[a, c]$ and $[b, c]$. Let us define the *upper angle* α of the triangle Δ at the vertex a. For every point x on the segment $[a, b]$ and for every point y on the segment $[a, c]$, we consider a triangle $\Delta_{x,y}$ in X with sides $[a, x]$, $[a, y]$ and $[x, y]$, where $[a, x]$ and $[a, y]$ are subsegments of $[a, b]$ and $[a, c]$ respectively and where $[x, y]$ is a geodesic segment joining the points x and y. Let $\Delta^{\star}_{x,y}$ be an associated comparison triangle in the Euclidean plane and let $\alpha_{x,y}$ be the angle of $\Delta^{\star}_{x,y}$ at the vertex that corresponds to the vertex a of $\Delta_{x,y}$. Then, the *upper angle* of the triangle Δ at the vertex a is defined as

$$\alpha = \limsup_{x,y \to a} \alpha_{x,y}.$$

[8] Busemann has some additional hypotheses on the metric spaces that he considers. One such hypothesis is the uniqueness of the prolongation of geodesics. There are interesting examples of spaces that do not satisfy this hypothesis and for which the results that we are interested in are valid, and for that reason we have tried to avoid this hypothesis in this book. We recall the precise definition of the spaces considered by Busemann (which he calls G-spaces) in the Notes on Chapter 2 below.

[9] One of the basic papers of Alexandrov on that theory is [2]. We also refer the reader to [3] and [4].

The *angular excess* of the triangle Δ is then defined as

$$\delta(\Delta) = \alpha + \beta + \gamma - \pi,$$

where α, β and γ are the upper angles of Δ at the vertices a, b and c.

Finally, the space X is said to be *nonpositively curved* in the sense of Alexandrov if every point in X possesses a geodesically convex neighborhood U such that the angular excess of any triangle in U is ≤ 0.

Let us note that in the case where the space X is a differentiable surface, the angular excess $\delta(\Delta)$ is the classical *total curvature*, that is, the integral of the Gaussian curvature over the region $R \subset S$ which is bounded by the triangle Δ. Indeed, the Gauss-Bonnet theorem applied to a disk D with piecewise geodesic boundary embedded in a differentiable surface $S \subset \mathbb{R}^3$ says that

$$\tau + \omega = 2\pi\chi,$$

where $\omega = \int_D dS$ is the total curvature of the disk D, that is, the integral of the Gaussian curvature with respect to the area element of that disk. In the special case considered, χ is the Euler characteristic of the disk, which is equal to 1, and τ is the sum of the *rotations* of the boundary of the disk at the vertices. At a vertex whose angle is α, the rotation is equal to $\pi - \alpha$. This gives

$$\omega = 2\pi - (\pi - \alpha) - (\pi - \beta) - (\pi - \gamma) = \delta(\Delta).$$

Thus, the notion of nonpositive curvature in the sense of Alexandrov generalizes the classical notion of nonpositive curvature for differentiable surfaces.

Complete Riemannian manifolds with nonpositive sectional curvature are also examples of nonpositively curved metric spaces in the sense of Alexandrov.

It should be noted that a metric space which is nonpositively curved in the sense of Alexandrov is also nonpositively curved in the sense of Busemann, but that the converse is not true. For instance, any finite-dimensional normed vector space whose unit ball is strictly convex is nonpositively curved in the sense of Busemann, but if the norm of such a space is not associated to an inner product, then this space is not nonpositively curved in the sense of Alexandrov. Alexandrov mentions this example in [3], p. 197.

Let us note finally that the techniques that are used by Alexandrov in all his works rely heavily on the notion of angle in a metric space, whereas the techniques of Busemann seldom use this notion.

Convexity

To end this introduction, we would like to make a few comments on convexity theory in relation with nonpositive curvature, but before that, we mention a particular link between the study of convex polyhedra and that of the differential geometry of surfaces.

In his historical report [46], Werner Fenchel traces back the origin of Alexandrov's work on the intrinsic geometry of convex surfaces to some early work on convex polyhedra. He first recalls Cauchy's rigidity result of 1812 stating that if two combinatorially equivalent convex polyhedra in $\mathbb{R}^3$ have congruent corresponding faces, then the polyhedra are themselves congruent.[10] Fenchel then reports that Cauchy, in a note he made for the French Academy of Sciences, wrongly announced that this result on polyhedra immediately implies that there is no closed convex surface that admits isometric deformations, a result that had already been claimed, also with a false proof, by Newton, around the year 1770.[11] Still, the problem was posed, and the relation between the rigidity of convex polyhedra and the rigidity of convex surfaces was clear: by replacing the condition on the isometry between the faces by a local (infinitesimal) condition, one is naturally led to the problem of finding local conditions on two closed convex surfaces under which these surfaces are isometric. The list of mathematicians who worked on this problem includes the names of J. H. Jellett, H. Liebmann, H. Weyl, S. Cohn-Vossen, A. D. Alexandrov, A. V. Pogorelov and others. We refer the reader to the paper by Fenchel for this fascinating story.

Of course, there are also relations between convexity and nonnegative curvature. We mention as an example and without further comment the following result of Alexandrov and Pogorelov: if X is a length metric space homeomorphic to the 2-sphere, then X has nonnegative curvature if and only if X is isometric to a convex surface S in $\mathbb{R}^3$, and in this case the surface S is unique up to rigid motions of $\mathbb{R}^3$.

We turn back to the relation between convexity theory and the theory of spaces of nonpositive curvature. First of all, as we have already said, it had already been realized by Hadamard that the convexity of the distance function in a nonpositively curved space is responsible for many of the global properties of that space. We also mentioned that this idea has been extensively explored by Busemann, who defined nonpositive curvature precisely by a convexity property of the distance function, and who showed, using this new definition, that most of the important properties of a nonpositively curved Riemannian manifolds are valid in a setting which is much wider than that of Riemannian geometry. Secondly, many of the basic results in convexity theory have the flavour of nonpositive curvature, and we mention as an important class of examples the "local-implies-global" properties, such as the fact that a locally convex function is globally convex, or the fact that a local geodesic in a Busemann space (*i.e.* in a simply connected metric space that is nonpositively curved in the sense of Busemann) is a global geodesic, and there are many others.

The aim of the chapters that follow is to describe these facts in some detail.

[10] Cauchy, in his paper *Sur les polygones et les polyèdres*, cf. [37], traces back this work on polyhedra to Euclid. He says that this rigidity statement is contained in Definition 9, of Book XI of the *Elements*.

[11] Cauchy's arguments were corrected later on by H. Lebesgue among others.

Chapter 1

Lengths of paths in metric spaces

Introduction

The axioms defining a metric space concern the notion of distance between two points. By using a limiting process, these axioms lead to the notion of length of a path. The function that associates its length to each path is at the basis of a one-dimensional theory of integration in metric spaces that are "connected by rectifiable paths", that is, metric spaces in which any two points can be joined by a path of finite length.

The definition of the length of a path (even in the case where the path is a plane curve) has been a delicate question for a long period of time, in particular because it involves a limiting process. The Ancient Greeks had already worked on it. For instance, Archimedes knew how to compute the lengths of a spiral, of a circle and of other conics. Of course, he had no precise definition for the length of a path, but he was close to such a definition (see the historical remarks by Guiseppe Peano in [121]).

The basis of the modern theory of paths in metric spaces and their lengths was developed at the beginning of the 20th century, in fact, at the same time as the theory of metric spaces itself. Several major mathematicians contributed to this subject, and we can mention the names of Jordan, Fréchet, Lebesgue and Menger.

This chapter contains basic material on paths and their lengths. The results that we present are classical. Some of these results are straightforward, but we state them as propositions because of their frequent use. References for this material include the books by Choquet and Blumenthal ([39] and [18]) and Chapter 1 of Busemann [28]. We have included a few historical notes at the end of the chapter.

The outline of this chapter is the following:

In Section 1, we establish a few elementary properties of lengths of paths in a metric space. We study change of parameter and concatenation of paths and we give a characterization of the length function defined on the set of paths as being the smallest function that is additive under concatenation and whose value on any path is bounded below by the distance between its endpoints.

Section 2 concerns paths that are parametrized by arclength and paths that are parametrized proportionally to arclength.

Section 3 is a brief discussion of the classical case of $\mathcal{C}^1$-paths in Euclidean space $\mathbb{R}^n$. The length of a $\mathcal{C}^1$-path $\gamma\colon [a,b] \to \mathbb{R}^n$ is given by $L(\gamma) = \int_a^b \|\gamma'(t)\|\, dt$, a formula that is well-known from calculus.

In Section 4, we consider the set $\mathcal{C}([a,b], X)$ of paths in a metric space X whose domain is a fixed interval $[a,b]$, and we equip this set with the topology of uniform convergence. The function "length of a path", defined on $\mathcal{C}([a,b],X)$, is not continuous in general, but it is lower semi-continuous. This function is a useful tool in geometry. Using Ascoli's theorem, we derive some properties of limits of sequences of paths and their lengths.

1.1 The length of a path

Let X be a metric space. We denote by $d(x,y)$ the distance between two points x and y of X. We shall also use the notations $d(x,y) = |x-y|_X$, $d(x,y) = |x-y|_d$, or, more simply, $d(x,y) = |x-y|$ if there is no possible ambiguity.

A *path* in X is a continuous map $\gamma\colon [a,b] \to X$, where a and b are two arbitrary real numbers satisfying $a \leq b$. If $\gamma(a) = x$ and $\gamma(b) = y$, then we say that x and y are the *endpoints* of γ, and that γ *joins* the points x and y.

A *subdivision* of a compact interval $[a,b]$ is a finite subset σ of $[a,b]$ containing a and b. If $n = \mathrm{card}(\sigma)$, we say that n is the *length* of the subdivision σ. An element of σ is called a *vertex* of this subdivision. If σ is a subdivision of length $n+1$, we obtain, by taking the vertices in increasing order, a finite sequence $(t_i)_{i=0,\dots,n}$ of real numbers satisfying $a = t_0 < t_1 < \cdots < t_n = b$. In all that follows, the notation $\sigma = (t_i)_{i=0,\dots,n}$ means that σ is a subdivision, that the t_i's are the vertices of σ, that they are ordered increasingly and that they are pairwise distinct. (This makes sense only if $a \neq b$.)

To avoid cumbersome remarks, we shall tacitly assume, in all that follows, that we have $a < b$ whenever this hypothesis is necessary to make things meaningful.

Definition 1.1.1 (The length of a path). The *length* of a path $\gamma\colon [a,b] \to X$ is the quantity

$$L_X(\gamma) = L(\gamma) = \sup_{\sigma} \sum_{i=0}^{n-1} |\gamma(t_i) - \gamma(t_{i+1})| \,,$$

where the supremum is taken over the set of subdivisions $\sigma = (t_i)_{i=0,\dots,n}$ of $[a,b]$. A path is said to be *rectifiable* if its length is finite.

We have, in the general case, $0 \leq L(\gamma) \leq \infty$. By considering the subdivision $\sigma = \{a,b\}$ of $[a,b]$, we obtain immediately the inequality

$$|x-y| \leq L(\gamma), \tag{1.1.1.1}$$

where $x = \gamma(a)$ and $y = \gamma(b)$ are the endpoints of γ. Thus, the length of a path is bounded below by the distance between its endpoints.

Let $\gamma\colon [a,b] \to X$ be a path in X and let $\sigma = (t_i)_{i=0,\dots,n}$ be a subdivision of $[a,b]$. The *total variation of* γ *with respect to* σ is defined as

$$V_\sigma(\gamma) = \sum_{i=0}^{n-1} |\gamma(t_i) - \gamma(t_{i+1})|.$$

Thus, we can write the following formula for the length of an arbitrary path $\gamma\colon [a,b] \to X$:

$$L(\gamma) = \sup_{\sigma} V_\sigma(\gamma), \tag{1.1.1.2}$$

where the supremum is taken over the set of subdivisions σ of $[a,b]$.

Proposition 1.1.2. *The length of a path $\gamma\colon [a,b] \to X$ is zero if and only if γ is a constant path, that is, if and only if there exists x_0 in X such that $\gamma(t) = x_0$ for all t in $[a,b]$.*

Proof. If γ is a constant path, then $V_\sigma(\gamma) = 0$ for any subdivision σ of $[a,b]$, which implies $L(\gamma) = 0$. Conversely, suppose that $L(\gamma) = 0$. Let t be a point in $]a,b[$ and let us consider the subdivision $\sigma = \{a,t,b\}$. We have

$$V_\sigma(\gamma) = |\gamma(a) - \gamma(t)| + |\gamma(t) - \gamma(b)| \le L(\gamma) = 0.$$

Thus, we have $|\gamma(a) - \gamma(t)| = 0$, that is, $\gamma(t) = \gamma(a)$ for all t in $]a,b[$, which implies that the path γ is constant. □

Example 1.1.3 (Discrete metric space). Let X be a discrete metric space, that is, a metric space in which every point is isolated. It is not hard to see that any path in X is constant and the length of such a path is zero.

We note that we can equip an arbitrary set E with a *canonical structure of a discrete metric space*, by setting

$$|x - y| = \begin{cases} 0 & \text{if } x = y, \\ 1 & \text{if } x \neq y. \end{cases}$$

Definition 1.1.4 (Affine paths in a vector space). Let E be a vector space. For all x and y in E, the *affine path joining x and y* is the path $\gamma\colon [0,1] \to E$ defined by $\gamma(t) = (1-t)x + ty$. We note that an affine path is indeed a path, that is, it is continuous, for any metric that we shall consider on E (in fact, we shall only consider metrics that are associated to norms).

Proposition 1.1.5 (The length of an affine path in a normed vector space). *Let E be a normed vector space and let $\gamma\colon [0,1] \to E$ be an affine path joining two points x and y in E. Then we have $L(\gamma) = \|x - y\|$.*

Proof. For all t_1 and t_2 satisfying $0 \le t_1 \le t_2 \le 1$, we have

$$\begin{aligned} |\gamma(t_1) - \gamma(t_2)| &= \|(1-t_1)x + t_1 y - (1-t_2)x - t_2 y\| \\ &= \|(t_2 - t_1)x + (t_1 - t_2)y\| \\ &= (t_2 - t_1)\|x - y\|. \end{aligned}$$

Thus, if $\sigma = (t_i)_{i=0,\dots,n}$ is an arbitrary subdivision of [0, 1], we have

$$V_\sigma(\gamma) = \sum_{i=0}^{n-1} |\gamma(t_i) - \gamma(t_{i+1})| = \sum_{i=0}^{n-1} (t_{i+1} - t_i)\|x - y\| = \|x - y\|.$$

Taking the infimum of $V_\sigma(\gamma)$ over all subdivisions σ, we obtain $L(\gamma) = \|x - y\|$. □

Example 1.1.6 (Non-rectifiable path in $\mathbb{R}$). We consider the set $\mathbb{R}$ of real numbers, equipped with its usual metric (where the distance between two points is equal to the absolute value of the difference). Let us prove that the path $\gamma\colon [0, 1] \to \mathbb{R}$ defined by

$$\gamma(t) = \begin{cases} 0 & \text{if } t = 0, \\ t \sin(1/t) & \text{otherwise} \end{cases}$$

is not rectifiable. To see this, we consider, for all $n \geq 1$, the following subdivision of [0, 1]:

$$\sigma_n = \left\{ \frac{2}{\pi(2i+1)}\ ,\ i = 0, \dots, n \right\} \cup \{0, 1\}.$$

Then we have, for all $n \geq 2$,

$$\begin{aligned} V_{\sigma_n}(\gamma) &= \left| 0 - \frac{2}{\pi(2n+1)} \sin \frac{\pi(2n+1}{2} \right| \\ &\quad + \sum_{i=0}^{n-1} \left| \frac{2}{\pi(2i+1)} \sin \frac{\pi(2i+1)}{2} - \frac{2}{\pi(2i+3)} \sin \frac{\pi(2i+3)}{2} \right| \\ &\qquad + \left| \frac{2}{\pi} \sin \frac{\pi}{2} - 1 \right| \\ &\geq \sum_{i=0}^{n-1} \left(\frac{2}{\pi(2i+1)} + \frac{2}{\pi(2i+3)} \right) \\ &= \frac{2}{\pi} + \frac{2}{\pi(2n+1)} + \frac{2}{\pi} \sum_{i=0}^{n-1} \frac{1}{2i+1}. \end{aligned}$$

Since $\sum_{i=0}^{n-1} 1/(2i+1)$ is the n-th partial sum of a divergent series, we have $V_{\sigma_n}(\gamma) \to \infty$ as $n \to \infty$, which shows, by (1.1.1.2), that $L(\gamma) = \infty$.

Observe that the path γ of Example 1.1.6 is a uniform limit of a sequence of rectifiable paths $\gamma_n\colon [0, 1] \to \mathbb{R}$, where for all $n \geq 1$, γ_n is defined by

$$\gamma_n(t) = \begin{cases} 0 & \text{if } 0 \leq t \leq 1/(n\pi)\ , \\ \gamma(t) & \text{if } 1/(n\pi) \leq t \leq 1. \end{cases}$$

Thus, the uniform limit of a sequence of rectifiable paths is not necessarily rectifiable.

Definition 1.1.7 (Change of parameter). Let $\gamma\colon [a,b] \to X$ and $\gamma'\colon [c,d] \to X$ be two paths in X. We say that *γ' is obtained from γ by a change of parameter* if there exists a map $\psi\colon [c,d] \to [a,b]$ that is monotonic (in the weak sense), surjective and that satisfies $\gamma' = \gamma \circ \psi$. The map ψ is called *the change of parameter*.

Remarks. (i) We do not require that the map ψ be a homeomorphism.

(ii) A monotonic and surjective map between two intervals of $\mathbb{R}$ is necessarily continuous. Thus, ψ is continuous.

Proposition 1.1.8 (Length is invariant under change of parameter). *Let $\gamma'\colon [c,d] \to X$ be a path obtained from a path $\gamma\colon [a,b] \to X$ by a change of parameter. Then $L(\gamma) = L(\gamma')$.*

Proof. First, let us show that $L(\gamma') \geq L(\gamma)$. Let $\psi\colon [c,d] \to [a,b]$ be the change of parameter. To any subdivision $\sigma = (t_i)_{i=0,\dots,n}$ of $[a,b]$, we associate a subdivision $\sigma' = (t'_i)_{i=0,\dots,n}$ of $[c,d]$ by choosing for each $i = 0,\dots,n$ an arbitrary point t'_i in the set $\psi^{-1}(t_i)$. We have $V_{\sigma'}(\gamma') = V_\sigma(\gamma)$. Taking the supremum over all subdivisions σ' of $[c,d]$, we obtain $L(\gamma') \geq V_\sigma(\gamma)$. Now taking the supremum over all subdivisions σ of $[a,b]$, we obtain $L(\gamma') \geq L(\gamma)$.

For the converse inequality, let us consider a subdivision σ of $[c,d]$. Its image $\psi(\sigma) = \sigma'$ is a subdivision of $[a,b]$ satisfying $V_\sigma(\gamma') = V_{\sigma'}(\gamma)$, since ψ is monotonic. Taking the supremum over all subdivisions σ' of $[c,d]$, we obtain $V_\sigma(\gamma') \leq L(\gamma)$, which implies, by taking the supremum over all subdivisions σ of $[a,b]$, that $L(\gamma') \leq L(\gamma)$. Thus, we have $L(\gamma) = L(\gamma')$. This proves Proposition 1.1.8. □

Lemma 1.1.9. *Let $\gamma\colon [a,b] \to X$ be a path and let σ and σ' be two subdivisions of $[a,b]$ satisfying $\sigma \subset \sigma'$. Then $V_\sigma(\gamma) \leq V_{\sigma'}(\gamma)$.*

Proof. This is a consequence of the triangle inequality. □

Let $\sigma = (t_i)_{i=0,\dots,n}$ be a subdivision of $[a,b]$. We call the *modulus of σ* the quantity

$$|\sigma| = \sup_{i=0,\dots n-1} (t_{i+1} - t_i).$$

Proposition 1.1.10. *For every path $\gamma\colon [a,b] \to X$, we have:*

$$L(\gamma) = \lim_{|\sigma| \to 0} V_\sigma(\gamma).$$

Proof. Let M be an arbitrary real number satisfying $M < L(\gamma)$. (Note that we did not assume that the path γ is rectifiable, and therefore, $L(\gamma)$ can be equal to ∞.) To prove the proposition, it suffices to prove that there exists a real number $\eta > 0$ such that for any subdivision σ of $[a, b]$ satisfying $|\sigma| < \eta$, we have

$$M \leq V_\sigma(\gamma) \leq L(\gamma). \tag{1.1.10.1}$$

It follows from the definition of $L(\gamma)$ that the right-hand side inequality in (1.1.10.1) is always satisfied. We prove the left-hand side inequality. Let $\epsilon = (L(\gamma) - M)/2$, let $M' = M + \epsilon$ and let $\tau = (t_i)_{i=0,\dots,n}$ be a subdivision of $[a, b]$ satisfying

$$M + \epsilon < V_\tau(\gamma). \tag{1.1.10.2}$$

Since the map $\gamma : [a, b] \to X$ is uniformly continuous, we can find a real number η satisfying

$$0 < \eta < \frac{1}{4} \inf_{i=0,\dots,n-1} (t_{i+1} - t_i)$$

and such that for every u and v in $[a, b]$ satisfying $|u - v| \leq \eta$, we have

$$|\gamma(u) - \gamma(v)| \leq \frac{\epsilon}{2(n-1)}.$$

We fix such a real number η and we let σ be a subdivision of $[a, b]$ satisfying $|\sigma| < \eta$. For all $i = 1, \dots, n-1$, let t_i' (respectively t_i'') be the vertex of σ that is closest to t_i and that satisfies $t_i' \leq t_i$ (respectively $t_i < t_i''$). The inequality $|\sigma| < \eta$ implies that for all $i = 1, \dots, n-1$, we have the following sequence of inequalities:

$$t_i < t_i'' < t_{i+1}' \leq t_{i+1}.$$

Now let us consider the subdivision $\sigma \cup \tau$ of $[a, b]$. We have:

$$\begin{aligned} V_{\sigma\cup\tau}(\gamma) - V_\sigma(\gamma) &= \sum_{i=1}^{n-1} \big(|\gamma(t_i) - \gamma(t_i')| + |\gamma(t_i) - \gamma(t_i'')| - |\gamma(t_i') - \gamma(t_i'')|\big) \\ &\leq \sum_{i=1}^{n-1} \big(|\gamma(t_i) - \gamma(t_i')| + |\gamma(t_i) - \gamma(t_i'')|\big) \\ &\leq 2(n-1) \times \frac{\epsilon}{2(n-1)}. \end{aligned}$$

Hence, we obtain

$$V_{\sigma\cup\tau}(\gamma) \leq V_\sigma(\gamma) + \epsilon.$$

Since $\tau \subset \sigma \cup \tau$, we have, by Lemma 1.1.9,

$$V_\tau(\gamma) \leq V_{\sigma\cup\tau}(\gamma).$$

The last two inequalities give

$$V_\tau(\gamma) \leq V_\sigma(\gamma) + \epsilon. \tag{1.1.10.3}$$

Inequalities (1.1.10.2) and (1.1.10.3) imply $M \leq V_\sigma(\gamma)$, which is inequality (1.1.10.1). This proves Proposition 1.1.10. □

Proposition 1.1.10 is familiar in the case of Euclidean space. It allows us to see the length of a curve in this space as the limit of the perimeters of an arbitrary sequence of polygonal curves that are obtained by joining by Euclidean segments consecutive points of the image of such a curve, provided the lengths of the sides of this sequence of polygons tend uniformly to 0.

Example 1.1.11 (Koch's curve). Koch's curve (cf. [92]; see Figure 1.2) is an example of a non-rectifiable path in $\mathbb{R}^2$. This path γ is the uniform limit of a sequence of paths $\gamma_n\colon [0,1] \to \mathbb{R}^2$ ($n \geq 0$), defined inductively as follows. The path γ_0 is defined by $\gamma_0(t) = (t, 0)$ for all t in $[0, 1]$ and for every $n \geq 0$, the map γ_n is linear on each interval of the subdivision $\sigma_n = \{i/4^n,\ i = 0, 1, \dots, 4^n\}$ of $[0, 1]$. For each $n \geq 0$, the path γ_{n+1} is obtained from γ_n by first cutting each interval I of the subdivision σ_n associated to γ_n into 3 pieces of equal lengths, and then modifying the path γ_n on each of these pieces in a piecewise-linear manner so that the image by γ_{n+1} of this interval I has the form shown in Figure 1.1. In other words, using the notations of that figure, if the image of the interval I by the map γ_n is the union of the three segments OA, AC and CD, then the image of the same interval I by γ_{n+1} is the union of the four segments OA, AB, BC and CD. For all $n \geq 0$, we have $L(\gamma_{n+1}) = 4/3\, L(\gamma_n)$. Figure 1.2, represents the images of γ_n for $i = 0, 1, 2$ and 3. To see that the path γ is non-rectifiable, we consider, for every $n \geq 0$, the subdivision

$$\sigma_n = \{i/4^n, i = 0, 1, \dots, 4^n\}.$$

Then $V_{\sigma_n}(\gamma) = V_{\sigma_n}(\gamma_n) = (4/3)^n$. Thus, we have

$$L(\gamma) \geq \lim_{n\to\infty} V_{\sigma_n}(\gamma) = \infty.$$

It is easy to see, by a similar reasoning applied to subpaths of γ that no subpath of γ is rectifiable (Koch's curve is an example of a "self-similar" curve).

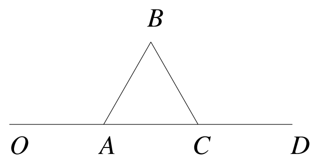


Figure 1.1. The building block of Koch's curve.

If $\gamma\colon I \to X$ is an arbitrary path and if I_0 is a closed sub-interval of I, we obtain, by restriction of γ to I_0, a path $\gamma_{|I_0}\colon I_0 \to X$. Let us note a direct consequence of Proposition 1.1.10:

Proposition 1.1.12 (Additivity of length). *Let $\gamma : [a, b] \to X$ be a path. For all c in $[a, b]$, we have*

$$L(\gamma) = L(\gamma|_{[a,c]}) + L(\gamma|_{[c,b]}).$$

Proof. We choose a sequence of subdivisions $(\sigma_n)_{n\geq 0}$ of $[a, b]$ such that $c \in \sigma_n$ for all $n \geq 0$ and such that $|\sigma_n| \to 0$ as $n \to \infty$. For all $n \geq 0$, let $\sigma'_n = \sigma_n \cap [a, c]$ and let $\sigma''_n = \sigma_n \cap [c, b]$. The set σ'_n (respectively σ''_n) is a subdivision of $[a, c]$ (respectively of $[c, b]$). We have

$$V_{\sigma_n}(\gamma) = V_{\sigma'_n}(\gamma|_{[a,c]}) + V_{\sigma''_n}(\gamma|_{[c,b]})$$

and $\lim_{n\to\infty} |\sigma'_n| = \lim_{n\to\infty} |\sigma''_n| = 0$. Using Proposition 1.1.10, we obtain:

$$\begin{aligned} L(\gamma) &= \lim_{n\to\infty} V_{\sigma_n}(\gamma) \\ &= \lim_{n\to\infty} \big(V_{\sigma'_n}(\gamma|_{[a,c]}) + V_{\sigma''_n}(\gamma|_{[c,b]})\big) \\ &= L(\gamma|_{[a,c]}) + L(\gamma|_{[c,b]}), \end{aligned}$$

which proves Proposition 1.1.12. □

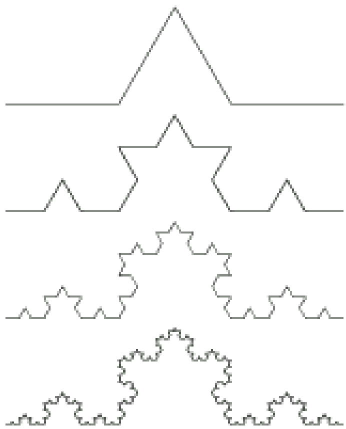


Figure 1.2. Koch's curve.

For every path $\gamma : [a, b] \to X$ and for every t in $[a, b]$, we denote by γ_t the path $\gamma|_{[a,t]}$. Then we have the following

Proposition 1.1.13. *For every rectifiable path $\gamma : [a, b] \to X$, the map $t \mapsto L(\gamma_t)$ defined on $[a, b]$ is increasing and continuous.*

Proof. The fact that this map is increasing follows from the additivity of length (Proposition 1.1.12). Let us show that it is continuous. We fix a real number $\epsilon > 0$.

Using respectively Proposition 1.1.8 (Length is invariant under change of parameter) and the uniform continuity of γ, we can find a positive real number η satisfying the following two properties:

- For every subdivision σ of $[a,b]$ satisfying $|\sigma| \leq \eta$, we have

$$L(\gamma) - V_\sigma(\gamma) \leq \epsilon; \tag{1.1.13.1}$$

- for every u and v in $[a,b]$ satisfying $|v-u| \leq \eta$, we have

$$|\gamma(u) - \gamma(v)| \leq \epsilon. \tag{1.1.13.2}$$

Now let u and v be two points in $[a,b]$ satisfying $0 \leq u - v \leq \eta$ and let σ be a subdivision of $[a,b]$ satisfying $|\sigma| \leq \eta$ and $\sigma \cap [u,v] = \{u,v\}$. We set $\sigma_1 = \sigma \cap [a,u]$ and $\sigma_2 = \sigma \cap [v,b]$. Then we have

$$\begin{aligned} L(\gamma_{|[a,u]}) + L(\gamma_{|[u,v]}) + L(\gamma_{|[v,b]}) &= L(\gamma) \\ &\leq V_\sigma(\gamma) + \epsilon \quad \text{(by 1.1.13.1)} \\ &= V_{\sigma_1}(\gamma_{|[a,u]}) + |\gamma(u) - \gamma(v)| + V_{\sigma_2}(\gamma_{|[v,b]}) + \epsilon \\ &\leq L(\gamma_{|[a,u]}) + \epsilon + L(\gamma_{|[v,b]}) + \epsilon \quad \text{(by 1.1.13.2).} \end{aligned}$$

Thus, we obtain $L(\gamma_{|[u,v]}) \leq 2\epsilon$. From this fact and the additivity of length, we conclude easily the proof of Proposition 1.1.13. □

Definition 1.1.14 (Concatenation of paths). Let a, b and c be real numbers satisfying $a \leq c \leq b$. If $\gamma_1 \colon [a,c] \to X$ and $\gamma_2 \colon [c,d] \to X$ are two paths satisfying $\gamma_1(c) = \gamma_2(c)$ then we can define the path $\gamma_1 * \gamma_2 \colon [a,b] \to X$ by setting

$$\gamma_1 * \gamma_2(t) = \begin{cases} \gamma_1(t) & \text{if } a \leq t \leq c, \\ \gamma_2(t) & \text{if } c \leq t \leq b. \end{cases}$$

This path $\gamma_1 * \gamma_2$ is called the *concatenation* of γ_1 and γ_2.

By the additivity of length (Proposition 1.1.12), we have

$$L(\gamma_1 * \gamma_2) = L(\gamma_1) + L(\gamma_2).$$

Conversely, the function "length of a path" can be characterized by some of the properties that we established:

Proposition 1.1.15 (Characterization of the length function). *Let $\mathcal{C}$ be the set of paths in X and let $L \colon \mathcal{C} \to [0,\infty]$ be the map that assigns to each path its length. Then L is the smallest map $\mathcal{L} \colon \mathcal{C} \to [0,\infty]$ satisfying the following two properties:*

(i) *for every path $\gamma \colon [a,b] \to X$, we have $\mathcal{L}(\gamma) \geq |\gamma(a) - \gamma(b)|$;*

(ii) *if a path γ is the concatenation of two paths γ_1 and γ_2, then*

$$\mathcal{L}(\gamma) = \mathcal{L}(\gamma_1) + \mathcal{L}(\gamma_2).$$

Proof. If $\mathcal{L}\colon \mathcal{C} \to [0, \infty[$ satisfies (i) and (ii) and if $\sigma = (t_i)_{i=0,\dots,n}$ is an arbitrary subdivision of $[a, b]$, then we have

$$\mathcal{L}(\gamma) \geq \sum_{i=0}^{n-1} \mathcal{L}(\gamma_{|[t_i,t_{i+1}]}| \geq \sum_{i=0}^{n-1} |\gamma(t_i) - \gamma(t_{i+1})| \geq V_\sigma(\gamma),$$

which implies that $\mathcal{L}(\gamma) \geq L(\gamma)$. Since the map L satisfies properties (i) and (ii), the proof of Proposition 1.1.15 follows. □

1.2 Arclength as parameter

Proposition 1.2.1. *Let $\gamma : [a, b] \to X$ be a rectifiable path. For each u in $[0, L(\gamma)]$, there exists a unique point x in X and a point t in $[a, b]$ such that $x = \gamma(t)$, where $L(\gamma_t) = u$. Furthermore, the set of such points t associated to u is a closed sub-interval of $[a, b]$, and the map γ is constant on this sub-interval.*

Proof. Since the map $t \mapsto L(\gamma_t)$ is continuous (Proposition 1.1.13), we can use the mean value theorem to obtain, for every u satisfying $0 \leq u \leq L(\gamma)$, a real number t satisfying $L(\gamma_t) = u$. Now if t and t' are two real numbers satisfying $a \leq t \leq t' \leq b$ and $L(\gamma_t) = L(\gamma_{t'})$, then, by Proposition 1.1.12 (Additivity of length), we have

$$L(\gamma_{|[t,t']}) = L(\gamma_{t'}) - L(\gamma_t) = 0.$$

Thus, we obtain $L(\gamma_{|[t,t']}) = 0$, which, by Proposition 1.1.2, implies that γ is constant on the interval $[t, t']$. This also proves that the set of real numbers t associated to u is a sub-interval of $[a, b]$. By the continuity of the map $t \mapsto L(\gamma_t)$, this sub-interval is closed. □

Proposition 1.2.2. *Let $\gamma\colon [a, b] \to X$ be a rectifiable path. Consider the map $\lambda\colon [0, L(\gamma)] \to X$ defined by $\lambda(u) = \gamma(t)$, where $\gamma(t)$ is the unique point of X provided by Proposition 1.2.1, satisfying $L(\gamma_t) = u$. Then the map λ is 1-Lipschitz. In particular, λ is continuous and therefore it is a path. Furthermore, γ is the path obtained from λ by the change of parameter $\psi\colon [a, b] \to [0, L(\gamma)]$ defined by $\psi(t) = L(\gamma_t)$.*

Proof. Let u and u' be two points in $[0, L(\gamma)]$ satisfying $u \leq u'$ and let t and t' be two points in $[a, b]$ satisfying $L(\gamma_t) = u$ and $L(\gamma_{t'}) = u'$. Then, $\lambda(u) = \gamma(t)$ and $\lambda(u') = \gamma(t')$. Using (1.1.1.1), we obtain

$$|\lambda(u) - \lambda(u')| = |\gamma(t) - \gamma(t')| \leq L(\gamma_{|[t,t']})$$

$$= |u - u'| = L(\gamma_{t'}) - L(\gamma_t) = u' - u,$$

which shows that λ is a 1-Lipschitz map.

The map ψ is increasing and surjective. The uniqueness of the point x in Proposition 1.2.1 implies that $\gamma = \lambda \circ \psi$. This completes the proof of Proposition 1.2.2. $\square$

Proposition 1.2.3. *Let $\gamma\colon [a,b] \to X$ be a rectifiable path and let $\lambda\colon [0, L(\gamma)] \to X$ be the associated map, defined in Proposition* 1.2.2. *Then we have, for all u in $[0, L(\gamma)]$, $u = L(\lambda_u)$.*

Proof. Let u be a point in $[0, L(\gamma)]$ and let t be a point in $[a,b]$ satisfying $L(\gamma_t) = u$. We consider the paths $\lambda_u = \lambda_{|[0,u]}$ and $\gamma_t = \gamma_{|[a,t]}$. By construction, we have $\gamma_t = \lambda_u \circ \psi_t$, where $\psi_t\colon [0,t] \to [0,u]$ is the map $x \mapsto L(\gamma_x)$. Thus, the path γ_t is obtained from the path λ_u by the change of parameter ψ_t. By Proposition 1.1.8, these two paths have the same length, which implies $L(\lambda_u) = u$. $\square$

Definition 1.2.4 (Path parametrized by arclength). Let $\gamma\colon [a,b] \to X$ be a rectifiable path. We say that γ is *parametrized by arclength* if for all u and v satisfying $a \leq u \leq v \leq b$, we have $v - u = L(\gamma_{|[u,v]})$.

In particular, if a path $\gamma\colon [a,b] \to X$ is parametrized by arclength, it satisfies $L(\gamma) = b - a$.

Proposition 1.2.5. *Let $\gamma\colon [a,b] \to X$ be a path parametrized by arclength. Then the map $t \mapsto L(\gamma_t)$, defined on the interval $[a,b]$, is strictly increasing.*

Proof. The proof follows directly from Definition 1.2.4. $\square$

Corollary 1.2.6. *Let $\gamma\colon [a,b] \to X$ be a rectifiable path. Then the path*

$$\lambda\colon [0, L(\gamma)] \to X$$

that is associated to it by Proposition 1.2.2 *is parametrized by arclength.*

Proof. By Proposition 1.2.3, for all u and v satisfying $0 \leq u \leq v \leq L(\gamma)$, we have

$$v - u = L(\lambda_{|[0,v]}) - L(\lambda_{|[0,u]}) = L(\lambda_{|[u,v]}).$$ $\square$

For every rectifiable path $\gamma\colon [a,b] \to X$, we call $\lambda\colon [0, L(\gamma)] \to X$ *the path parametrized by arclength that is associated to γ.*

Proposition 1.2.7 (Concatenation of paths parametrized by arclength). *If γ is the concatenation of two paths γ_1 and γ_2 that are parametrized by arclength, then γ is also parametrized by arclength.*

Proof. We use the notations of Definition 1.1.14, with $\gamma = \gamma_1 * \gamma_2$. Let u and v be two real numbers satisfying $a \leq u \leq v \leq b$. We distinguish three cases:

If $a \leq u \leq v \leq c$, then, since γ_1 is parametrized by arclength, we have

$$v - u = L(\gamma_{1|[u,v]}) = L(\gamma_1 * \gamma_{2|[u,v]}).$$

If $c \leq u \leq v \leq b$, then, since γ_2 is parametrized by arclength, we have

$$v - u = L(\gamma_{2|[u,v]}) = L(\gamma_1 * \gamma_{2|[u,v]}).$$

If $a \leq u \leq c \leq v \leq b$, then we write

$$\begin{aligned} v - u &= (v - c) + (c - u) \\ &= L(\gamma_{1|[u,c]}) + L(\gamma_{2|[c,v]}) \\ &= L(\gamma_1 * \gamma_{2|[u,c]}) + L(\gamma_1 * \gamma_{2|[c,v]}) \\ &= L(\gamma_1 * \gamma_{2|[u,v]}). \end{aligned}$$

This proves Proposition 1.2.7. □

Definition 1.2.8 (Path parametrized proportionally to arclength). Let $\gamma : [a, b] \to X$ be a path, with $a < b$. We say that γ is *parametrized proportionally to arclength* if either γ is a constant path, or there exists a path $\gamma' : [c, d] \to X$ that is parametrized by arclength and that satisfies $\gamma = \gamma' \circ \psi$, where $\psi : [a, b] \to [c, d]$ is the unique affine homeomorphism between these two intervals, that is, the map defined by $\psi(x) = \big((d - c)x + (bc - ad)\big)/(b - a)$.

La following property will be useful for us later on.

Proposition 1.2.9. *Let $\gamma : [0, 1] \to X$ be a path parametrized proportionally to arclength. Then γ is an $L(\gamma)$-Lipschitz map.*

Proof. If $\gamma : [0, 1] \to X$ is a constant path, then the conclusion follows trivially. Now suppose that $\gamma = \gamma' \circ \psi$, where $\gamma' : [c, d] \to X$ is a path that is parametrized by arclength and where $\psi : [0, 1] \to [c, d]$ is defined by $\psi(x) = (d - c)x + c$. Then, for all u and v satisfying $0 \leq u \leq v \leq 1$, we have

$$\begin{aligned} |\gamma(u) - \gamma(v)| &= |\gamma'\big(\psi(u)\big) - \gamma'\big(\psi(v)\big)| \\ &\leq L(\gamma'_{|[\psi(u),\psi(v)]}) \text{ (using (1.1.1.1))} \\ &= \psi(v) - \psi(u) \text{ (since } \gamma' \text{ is parametrized by arclength)} \\ &= (d - c)(v - u) \\ &= L(\gamma')(v - u) \\ &= L(\gamma)(v - u) \end{aligned}$$

This proves Proposition 1.2.9. □

1.3 Differentiable paths in Euclidean space

For all $n \geq 1$, we denote by $\mathbb{E} = \mathbb{E}^n$ the Euclidean space of dimension n, that is, the space $\mathbb{R}^n$ equipped with the norm $\|(x_1, \dots, x_n)\| = \sqrt{x_1^2 + \cdots + x_n^2}$ and with the metric induced by that norm.

Proposition 1.3.1. *Let $\gamma \colon [a, b] \to \mathbb{E}$ be a $\mathcal{C}^1$-path and let $\gamma' \colon [a, b] \to \mathbb{E}$ be its derivative. Then we have:*

(1.3.1.1)
$$L(\gamma) = \int_a^b \|\gamma'(t)\| \, dt.$$

Proof. Let us equip the space $\mathbb{E}$ with an orthonormal basis and let $\gamma_1, \dots, \gamma_n$ be the components of γ in this basis. Then we can write

$$\|\gamma'(t)\| = \sqrt{\sum_{j=1}^{n} \gamma_j'(t)^2}.$$

For all t in $[a, b]$, we set, as usual, $\gamma_t = \gamma_{|[a,t]}$, we consider the map $s \colon [a, b] \to \mathbb{R}$ defined by $s(t) = L(\gamma_t)$ and we show that this map is differentiable and that its derivative is equal to $\|\gamma'(t)\|$. This, combined with the fact that $s(a) = 0$, will give Formula (1.3.1.1).

Let t and t' be two real numbers satisfying $a \leq t < t' \leq b$. We have

(1.3.1.2)
$$s(t') - s(t) = L(\gamma_{|[t,t']}).$$

Let us fix a real number $\epsilon > 0$. For all $j = 1, \dots, n$, the map $\gamma_j' \colon [a, b] \to \mathbb{R}$ is uniformly continuous. Therefore, there exists $\eta > 0$ such that for all t and t' satisfying $a \leq t < t' \leq b$ and $t' - t < \eta$, we have, for all $j = 1, \dots, n$ and for all τ in $[t, t']$,

$$\gamma_j'(\tau)^2 \leq \gamma_j'(t)^2 + \epsilon.$$

Now, let us take t and t' satisfying $a \leq t < t' \leq b$ and $t' - t < \eta$ and let $\sigma = (t_i)_{i=0,\dots,k}$ be an arbitrary subdivision of $[t, t']$. We have

$$V_\sigma(\gamma_{|[t,t']}) = \sum_{i=0}^{k-1} |\gamma(t_i) - \gamma(t_{i+1})| = \sum_{i=0}^{k-1} \sqrt{\sum_{j=1}^{n} \big(\gamma_j(t_i) - \gamma_j(t_{i+1})\big)^2}.$$

By the mean value theorem, for all $i = 0, \dots, k$ and for all $j = 1, \dots, n$, we can find $\tau_{i,j} \in [t_i, t_{i+1}]$ such that

$$\gamma_j(t_i) - \gamma_j(t_{i+1}) = \gamma_j'(\tau_{i,j})(t_i - t_{i+1}),$$

which implies

$$\big(\gamma_j(t_i) - \gamma_j(t_{i+1})\big)^2 = \gamma_j'(\tau_{i,j})^2 (t_{i+1} - t_i)^2,$$

whence

$$\begin{aligned}\sum_{j=1}^{n} \big(\gamma_j(t_i) - \gamma_j(t_{i+1})\big)^2 &\leq \sum_{j=1}^{n} (\gamma_j'(t)^2 + \epsilon)(t_{i+1} - t_i)^2 \\ &= (n\epsilon + \sum_{j=1}^{n} \gamma_j'(t)^2)(t_{i+1} - t_i)^2 \\ &= (n\epsilon + \|\gamma'(t)\|^2)(t_{i+1} - t_i)^2.\end{aligned}$$

Thus, we obtain

$$V_\sigma(\gamma_{|\,[t,t']}) \leq \sum_{i=0}^{k-1} \sqrt{n\epsilon + \|\gamma'(t)\|^2}(t_{i+1} - t_i) = \sqrt{n\epsilon + \|\gamma'(t)\|^2}(t' - t).$$

The right hand side in the last expression does not depend on the choice of σ. Thus, we have:

$$L(\gamma_{|\,[t,t']}) \leq \sqrt{n\epsilon + \|\gamma'(t)\|^2}(t' - t).$$

By (1.3.1.2), we therefore obtain:

$$\frac{s(t') - s(t)}{t' - t} \leq \sqrt{n\epsilon + \|\gamma'(t)\|^2}. \tag{1.3.1.3}$$

On the other hand, we have, using (1.1.1.1) and (1.3.1.2),

$$|\gamma(t') - \gamma(t)| \leq |s(t') - s(t)|.$$

Since s is increasing (Proposition 1.1.13), we obtain

$$\frac{|\gamma(t') - \gamma(t)|}{t' - t} \leq \frac{s(t') - s(t)}{t' - t},$$

or, equivalently,

$$\left\| \frac{\gamma(t') - \gamma(t)}{t' - t} \right\| \leq \frac{s(t') - s(t)}{t' - t}. \tag{1.3.1.4}$$

Inequations (1.3.1.3) and (1.3.1.4), that we proved for $a \leq t < t' \leq b$ with $t - t' < \eta$, are valid for all t and t' in $[a, b]$ satisfying $t \neq t'$ and $|t - t'| < \eta$ (to see this, use the fact that s is increasing).

We have

$$\lim_{|t-t'|\to 0} \left\| \frac{\gamma(t') - \gamma(t)}{t' - t} \right\| = \|\gamma'(t)\|.$$

Therefore, there exists $\eta' > 0$ such that for all $|t - t'| < \eta'$, we have

$$\|\gamma'(t)\| - \epsilon \leq \frac{s(t') - s(t)}{t' - t}.$$

The last inequality, together with (1.3.1.3), imply that for all $\epsilon > 0$ and for all t and t' in $[a, b]$ satisfying $t \neq t'$ and $|t - t'| < \inf\{\eta, \eta'\}$, we have

$$\|\gamma'(t)\| - \epsilon \leq \frac{s(t') - s(t)}{t' - t} \leq \sqrt{n\epsilon + \|\gamma'(t)\|^2}.$$

Letting ϵ tend to 0, we obtain $s'(t) = \|\gamma'(t)\|$, which completes the proof of Proposition 1.3.1. □

1.4 The space of paths

At several places in these notes, we need to take limits of paths and for that reason, we begin by defining a space of paths. In all this section, X is a metric space and $[a, b]$ is a compact interval of $\mathbb{R}$.

We denote by $\mathcal{C}([a, b])$ the set of paths in X with domain $[a, b]$ and we equip this set with the topology of uniform convergence. We note that this topology is associated to the metric defined by

$$|\gamma_1 - \gamma_2| = \sup_{t \in [a,b]} |\gamma_1(t) - \gamma_2(t)|,$$

for every γ_1 and $\gamma_2 \in \mathcal{C}([a, b], X)$.

We start by observing that the map L from $\mathcal{C}([a, b], X)$ to the extended ray $[0, \infty] \cup \{\infty\}$ that associates to each path γ its length $L(\gamma)$ is not continuous in general, as the following examples show:

Examples 1.4.1 (Non-continuity of the length function).

(i) For every nonpositive integer n, let $F_n : [0, 1] \to \mathbb{R}$ be the piecewise-affine continuous map whose graph is the union of the affine segments joining pairwise (in consecutive order with respect to the first coordinate) the points of the following sequence in the plane $\mathbb{R}^2$

$$\left(\frac{p}{2^n}, \frac{\epsilon(p)}{2^n}\right) \quad (0 \leq p \leq 2^n),$$

where $\epsilon(p) = 0$ if p is odd and 1 if p is even.

For all $n \geq 0$ and for all t in $[0, 1]$, let $\gamma_n : [0, 1] \to \mathbb{E}^2$ be the map defined by

$$\gamma_n(t) = (t, F_n(t)).$$

We have represented in Figure 1.3 the images of the paths γ_1, γ_2, γ_3 and γ_4.

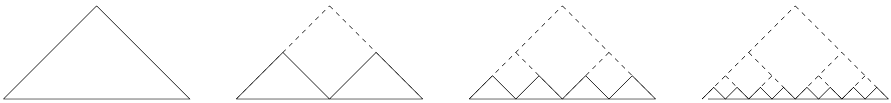


Figure 1.3. The images of the paths γ_1, γ_2, γ_3 and γ_4 of Example 1.4.1 (i).

The sequence of paths (γ_n) converges to the path $\gamma\colon [0,1] \to \mathbb{E}^2$ defined by $\gamma(t) = (t,0)$, whose length is 1, whereas $L(\gamma_n) = \sqrt{2}$ for every $n \geq 0$. Thus, $L(\gamma_n) \not\to L(\gamma)$.

(ii) Let $\gamma_n\colon [0,\pi] \to \mathbb{R}$ be the path defined by

$$\gamma_n(t) = (1/n)\cos(n^2 t).$$

When t varies in the interval $[0,\pi]$, $\cos(n^2 t)$ takes n^2 times the values 1 and -1. Hence,

$$L(\gamma_n) \geq \frac{1}{n} \times 2n^2 = 2n.$$

The sequence of paths γ_n converges uniformly to the constant path $\gamma\colon [0,1] \to \mathbb{R}$ defined by $\gamma(t) = 0$ for all t, whose length is zero, whereas $L(\gamma_n) \to \infty$. Thus, we also have $L(\gamma_n) \not\to L(\gamma)$.

(iii) For every $n \geq 1$, consider the path $\gamma_n\colon [0,1] \to \mathbb{R}$ defined by

$$\gamma_n(t) = (t/n)\sin(1/t).$$

We have $L(\gamma_n) = \infty$ for all n (Example 1.1.6). On the other hand, we have $|\gamma_n(t)| \leq 1/n$ for all t in $[0,1]$, which shows that the sequence γ_n converges uniformly to the constant path $\gamma(t) = 0$, whose length is zero. Thus, we again have $L(\gamma_n) \not\to L(\gamma)$.

We note that we have in all these examples $0 = L(\gamma) \leq \liminf_{n\to\infty} L(\gamma_n) = \infty$. This is related to the fact that the length function is lower semi-continuous, as we shall see next. Before that, we recall the following classical definition:

Let E be a topological space and let x_0 be a point in E. A map $f\colon E \to \overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ is said to be *lower semi-continuous at* x_0 if for any real number $m < f(x_0)$, there exists a neighborhood W of x_0 in E such that for all x in W, we have $m \leq f(x)$. The map f is said to be *lower semi-continuous* on E if it is lower semi-continuous at every point in E.

Let us note the following classical examples of lower semi-continuous maps.

Examples 1.4.2 (Lower semi-continuous maps).

(i) If $f\colon E \to \overline{\mathbb{R}}$ is continuous at x_0 in E, then f is lower semi-continuous at this point.

(ii) The function $x \mapsto E(x)$ (integral value of x) is lower semi-continuous on $\mathbb{R}$. One can easily deduce this fact from the definitions, but we can also derive it from the following proposition, which is often useful in the context of lower semi-continuous functions.

Proposition 1.4.3 (Upper limit). *Let $(f_i)_{i\in\mathcal{I}}$ be a family of maps from E to $\overline{\mathbb{R}}$ and let $f\colon E \to \overline{\mathbb{R}}$ be the upper limit of this family, that is, the map defined by $f(x) = \sup_{i\in\mathcal{I}} f_i(x)$ for all x in E. Let x_0 be a point in E. If each f_i is lower semi-continuous at x_0, then f is also lower semi-continuous at x_0.*

Proof. For every $m < f(x_0)$, there exists i_0 in $\mathcal{I}$ such that $m < f_{i_0}(x_0)$. Since f_{i_0} is lower semi-continuous at x_0, we can find a neighborhood W of x_0 such that for all x in W, we have $m \leq f_{i_0}(x)$. Therefore, for all x in W, we have $m \leq f_{i_0}(x) \leq f(x)$. This completes the proof of Proposition 1.4.3. □

We deduce the following result, which is due to Menger.

Theorem 1.4.4 (Lower semi-continuity of the length function). *The length function $L\colon \mathcal{C}([a,b],X) \to \mathbb{R} \cup \{\infty\}$ is lower semi-continuous.*

Proof. Let us set $E = \mathcal{C}([a,b],X)$ and let us fix a point t in $[a,b]$. For any γ and γ' in E, we have $|\gamma(t) - \gamma'(t)| \leq |\gamma - \gamma'|_E$ and therefore the map $E \to \mathbb{R}$ defined by $\gamma \mapsto \gamma(t)$ is continuous. For each subdivision $\sigma = (t_i)_{i=0,\dots,n}$ of $[a,b]$, the map $V_\sigma\colon E \to \mathbb{R}$ defined by $\gamma \mapsto V_\sigma(\gamma) = \sum_0^{n-1} |\gamma(t_{i-1}) - \gamma(t_i)|$ is a sum of continuous maps, and therefore it is continuous. Therefore, by Proposition 1.4.3, the map $\gamma \mapsto L(\gamma) = \sup_\sigma V_\sigma(\gamma)$ is lower semi-continuous. This proves Theorem 1.4.4. □

Let us note a few consequences of Theorem 1.4.4:

Corollary 1.4.5. *Let $(\gamma_n\colon [a,b] \to X)_{n\geq 0}$ be a sequence of paths converging uniformly to a path $\gamma\colon [a,b] \to X$. Then $L(\gamma) \leq \liminf_{n\to\infty} L(\gamma_n)$.*

Proof. This follows from Theorem 1.4.4 and from the following lemma:

Lemma 1.4.6. *Let E be a topological space, let $f\colon E \to \overline{\mathbb{R}}$ be a map that is lower semi-continuous at some point x in E and let $(x_n)_{n\geq 0}$ be a sequence in E converging to x. Then $f(x) \leq \liminf_{n\to\infty} f(x_n)$.*

Proof. Since f is lower semi-continuous at x, then, for every real number m satisfying $m < f(x)$, there exists a neighborhood W of x in E such that for all y in W, we have $m \leq f(y)$. Since x_n converges to x as $n \to \infty$, for every n large enough, x_n is in W, which implies $m \leq f(x_n)$. Thus, $f(x) \leq \liminf_{n\to\infty} f(x_n)$. □

Corollary 1.4.7. *If $(\gamma_n\colon [a,b] \to X)_{n\geq 0}$ is a sequence of rectifiable paths that converges uniformly to a path γ and if there exists a real number M such that $L(\gamma_n) \leq M$ for all $n \geq 0$, then γ is rectifiable.*

Proof. This is an immediate consequence of Corollary 1.4.5. □

Corollary 1.4.8. *For any real number M, the set of paths $\gamma\colon [a,b] \to X$ such that $L(\gamma) \leq M$ is a closed subset of $\mathcal{C}([a,b], X)$.*

Proof. The uniform limit of a sequence of continuous maps is continuous. Furthermore, if $\gamma_n\colon [a,b] \to X$ $(n \geq 0)$ is a sequence of paths converging uniformly to a path γ and if $L(\gamma_n) \leq M$ for all n, then we have, by Corollary 1.4.5, $L(\gamma) \leq M$. □

Before going through other results on limits of sequences of paths, we recall the notion of uniform equicontinuity of a sequence of maps, and the theorem of Ascoli that is intimately related to this notion.

Let X and Y be two metric spaces. A sequence of maps $(f_n)_{n\geq 0}$ from X to Y is said to be *uniformly equicontinuous* [1] if for every $\epsilon > 0$, there exists an $\eta > 0$ such that for every integer $n \geq 0$ and for every x and y in X, we have

$$|x - y| < \eta \Rightarrow |f_n(x) - f_n(y)| < \epsilon.$$

An example of a uniformly equicontinuous sequence is a sequence $(f_n)_{n\geq 0}$ of K-Lipschitz maps, with K independent of n or, more generally, a sequence of uniformly Hölder maps. More generally, if there exist two constants $\alpha > 0$ and $K \geq 0$ such that for any integer $n \geq 0$ and for all x and y in X we have $|f_n(x) - f_n(y)| \leq K|x - y|^\alpha$, then the sequence (f_n) is uniformly equicontinuous.

The notion of equicontinuity is due to Ascoli. Theorem 1.4.9 below is part of a series of results on the convergence of sequences of continuous maps between topological spaces, that are due to Ascoli and Arzelà. In [6], Ascoli gave sufficient conditions for certain spaces of functions to be compact, and Arzelà showed in [5] that these conditions are necessary. Versions of the theorems of Ascoli-Arzelà adapted to the setting of metric spaces were obtained by Fréchet in his thesis [49] that we already mentioned.

Before stating that theorem, we recall two definitions.

A metric space X is said to be *proper* if every closed bounded subset of X is compact.[2] Equivalently, a space X is proper if from every infinite bounded sequence

[1] Some authors call this property "equicontinuity". Even though we shall only use the property of uniform equicontinuity, it is preferable to reserve the term "equicontinuity" to the following local property: we say that the sequence of maps $(f_n)_{n\geq 0}$ from X to Y is equicontinuous at some point x of X if for all $\epsilon > 0$, there exists an $\eta > 0$ such that for every integer $n \geq 0$ and for all y in X satisfying $|x - y| < \eta$, we have $|f_n(x) - f_n(y)| < \epsilon$. Then the sequence of maps $(f_n)_{n\geq 0}$ is said to be equicontinuous if it is equicontinuous at every point of X.

[2] A more classical terminology, for proper, is "finitely compact". For instance, this is the terminology used by Busemann. We note also that E. Cartan, in [36], calls such a space a "normal space".

of points in X we can extract a convergent subsequence. Thus, a proper space satisfies an analog of the Bolzano–Weierstrass property for sequences of points in $\mathbb{R}^n$ (or, more generally, in a finite-dimensional normed vector space), namely, that any infinite bounded sequence has a convergent subsequence.

An example of a non-proper space is the set $\mathbb{Q}$ of rational numbers equipped with the metric induced from that of $\mathbb{R}$.

We also note the following equivalent property, which explains the use of the word "proper": a metric space X is proper if for every x_0 in X, the map $d_{x_0}: X \to [0, \infty[$ defined by $x \mapsto |x_0 - x|$ is proper in the usual sense (the inverse image of a compact subset of $[0, \infty[$ is a compact subset of X).

Clearly, a closed subset of a proper metric space, equipped with the induced metric, is proper and a proper metric space is complete.

We recall that a metric space X is said to be *separable* if it contains a countable dense subset. It is easy to see that a proper metric space is separable. (First, prove this for a compact space.)

The following result is part of a series of results which, as we said above, are known under the name of "Ascoli's theorem" or "Arzelà–Ascoli's theorem". The particular case that we consider here is sufficient for our needs. We shall use it essentially in two contexts: in the study of limits of geodesic paths in metric spaces and in the study of isometries of metric spaces.

Theorem 1.4.9 (Ascoli). *Let Y be a separable metric space, let X be a proper metric space and let $(f_n)_{n\geq 0}$ be a uniformly equicontinuous sequence of maps from Y to X such that for each y in Y, the sequence $(f_n(y))_{n\geq 0}$ is bounded. Then there exists a subsequence of (f_n) that converges uniformly on every compact subset of Y to a map $f : Y \to \mathbb{R}$, and the limit map f is uniformly continuous.*

Proof. We can assume that the set Y is infinite, otherwise the theorem is obvious. Then, let D be an infinite dense countable subset of Y. We start by showing that up to passing to a subsequence, the sequence $(f_n(x))_{n\geq 0}$ is convergent for every x in D. To this end, we use an argument which is known as "Cantor's diagonal process". We start by indexing the points of D by the set of integers that are ≥ 1 and we write henceforth $D = \{x_1, x_2, \dots\}$. We first consider the point x_1. Since X is proper and since the sequence $(f_n(x_1))_{n\geq 0}$ is bounded, it has a convergent subsequence. Let $(f_{n_1}(x_1))_{n\geq 0}$ be such a convergent subsequence. Next, we consider the point x_2. The sequence $(f_{n_1}(x_2))_{n\geq 0}$ being also bounded, it has a convergent subsequence. Let $(f_{n_2}(x_2))_{n\geq 0}$ be such a convergent subsequence. Continuing in the same way, we obtain, for every integer $k \geq 1$, a subsequence $(f_{n_k})_{n\geq 0}$ of $(f_{n_{k-1}})_{n\geq 0}$ that has the property that for all $i = 1, \dots, k$, the sequence $(f_{n_k}(x_i))_{n\geq 0}$ in X is convergent. The "diagonal sequence" $(f_{n_n})_{n\geq 0}$ is then a subsequence of $(f_n)_{n\geq 0}$ that has the property that for all x in D, the sequence $(f_{n_n}(x))_{n\geq 0}$ is convergent.

Thus we can assume, up to replacing the sequence $(f_n)_{n\geq 0}$ by a subsequence, that the sequence $(f_n(x))_{n\geq 0}$ is convergent for every x in D.

Let ϵ be a positive real number. Since the sequence $(f_n)_{n\geq 0}$ is uniformly equicontinuous, there exists a positive real number δ such that for all x and y in Y satisfying $|x-y| < \delta$ and for every integer $n \geq 0$, we have $|f_n(x) - f_n(y)| < \epsilon$.

Given a point x in Y, let us choose a point y in D satisfying $|x-y| < \delta$. We have, for all nonnegative integers m and n,

$$(1.4.9.1)\quad |f_m(x) - f_n(x)| \leq |f_m(x) - f_m(y)| + |f_m(y) - f_n(y)| + |f_n(y) - f_n(x)|.$$

Since the sequence $(f_n(y))_{n\geq 0}$ is convergent, there exists an integer n_0 such that for all $m \geq n_0$ and $n \geq n_0$, we have $|f_m(y) - f_n(y)| < \epsilon$. Using (1.4.9.1), if $m \geq n_0$ and $n \geq n_0$, we obtain $|f_m(x) - f_n(x)| \leq 3\epsilon$. Therefore, the sequence $(f_n(x))_{n\geq 0}$ is a Cauchy sequence. Since X is proper, this sequence converges. Now we set, for every x in X, $f(x) = \lim_{n\to\infty} f(x)$.

Let us show that the map f is uniformly continuous. By our choice of the real number δ, we have, for every x and y in X satisfying $|x-y| < \delta$ and for every integer $n \geq 0$, $|f_n(x) - f_n(y)| < \epsilon$. Letting n tend to infinity, we obtain $|f(x) - f(y)| \leq \epsilon$. This shows that f is uniformly continuous.

Finally, let us prove that the convergence of $(f_n)_{n\geq 0}$ is uniform on every compact subset of Y.

Let $K \subset Y$ be a compact set. Since D is dense in Y, there exists an integer $n_1 > 0$ such that for all x in K, we can find a point x^* in the subset $\{x_1, \ldots, x_{n_1}\}$ of D satisfying $|x - x^*| < \delta$.

Since $\{x_1, \ldots, x_{n_1}\}$ is a finite set, we can find an integer n_2 such that for all $n \geq n_2$ and for all i satisfying $1 \leq i \leq n_1$, we have $|f_n(x_i) - f(x_i)| < \epsilon$.

The integer n_2 does not depend on the choice of x in K, and we have, for all x in K and for all $n \geq n_2$,

$$|f(x) - f_n(x)| < |f(x) - f(x^*)| + |f(x^*) - f_n(x^*)| + |f_n(x^*) - f_n(x)| \leq 3\epsilon.$$

This proves the uniform convergence of $(f_n(x))_{n\geq 0}$ to f on K, and this completes the proof of Theorem 1.4.9. □

Let us note a few consequences of Theorem 1.4.9, applied to the cases where the maps f_n are paths:

Proposition 1.4.10. *Let X be a compact metric space, let M be a nonnegative real number and for all $n \geq 0$, let $\gamma_n \colon [a,b] \to X$ be a path parametrized proportionally to arclength and satisfying $L(\gamma_n) \leq M$. Then the sequence of paths $(\gamma_n)_{n\geq 0}$ has a uniformly convergent subsequence, and the length of the limit path is $\leq M$.*

Proof. By Proposition 1.2.9, the map $\gamma_n \colon [a,b] \to X$ is M-Lipschitz, and the sequence of maps $(\gamma_n)_{n\geq 0}$ is therefore uniformly equicontinuous. Furthermore, for all t in $[a,b]$, the sequence $(\gamma_n(t))_{n\geq 0}$ is bounded, since X is compact. Thus, the existence of a convergent subsequence follows from Ascoli's Theorem (Theorem 1.4.9), and the fact that the length of the limit path is $\leq M$ follows from Corollary 1.4.8. □

Proposition 1.4.11. *Let X be a proper metric space, let M be a nonnegative real number and for all $n \geq 0$, let $\gamma_n : [a, b] \to X$ be a path parametrized proportionally to arclength and satisfying $L(\gamma_n) \leq M$. Suppose furthermore that the subset $\{\gamma_n(a), n \geq 0\}$ of X is bounded. Then the sequence of paths $(\gamma_n)_{n\geq 0}$ has a subsequence that converges uniformly to a path γ whose length is also $\leq M$.*

Proof. Since γ_n is an M-Lipschitz map for all $n \geq 0$, the sequence of maps $(\gamma_n)_{n\geq 0}$ is uniformly equicontinuous. Since the sequence $(\gamma_n(a))_{n\geq 0}$ is bounded and since the space X is proper, we can assume, up to replacing the sequence $(\gamma_n)_{n\geq 0}$ by a subsequence, that $(\gamma_n(a))_{n\geq 0}$ is convergent. Let x be the limit of this sequence. Then for all $n \geq 0$ and for all t in $[0, 1]$, we have

$$|x - \gamma_n(t)| \leq L(\gamma_n)|\gamma_n(0) - \gamma_n(t)| \leq L(\gamma_n)|t - a| \leq M(b - a).$$

Thus, for all t in $[0, 1]$, the sequence $(\gamma_n(t))_{n\geq 0}$ is bounded. By Theorem 1.4.9, there exists a subsequence of γ_n that converges uniformly to a path γ. By Corollary 1.4.8, we have $L(\gamma) \leq \liminf_{n\to\infty} L(\gamma_n) \leq M$. □

Proposition 1.4.12 (Existence of paths of minimal length). *Let X be a proper metric space, let x and y be two points in X and suppose that there exists a rectifiable path in X joining x and y. Then there exists such a path whose length is equal to the infimum of the lengths of paths that join x and y.*

Proof. Let $\alpha = \inf\{L(\gamma)$ where γ is a path in X joining x and $y\}$, and let $(\gamma_n)_{n\geq 0}$ be a sequence of paths joining x and y and satisfying $L(\gamma_n) \to \alpha$. Without loss of generality, we can assume that for every $n \geq 0$, γ_n is parametrized proportionally to arclength and that its domain is the interval $[0, 1]$. By Proposition 1.4.11, there exists a subsequence of $(\gamma_n)_{n\geq 0}$ that converges uniformly to a path γ. By taking limits, it is clear that the path γ joins x and y. By Corollary 1.4.5, we have $L(\gamma) \leq \liminf_{n\to\infty} L(\gamma_n) = \alpha$. From the definition of α, we also have $L(\gamma) \geq \alpha$, which shows that $L(\gamma) = \alpha$. □

The existence of a path of minimal length joining two given points in a metric space is interesting information and it leads to non-trivial properties. In particular, such a path is necessarily injective. An injective path is usually called a *Jordan path*. We shall see in Chapter 2 that Proposition 1.4.12 implies directly the theorem of Hopf–Rinow on the existence of geodesic paths joining two arbitrary points in a proper length space (Theorem 2.4.6).

The following result which has the same flavour as Proposition 1.4.12 will be used later on in the proof of the existence of local geodesics in homotopy classes of paths with fixed endpoints.

Proposition 1.4.13. *let X be a proper metric space, let x and y be two points in X and let γ be a rectifiable path joining x and y. Then, there exists a path that is homotopic* *to γ with endpoints fixed, whose length is equal to the infimum of the lengths of paths in this homotopy class.*

Proof. We set $\alpha = \inf\{L(\gamma')\}$ where γ' varies over the paths joining x and y that are homotopic with fixed endpoints to γ. The proof of the proposition then follows the same outline as that of Proposition 1.4.12. Clearly, the limit of a sequence of paths joining x and y that are in the same homotopy class with endpoints fixed than γ is itself in that class. □

Notes on Chapter 1

Computing lengths of curves in Greek antiquity. The idea of computing the length of a curve in the plane by approximating it by polygonal lines can be traced back to Greek antiquity. In the third century B.C., Archimedes used this method to obtain good approximations of the perimeter of a circle and of the length of a spiral. In his paper *On the sphere and cylinder* (see [67]), Archimedes proves that the perimeter of a circle is comprised between the perimeters of the inscribed and the circumscribed polygons. In his paper *On the measure of the circle* (cf. [67] p. 98), he shows that the ratio between the perimeter of a circle and its diameter is comprised between $3 + \frac{10}{71}$ and $3 + \frac{1}{7}$. These bounds are obtained by calculating the perimeter of an inscribed regular polygon having 96 sides and that of a circumscribed regular polygon having the same number of sides.

Concerning the notion of length in Greek antiquity, the mathematician Giuseppe Peano writes in his paper [121]: "The Greek geometers, for what concerns the lengths of lines and the areas of surfaces (spheres, cylinders and so on), started from *postulates* instead of *definitions*. But the difference is only formal. The postulates that were stated by Archimedes in *On the sphere and the cylinder* are equivalent to the following definitions:

1) the length of a curvilinear plane convex arc is the common value of the least upper bound of the length of the polygonal inscribed arcs and the greatest lower bound of the circumscribed ones;

2) the area of a convex surface is the common value of the least upper bound of the areas of the inscribed polyhedral convex surfaces, and the greatest lower bound of the areas of the circumscribed ones;

3) the length of a curvilinear arc is the least upper bound of the lengths of the polygonal inscribed arcs."

Paths and their lengths after the 17th century. The ideas of Archimedes were taken up by 17th century geometers such as R. Descartes, E. Torricelli, C. Huygens and J. Wallis. Proposition 1.3.1 that says that in Euclidean space, the "line element" ds is equal to $\sqrt{dx_1^2 + \cdots + dx_n^2}$, is sometimes attributed to Wallis, because an analogous formula is contained in his *Arithmetica Infinitorum* (see [142]).

We note that the parametrized curves that were considered by the 17th century mathematicians were not arbitrary paths, but curves defined by algebraic equations. One of the earliest definitions of a curve as an arbitrary continuous map from an interval of $\mathbb{R}$ to Euclidean space $\mathbb{E}^2$ or $\mathbb{E}^3$ is contained in the *Cours d'Analyse* of Camille Jordan, written in 1882 (see [79]). In this treatise, Jordan takes, as a definition of the length of a curve, the limit of the perimeters of sequences of inscribed polygons. He uses the term *rectifiable* to denote a curve whose length is finite and he establishes a formula for the length of a differentiable curve that is analogous to the one given by Proposition 1.3.1 above.

Besides Jordan, there are several mathematicians who contributed to the notion of paths and their lengths. For instance, Peano, in [119], has a section on the definition of the length of a curve (p. 161) and on the area of a surface in Euclidean space (p. 164). In his paper [120], Peano gives an example of a path whose image is equal to the unit square in $\mathbb{E}^2$, showing that the notion of "curve" in the sense of Jordan could be contrary to the intuition that considers a curve as a one-dimensional object. Peano constructs such a path (which is usually referred to as the *Peano curve*), as a limit of paths for which he gives explicit formulas that involve the triadic development of an arbitrary number in [0, 1].

David Hilbert, in [68], gave an example of a Peano curve which is defined in a geometric manner. In fact, this paper by Hilbert contains the pictures of Figure 1.4. The limiting curve in this example is usually called "Hilbert's curve" or the "Hilbert-Peano curve". In his paper, Hilbert uses a step-by-step definition where at step n he cuts the initial square into 2^{2n} boxes, and he describes through a picture the image of the curve in these boxes. Hilbert's pictures replace Peano's formulas involving the triadic development of real numbers.

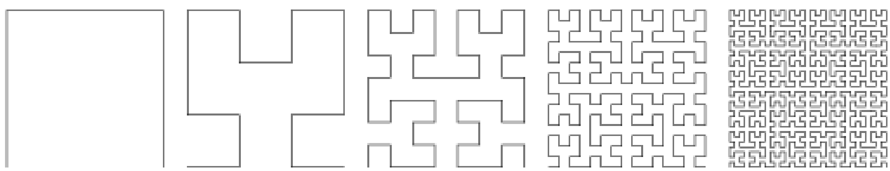


Figure 1.4. The Hilbert curve.

Felix Klein, in his 1898 paper [89], notes that there are various existing definitions of a curve, none of them being completely satisfactory, and he poses the problem of making a coherent definition.

Metric spaces. The first axiomatic definition of a metric space is due to Maurice Fréchet. In fact, until the 1930s, it was not unusual to write expressions like "metric space in the set of Fréchet" instead of "metric space". The definition of a metric space is contained in Fréchet's paper [49], written in 1906, which constitutes his doctoral dissertation. In this paper, the distance between two points is called "écart". In the same paper, Fréchet also defines the notion of complete metric space.[3]

In his 1925 paper [102], Karl Menger introduces the idea of a path in a general metric space, generalizing the notion of curve introduced by Jordan. In a series of papers on the subject (see [102], [103], [106] and [107]), Menger establishes the fundamental properties of lengths of paths. He proves for instance that the length of a curve is the limit of lengths of distances of points associated to a sequence of subdivisions whose moduli tend to zero (Proposition 1.1.10 above). He establishes the lower semi-continuity of the length function (Theorem 1.4.4) and other related basic results. One of the earliest instances of a generalization of an important theorem in Riemannian geometry to the context of metric spaces is due to Cohn-Vossen. In [41], Cohn-Vossen generalizes the theorem of Hopf and Rinow that we shall see in Chapter 2.

Finally, we note that the theory of lengths of paths in metric spaces allows one to define and use, in an efficient way, notions like angle, curvature, tangent space and other related objects in a setting that is much more general than that of Riemannian geometry where these notions are usually defined. As we already mentioned in the introduction to this book, for these generalizations, one has to mention especially the work done by A. D. Alexandrov and his school.

[3] In fact, several basic notions in topology are due to Fréchet. In an obituary presented to the French Academy of Sciences [97], S. Mandelbrojt writes the following: "One must note that in his book written in 1914, Hausdorff, while exposing some properties of Fréchet abstract spaces, introduces a terminology, and this fact made the attribution of the corresponding notions to Hausdorff. For instance: compact sets, separable sets, semi-compact sets, metric spaces. But these notions are due to Fréchet." In the same obituary, Mandelbrojt cites the following remark that Hadamard made in a report to the Academy in 1934: "It seems to us that the daring that is displayed, the abstraction effort accomplished by M. Fréchet, are unprecedented in all that was published since the work by Galois".

Chapter 2

Length spaces and geodesic spaces

Introduction

A *length space* is a metric space in which the distance between any pair of points is equal to the greatest lower bound of the set of lengths of paths joining them. If the distance between any pair of points is equal to the length of some path joining them, then the space is said to be a *geodesic space*. In this chapter, we review some basic properties of length spaces and of geodesic spaces and we examine several examples in some detail. Some of these examples (namely, the Carathéodory, the Kobayashi, and the Teichmüller metrics) belong to the realm of complex analysis, and it is an interesting fact that some of the important ideas of complex analysis can be formulated in terms of metric spaces and maps between them. We shall again encounter these examples several times in later chapters.

A *geodesic path* in a metric space is a path whose length is equal to the distance between its endpoints. Thus, a geodesic space is a space in which two arbitrary points can be joined by a geodesic path. Formulated this way, the property for a metric space of being geodesic is reminiscent of the property for a subset of $\mathbb{R}^n$ of being convex (any two points can be joined by a Euclidean segment contained in the subset). In fact, the notion of geodesic space is one among several notions that we shall study in these notes and that play the role of convexity in general metric spaces.

We also introduce in this chapter the notion of *betweenness* in a metric space. The definition is as follows. Let X be a metric space and let x and y be two distinct points in X. A point z in X is said to *lie between x and y* if z is distinct from x and from y and if the following equality (which is a degenerate triangle inequality) is satisfied:

$$|x - z| + |z - y| = |x - y|.$$

We already mentioned that the notion of geodesic in an arbitrary metric space is due to Menger. Betweenness has also been introduced by Menger, and it leads to a new notion of convexity: a metric space X is said to be *Menger convex* if for every pair of distinct points in that space, there exists a point that lies between them. The existence, for an arbitrary pair of distinct points in X, of a point that lies between them is a discrete analog of the existence of a geodesic path joining two arbitrary points in that space. Thus, Menger's theory of betweenness allows to build a theory of discrete geodesics in metric spaces. We prove (Theorem 2.6.2) that if X is a proper metric space, then it is Menger convex if and only if it is geodesic.

The outline of this chapter is as follows.

In Section 1, we study a few general properties of length spaces. If (X, d) is a metric space that is "connected by rectifiable paths", then there is a natural metric d_ℓ on X which is called the "length metric of X" and which has the property that the new metric space (X, d_ℓ) is a length space. In the case where (X, d) is already a length space, then the associated length metric d_ℓ coincides with the original metric d.

We prove that a complete locally compact length space is proper. This result, which is attributed to H. Hopf and W. Rinow, gives a characterization of compact subsets of complete locally compact length spaces that is analogous to the characterization of compact subsets of $\mathbb{R}^n$ that is given by the Bolzano–Weierstrass Theorem, which says that the compact subsets of $\mathbb{R}^n$ are the closed and bounded subsets.

In Section 2, we introduce geodesics in metric spaces and we study some of their basic properties. We make the relation with Menger's notion of Betweenness.

Section 3 contains a few results on limits of geodesic paths that we shall use in later chapters.

Section 4 covers geodesic spaces, that is, metric spaces where the distance between two points is equal to the length of a geodesic path joining them. A geodesic space is a length space and we shall see the following converse, which is also due to Hopf and Rinow: every complete locally compact length space is a geodesic space. We study a particularly interesting class of geodesic spaces, which is the class of uniquely geodesic metric spaces. For each pair of points in such a space, there is (up to reparametrization) a unique geodesic path joining them. Busemann metric spaces, which we shall consider more thoroughly in Chapter 8, are uniquely geodesic.

There is a theory of geodesic convexity that can be naturally developed in the context of uniquely geodesic spaces, and we consider in Section 5 a few basic elements of this theory.

In Section 6, we introduce the notion of Menger convexity and we establish several necessary and sufficient conditions for a metric space to be Menger convex. We use this notion to study geodesics in products of metric spaces.

2.1 Length spaces

Definition 2.1.1 (Space connected by rectifiable paths). We say that a metric space X is *connected by rectifiable paths* if for every x and y in X, there exists a rectifiable path $\gamma : [a, b] \to X$ such that $\gamma(a) = x$ and $\gamma(b) = y$.

Definition 2.1.2 (Length space). We say that a metric space X is a *length space* if for every x and y in X, we have

$$|x - y| = \inf_{\gamma} L(\gamma),$$

where the infimum is taken over the set of paths γ that join x and y. The metric of a length space is called a *length metric*.

In particular, a length space is connected by rectifiable paths.

We give now a list of examples of familiar length spaces (or of length metrics on familiar spaces). We also include examples of length *pseudo-metrics*, namely the Carathéodory, the Kobayashi and the Thurston metrics, that are not genuine metrics. We recall that a *pseudo-metric* on a space X is a map $d \colon X \times X \to \mathbb{R}$ satisfying, for all x, y and z in X, $d(x,x) = 0$, $d(x,y) \geq 0$, $d(x,y) = d(y,x)$ and $d(x,y) \leq d(x,y) + d(y,z)$. The three pseudo-metrics mentioned above do not satisfy the axiom $d(x,y) = 0 \Rightarrow x = y$, but they are important examples that arise naturally in the context of complex manifolds, and their definitions and basic properties are much in the same spirit as those of length metrics.

Examples 2.1.3 (Length spaces).

(i) *Euclidean spaces.* For all $n \geq 1$, let $\mathbb{E}^n$ be the n-dimensional Euclidean space, that is, the space $\mathbb{R}^n$ equipped with the Euclidean metric. From classical Euclidean geometry, we know that $\mathbb{E}^n$ is a length space: the distance between two points is equal to the length of the affine path joining them. Likewise, a convex subset of $\mathbb{E}^n$ (that is, a subset for which the affine segment joining any two points in this subset is contained in the subset), equipped with the metric induced from that of $\mathbb{E}^n$, is a length space. For every $n \geq 2$, the space $\mathbb{E}^n$ with a finite number of points removed is also a length space. On the contrary, the space $\mathbb{E}^n$ with an open ball B of positive radius r removed is not a length space. To prove the last assertion, let x and y be two opposite points on the boundary of B. We can use the fact (which we prove in Chapter 5 below) that the projection map on the closed ball $\overline{B}$ is 1-Lipschitz to show that the length of any path γ joining x and y in $\mathbb{E}^n \setminus B$ is bounded below by the length of a path joining x and y and whose image is contained in the sphere $S = \partial B$. Since the length of such a path is bounded below by πr, we have, for any path $\gamma \colon [a,b] \to \mathbb{E}^n \setminus B$ joining x and y, $L(\gamma) \geq \pi r$, whereas $|x - y| = 2r$. This shows that $\mathbb{E}^n \setminus B$ is not a length space. The same result holds for $\mathbb{E}^n$ with a closed ball removed.

(ii) *Normed vector spaces.* Normed vector spaces constitute a vast class of examples of length spaces. The length of an affine path in such a space is equal to the distance between its endpoints. More generally, any affinely convex subset in a normed vector space equipped with the induced metric is a length space.

(iii) *Spheres.* For $n \geq 3$, let $S = S^{n-1}$ be the unit sphere in Euclidean space $\mathbb{E}^n$ and let d be the metric induced on S by the metric of $\mathbb{E}^n$. Let us show that (S,d) is not a length space. From elementary geometry, we know that for every x and y in S, we have $d(x,y) = 2\sin(\alpha/2)$, where α is the angle, with value in the interval $[0,\pi]$, formed by the two rays issuing from the origin and passing through x and y. On the other hand, the length $L(\gamma)$ of any path $\gamma \colon [a,b] \to S$ joining x and y is bounded below by the length of the smallest arc of a great circle in S (that is, a Euclidean circle of maximal diameter that is contained in the sphere) joining x and y, that is, by α. Thus, we have $L(\gamma) \geq \alpha > 2\sin(\alpha/2)$, which implies $\inf_\gamma L(\gamma) > d(x,y)$ for all x and y satisfying $d(x,y) \neq 0$. This shows that (S,d) is not a length space. On the contrary, the *intrinsic metric* of S is a metric for which the distance between x and y is equal to α, and it is a length metric.

(iv) *Length metrics on simplicial graphs.* We start by recalling the definition of an abstract simplicial graph. This is an object G consisting of a pair (S, A) where S is a set (which could be finite or infinite) whose elements are called the *vertices* of G, and A a set whose elements are called the *edges* of G, where an element of A consists of an unordered (not necessarily distinct) pair of points in S which are called the *vertices* of the given edge. Thus, each vertex of G possesses two vertices (which could be identical). We equip an abstract graph with a pseudo-metric. For each edge a of G, we consider a nonempty closed interval I_a of $\mathbb{R}$ and a surjective map from the set of endpoints of I_a to the set of vertices of a. We say that I_a is a *geometric realization* of the edge a. We then define a topological space X by taking the quotient topology on the disjoint union $\bigcup_{a \in A} I_a$ by the equivalence relation that identifies two vertices if and only if these vertices are identical as elements of S. In all that follows, we suppose that the topological space X that we obtain in this way is connected (or, equivalently, arcwise connected).

To define the pseudo-metric on X, we use the notion of *affine path* in X. This is a path whose restriction to each image of an interval I_a in X, equipped with the affine structure that is induced from that of the interval I_a, is affine. Now we can define a map d on $X \times X$, by setting $d(x, y) = \inf_\gamma L(\gamma)$ for all x and y in X, where the infimum is taken over the set of affine paths joining x and y and where the length of such a path γ is defined as the sum of the lengths of the restrictions of γ to the various edges that it traverses. In each edge a of X, the length is computed with respect to the Euclidean metric induced from that of the interval I_a that is its geometric realization. We can easily check that d is a pseudo-metric. In order for this pseudo-metric to be a metric (that is, in order that it satisfies the axiom $d(x, y) = 0 \Rightarrow x = y$), it suffices that the set of lengths of each intervals I_a, for a in A, be bounded below by a positive constant that does not depend on a. With this condition, it is easy to check that the space X equipped with this metric is a length space, which is called a *metric simplicial graph*.

(v) *Length metrics on Cayley graphs.* Let Γ be a finitely generated group and let S be a finite generating set for Γ. The associated *Cayley graph* $C(\Gamma, S)$ is the simplicial graph whose vertices are the elements of Γ and whose edges are the pairs of vertices (a, b) such that there exists an element s in S satisfying $b = sa$. The space $C(\Gamma, S)$ is equipped with the length metric in which the length of each edge is equal to one. The induced metric on Γ is called a *word metric*. The group Γ acts isometrically by left translations on the Cayley graph $C(\Gamma, S)$, and this action of Γ is a fundamental object that is used in the geometrical study of Γ.

(vi) *Riemannian metric.* Let M be a differentiable manifold. (For our purpose, it suffices to suppose that M is of class $\mathcal{C}^1$, although in order to use efficiently the language and the techniques of Riemannian geometry, one has to suppose that M is at least of class $\mathcal{C}^2$.) We recall that a *Riemannian metric* on M is a map that assigns to each point x in M an inner product g_x on the tangent vector space $T_x M$ such that for each open subset U of M and for each tangent vector fields X and Y on U, the map $g\colon U \to \mathbb{R}$ defined by $x \mapsto g_x(X(x), Y(x))$ is of class $\mathcal{C}^1$. For every $\mathcal{C}^1$ path $\gamma\colon [a,b] \to M$, its "length" $L_g^*(\gamma)$ is defined as

$$L_g^*(\gamma) = \int_a^b \|\gamma'(t)\| \, dt,$$

where, for each t in $[a,b]$, we have

$$\|\gamma'(t)\| = g_{\gamma(t)}(\gamma'(t), \gamma'(t))^{1/2}.$$

Then, the metric d of M is defined by setting the distance $d(x,y)$ between any two points x and y in M to be equal to the infimum of the "lengths" of $\mathcal{C}^1$ paths in M joining them. The Riemannian manifold M, equipped with this metric, is a length metric space. (This is easy to see but it does not follow trivially from the definition; one has to prove that the length $L_d(\gamma)$ of a $\mathcal{C}^1$ path γ is indeed equal to $L_g^*(\gamma)$.)

(vii) *Hyperbolic space.* For each $n \geq 2$, we can define the n-dimensional hyperbolic space $\mathbb{H}^n$ by equipping the open unit ball B^n of $\mathbb{R}^n$,

$$B^n = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : \sum_{i=0}^{n} x_i^2 < 1\},$$

with the metric whose "line element" ds is given by

$$ds = 2\frac{\sqrt{dx_1^2 + \cdots + dx_n^2}}{1 - \|x\|^2},$$

where $\|x\|$ denotes the Euclidean norm of the element x. This means that the length of any piecewise $\mathcal{C}^1$-path $\gamma\colon [a,b] \to B^n$ is equal to

$$L(\gamma) = \int_a^b ds = \int_a^b 2\,\frac{\sqrt{\gamma_1'^2(t) + \cdots + \gamma_n'^2(t)}}{1 - (\gamma_1(t)^2 + \cdots + \gamma_n(t)^2)}\, dt,$$

where for each $i = 1, \ldots, n$, $\gamma_i(t)$ is the i-th coordinate of $\gamma(t)$. (This is similar to the formula that allows one to compute the length of a $\mathcal{C}^1$-path in Euclidean space; cf. Proposition 1.3.1.) The distance between two points in B^n is set to be the infimum of the lengths of $\mathcal{C}^1$-paths joining them. This defines a length metric on B^n, because the sum of the distances appearing in the formula giving the total variation of a path γ with respect to a subdivision can always be approximated by the length of a $\mathcal{C}^1$-path joining the endpoints of γ. The ball B^n, equipped with this metric, is a model of the n-dimensional hyperbolic space $\mathbb{H}^n$ that is called the *conformal* (or *Poincaré*) *ball model* . This metric on B^n is complete and it is usually referred to as the *Poincaré* *metric* of the ball. There are several nice formulas for the hyperbolic distance $d_{B^n}(x, y)$ between two points x and y in B^n. For instance, we have

$$\cosh d_{B^n}(x, y) = 1 + \frac{2\|x - y\|^2}{(1 - \|x\|^2)(1 - \|y\|^2)}.$$

The case of dimension 2 is particularly interesting and is a central object in complex analysis. To see some of its properties, consider B^2 as the subset of the complex plane defined by

$$B^2 = \{\zeta \in \mathbb{C}, |\zeta| = 1\}.$$

The following formula gives the Poincaré distance in terms of moduli of complex numbers: given two complex numbers ζ and ζ' in B^2, we have

$$d_{B^2}(\zeta, \zeta') = \frac{1}{2} \ln \frac{1 + \left|\frac{\zeta' - \zeta}{1 - \zeta'\overline{\zeta}}\right|}{1 - \left|\frac{\zeta' - \zeta}{1 - \zeta'\overline{\zeta}}\right|}.$$

The Poincaré model of hyperbolic geometry appeared for the first time in [122]. For a concise exposition of hyperbolic geometry, leading directly to modern developments, we refer the reader to W. P. Thurston's mimeographed notes [134], his book [137], and the book [14] by R. Benedetti and C. Petronio. See also [11] and [124]

Let us also note that there is a formula for the Poincaré metric of the ball that makes use of the cross-ratio of four points; we shall return to this matter in Chapter 5.

There is an explicit parametrization of geodesic segment joining two points in B^2, similar in some sense to the parametrization of geodesic segments joining points in Euclidean space: for ζ and ζ' in B^2, the map $\gamma : [0, 1] \to B^2$ defined by

$$\gamma(t) = \frac{t\frac{\zeta' - \zeta}{1 - \zeta'\overline{\zeta}} + \zeta}{1 + t\overline{\zeta}\frac{\zeta' - \zeta}{1 - \zeta'\overline{\zeta}}}$$

is an affinely reparametrized geodesic joining them.

For the record, we also mention the *upper half-space* model of $\mathbb{H}^n$. This is the subset

$$H^n = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : x_n > 0\}$$

of $\mathbb{R}^n$ equipped with the metric for which the line element ds is given by

$$ds = \frac{\sqrt{dx_1^2 + \cdots + dx_n^2}}{x_n}.$$

In this model, the hyperbolic distance $d_{H^n}(x, y)$ between two points $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ satisfies the relation

$$\cosh d_{H^n}(x, y) = 1 + \frac{\|x - y\|^2}{2x_n y_n}.$$

Again, in the case $n = 2$, there is an explicit formula that gives the distance between two points in H^2 considered as a subset of the complex plane,

$$H^2 = \{\zeta \in \mathbb{C} \mid \mathrm{Im}(\zeta) > 0\},$$

which is most useful in complex analysis. For ζ and ζ' in H^2, we have

$$d_{H^2}(\zeta, \zeta') = \frac{1}{2} \ln \left(\frac{|\zeta - \overline{\zeta'}| + |\zeta - \zeta'|}{|\zeta - \overline{\zeta'}| - |\zeta - \zeta'|} \right).$$

(viii) *Finsler metric.* A Finsler metric on a differentiable manifold M is a map that assigns to each point x in M a norm on the tangent space $T_x M$. Such a norm does not necessarily come from an inner product as in the case of a Riemannian manifold. The "length" of a piecewise $\mathcal{C}^1$-path $\gamma\colon [a, b] \to M$ is defined as $\int_a^b \|\gamma'(t)\|\, dt$. From that, one defines a length metric on M, in the same way as we did in Example (vi) above, for the case of a Riemannian manifold. Thus, Finsler manifolds generalize Riemannian manifolds. Considering the tangent vectors as "infinitesimal elements", one can consider that at the infinitesimal level, a Riemannian metric is a Euclidean vector space whereas a Finsler metric is a general finite-dimensional normed vector space.

Finsler spaces are named after Paul Finsler who introduced them in 1918 (see [47]).

(ix) *The Carathéodory pseudo-distance for complex manifolds.* Although the Carathéodory pseudo-distance, which we recall now, is not always a distance, we mention it here because of the importance of the idea that lies behind its definition, an idea which is close to the one that is at the basis of the definition of a length space. (In fact, as we shall see, the two ideas are in some sense dual to each other.) Furthermore, the definition of the Carathéodory pseudo-distance is at the origin of the Kobayashi pseudo-distance, which is of paramount importance in the study of complex manifolds and of maps between them. In the general case, the Carathéodory pseudo-distance is a length pseudo-distance, and in some cases it is a length metric. The idea behind this definition of the Carathéodory pseudo-distance is at the heart of many constructions in the theory of length spaces like, for instance, that of the length metric defined in Proposition 2.1.5 below and the discussion that follows it. To define the Carathéodory pseudo-distance, we let $D = B^2$ be the 2-dimensional disk equipped with its Poincaré metric d_D. For any complex manifold X and for any x and y in X, we set

$$d_{\mathcal{C}}(x, y) = \sup_f d_D\big(f(x), f(y)\big),$$

where the supremum is taken over the set of holomorphic maps $f\colon X \to D$. This clearly defines a pseudo-metric on X. A classical argument that uses Ascoli's theorem shows that the supremum is achieved by a holomorphic map.

In the case where X is the unit disk D, using the Schwarz–Pick lemma, it is easy to see that the Carathéodory pseudo-distance of D coincides with the Poincaré metric of that disk (see [91] p. 49). On the other hand, is X is the n-dimensional complex plane $\mathbb{C}^n$, the Carathéodory pseudo-distance vanishes identically. (This is a consequence of Liouville's theorem which says that a bounded holomorphic function on $\mathbb{C}$ is constant.)

Let us also note that the Carathéodory pseudo-distance can be defined at the infinitesimal level, and in fact, it is a *pseudo-Finsler* metric, that is, lengths of paths are obtained by integrating *pseudo-norms* of tangent vectors along these paths. More precisely, if v is a tangent vector at a point x in X, then the pseudo-norm of v is defined as

$$\|v\| = \sup_f \{df_x(v)\}$$

where the supremum is again taken again over the set of holomorphic maps $f : X \to D$.

The Carathéodory pseudo-distance was introduced by Constantin Carathéodory in [33] and [34]. Carathéodory also gave the infinitesimal version of this pseudo-distance.

(x) *The Kobayashi pseudo-distance for complex manifolds.* Let X be again a complex manifold. The Kobayashi pseudo-distance on X is some sort of "dual" to the Carathéodory pseudo-distance (the word "dual" is used by Kabayashi himself) where instead of considering holomorphic maps $f : X \to D$, one considers holomorphic maps $f : D \to X$. Before defining the Kobayashi pseudo-distance $d_{\mathcal{K}}$, one defines, for all x and y in X, the map $d'_{\mathcal{K}} : X \times X \to \mathbb{R}$ by

$$d'_{\mathcal{K}}(x, y) = \inf_f d_D(a, b),$$

where the infimum is taken over all holomorphic maps $f : D \to X$ and over all a and b in D satisfying $f(a) = x$ and $f(b) = y$. The map $d'_{\mathcal{K}}$ is not a pseudo-distance because it does not satisfy the triangle inequality. The Kobayashi pseudo-metric $d_{\mathcal{K}}$ is then defined as the *largest* pseudo-distance satisfying $d_{\mathcal{K}} \leq d'_{\mathcal{K}}$.

There are some very nice equivalent definitions of the Kobayashi pseudo-distance, and we now recall some of them, because they follow the scheme of some general definitions in the theory of length spaces.

One of these definitions uses the notion of analytic chain in X. If x and y are two arbitrary points in X, then, an *analytic chain joining x and y* is a sequence $f_1, \ldots, f_n$ of holomorphic maps $f_i : D \to X$, together with two sequences of points $\zeta_1, \ldots, \zeta_n$ and $\zeta'_1, \ldots, \zeta'_n$ in D satisfying $f_1(\zeta_1) = x$, $f_n(\zeta'_n) = y$ and for all $j = 1, \ldots, n-1$, $f_j(\zeta'_j) = f_{j+1}(\zeta_{j+1})$. The Kobayashi pseudo-distance is then equal to

$$d_{\mathcal{K}}(x, y) = \inf\{d_D(\zeta_1\zeta'_1) + \cdots + d_D(\zeta_n\zeta'_n)\}$$

where the infimum is taken over all analytic chains joining x and y.

The Kobayashi pseudo-distance is also characterized by the fact that it is the largest pseudo-distance on X such that any holomorphic map $f : D \to X$ is non-expanding (that is, f satisfies $|f(x) - f(y)| \leq |x - y|$).

There is also an infinitesimal version of the Kobayashi pseudo-distance, defined as follows: for any x in X and for any tangent vector v at x, the *pseudo-norm* of v is defined as

$$\|v\| = \inf_{v'}\{\|v'\|_D\}$$

where v' is a tangent vector to the disk D, $\|v'\|$ is the norm of v' with respect to the metric of D, and where the infimum is taken over all vectors v' in D such that there is a holomorphic maps $f: D \to X$ whose derivative sends v' to v. The Kobayashi pseudo-distance between two points is equal to the infimum of this pseudo-norm integrated over all $\mathcal{C}^1$ paths joining the two points, as in the case of a Finsler metric.

The first examples of Kobayashi pseudo-distances are those of the complex plane $\mathbb{C}$, of the punctured complex plane $\mathbb{C} \setminus \{0\}$ and of the unit disk D. For the case of the complex plane $\mathbb{C}$, using the fact that homotheties of this plane are holomorphic, we can find, for any x and y in $\mathbb{C}$ and for any $\epsilon > 0$, a holomorphic map $f: D \to \mathbb{C}$ satisfying $f(0) = x$ and $f(\epsilon) = y$, which implies $d_{\mathbb{C}}(x, y) \leq \epsilon$ for any $\epsilon \geq 0$, and therefore $d_{\mathbb{C}}(x, y) = 0$. For the case of the punctured complex plane $\mathbb{C} \setminus \{0\}$, one uses the existence of a surjective holomorphic map $\mathbb{C} \to \mathbb{C} \setminus \{0\}$, for instance, the exponential map $z \mapsto e^z$, and the fact that holomorphic maps are distance-non-expanding for the Kobayashi pseudo-metric. Since the Kobayashi-pseudo-metric of $\mathbb{C}$ is identically zero, the Kobayashi pseudo-metric of $\mathbb{C} \setminus \{0\}$ is also identically zero. The Kobayashi pseudo-metric of $\mathbb{C} \setminus \{0, 1\}$ is a metric. It is also easy to see, using the Schwarz–Pick lemma (which is one way of saying that holomorphic mappings are distance non-expanding), that the Kobayashi pseudo-metric of the disk D coincides with its Poincaré metric. Likewise, the Kobayashi pseudo-metric associated to a complex structure on a (real) surface coincides with its hyperbolic metric. From this point of view, one can consider the Kobayashi pseudo-metric associated to a complex manifold as a generalization of the hyperbolic (or Poincaré) metric associated to a 1-dimensional complex manifold.

The Kobayashi pseudo-distance was introduced by Shishichi Kobayashi in 1967 (see [90]). The infinitesimal description of this pseudo-distance was given by H. L. Royden in [126].

Let us finally note that the Kobayashi pseudo-distance is always bounded below by the Carathéodory pseudo-distance.

(xi) *Thurston's pseudo-metric on complex projective surfaces.* We describe here Thurston's construction of a pseudo-metric for complex projective surfaces, because it is in the same spirit as that of the Carathéodory and the Kobayashi pseudo-metrics. We recall that a complex projective structure on a surface is a maximal atlas of charts with values in the complex projective plane $\mathbb{C}P^1$ and whose coordinate changes are restrictions of projective transformations of $\mathbb{C}P^1$, that is to say, elements of $\mathrm{PSL}(2, \mathbb{C})$. At the infinitesimal level, this metric is defined as follows: for any $\mathbb{C}P^1$-manifold X, for any point x in X and for any tangent vector v at x, the pseudo-norm of v is defined as

$$\|v\| = \inf \|w\|_D,$$

where the infimum is taken over all vectors v in the Poincaré disk D such that there exists a projective map $f: D \to X$ satisfying $df(w) = v$. From the definitions, it is easy to see that Thurston's pseudo-metric is always bounded below by the Kobayashi pseudo-metric. In the 1970s, Thurston introduced new ideas and techniques, including the above definition, in the study of projective structures, but he did not publish anything on that subject. For interesting properties and applications of Thurston's pseudo-metric, we refer the reader to the paper [130] by H. Tanigawa.

(xii) *Gluing.* Finally, let us give an example of a length space that is obtained by gluing two metric spaces of apparently different nature. Consider a (connected or non-connected) hyperbolic surface S with geodesic boundary, this boundary consisting of two closed geodesics having the same length L, and consider a compact Euclidean cylinder whose boundary curves have equal length L. Connecting the boundary curves of C *via* the cylinder C by an isometric map (see Figure 2.1), we obtain a surface which

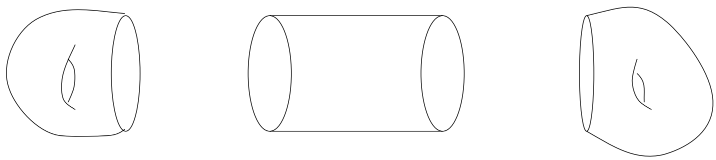


Figure 2.1. The cylinder is Euclidean and the two other surfaces are hyperbolic.

we shall call X. We can equip X with a length metric by defining the distance between two arbitrary points as the infimum of the lengths of paths joining them. Here, the length of a path γ is defined to be the sum of the lengths of subpaths $\gamma_1, \dots, \gamma_n$, where γ is decomposed as a concatenation $\gamma_1 * \dots * \gamma_n$, with the image of each γ_i lying either in the hyperbolic or in the Euclidean part of X, and where the length of γ_i is computed with respect to the original metric of that component. (For this definition, we can restrict to paths γ that can be written as finite concatenations $\gamma = \gamma_1 * \dots * \gamma_n$, with each γ_i, as above, having its image in the Euclidean part or in the hyperbolic part.) Clearly, the resulting metric on M is a length metric, and it is not a Riemannian metric. Such structures on surfaces appear in Thurston's work on complex projective structures (see for instance the papers [130] and [101] by Tanigawa and by McMullen). In that theory, the operation of inserting the Euclidean cylinder into the hyperbolic surface is called "grafting". The projective structure obtained by this operation is used in an essential way, but so far in this theory the underlying length structure does not play any important role.

Let us now return to general metric spaces. If (X, d) is a metric space, we define a map d_ℓ from $X \times X$ to the extended real line $\mathbb{R} \cup \{\infty\}$ by setting

$$d_\ell(x, y) = \inf_{\gamma} L(\gamma),$$

where the infimum is taken over the set of paths γ joining x and y.

The following lemma will be useful below.

Lemma 2.1.4. *Let X be a length space, let x and y be two points in X and let α and β be two nonnegative real numbers satisfying $\alpha + \beta \geq |x - y|$. Then for any $\epsilon > 0$, we can find a point z in X satisfying $|x - z| \leq \alpha$ and $|z - y| \leq \beta + \epsilon$.*

Proof. Let $\gamma: [a, b] \to X$ be a path joining x and y and satisfying $L(\gamma) \leq |x-y|+\epsilon$. Without loss of generality, we can suppose that $\alpha \leq L(\gamma)$. By Corollary 1.2.6, we can assume that γ is parametrized by arclength. If we set $z = \gamma(a + \alpha)$, then

$$|x - z| \leq L(\gamma_{|[a,a+\alpha]}) = \alpha$$

and

$$|z - y| \leq L(\gamma_{|[a+\alpha,b]}) = L(\gamma) - \alpha \leq |x - y| + \epsilon - \alpha \leq \beta + \epsilon. \qquad \square$$

Proposition 2.1.5. *Let (X, d) be a metric space that is connected by rectifiable paths. Then d_ℓ is a metric on X and we have $|x - y|_d \leq d_\ell(x, y)$ for every x and y in X.*

Proof. Let x and y be two arbitrary points in X and let $\gamma: [a, b] \to X$ be a path joining them. By inequality (1.1.1.1) of Chapter 1, we have $|x - y|_d \leq L(\gamma)$, which, by taking the infimum over all paths γ joining x and y, implies $|x - y| \leq d_\ell(x, y)$.

Now let us prove that $d_\ell(x, y)$ is a metric. To every path $\gamma: [a, b] \to X$, we associate a path $\gamma': [a, b] \to X$ defined, for t in $[a, b]$, by $\gamma'(t) = \gamma(a+b-t)$. It is clear that γ' joins y and x and that $L(\gamma) = L(\gamma')$. Thus, we have $d_\ell(y, x) \leq d_\ell(x, y)$ for all x and y, which, by symmetry, implies $d_\ell(x, y) = d_\ell(y, x)$. Now suppose that $d_\ell(x, y) = 0$. From the inequality $|x - y| \leq d_\ell(x, y)$ we obtain $|x - y| = 0$, which implies $x = y$. To prove that d_ℓ is a metric, it remains to show that the triangle inequality is satisfied. Let x, y and z be three arbitrary points in X, let $k = d_\ell(x, y)$ and let $k' = d_\ell(y, z)$. For any $\epsilon > 0$, let $\gamma: [a, b] \to X$ be a path joining x and y and satisfying $L(\gamma) \leq d_\ell(x, y) + \epsilon/2$ and let $\gamma': [a, b] \to X$ be a path joining y and z and satisfying $L(\gamma') \leq d_\ell(y, z) + \epsilon/2$. We define the path $\gamma'': [a, b + b' - a'] \to X$ by setting

$$\gamma''(t) = \begin{cases} \gamma(t) & \text{if } t \in [a, b], \\ \gamma'(t - b + a') & \text{if } t \in [b, b + b' - a']. \end{cases}$$

Obviously, we have $L(\gamma''_{|[a,b]}) = L(\gamma)$. On the other hand, the path $\gamma''_{|[b,b+b'-a']}$ is obtained from γ' by the change of parameters $\psi: [b, b+b'-a'] \to [a', b']$ defined by $\psi(t) = t + a' - b$. Therefore, we have $L(\gamma''_{|[b,b+b'-a']}) = L(\gamma')$. This implies

$$L(\gamma'') = L(\gamma) + L(\gamma') \leq d_\ell(x, y) + d_\ell(y, z) + \epsilon.$$

Taking the infimum over the paths γ'' joining x and z, we obtain $d_\ell(x, z) \leq d_\ell(x, y) + d_\ell(y, z) + \epsilon$. Since this holds for every $\epsilon > 0$, we obtain $d_\ell(x, z) \leq d_\ell(x, y) + d_\ell(y, z)$. This proves Proposition 1.5. $\square$

Definition 2.1.6 (The length metric associated to a metric). We call the metric d_ℓ, that is associated by Proposition 2.1.5 to a metric space (X, d) that is connected by rectifiable paths, the *length metric of X associated to d*.

Proposition 2.1.7. *Let (X, d) be a metric space that is connected by rectifiable paths. Then the identity map $(X, d_\ell) \to (X, d)$ is continuous.*

Proof. This follows from the fact that for all x and y in X, we have, by Proposition 2.1.5, $d(x, y) \leq d_\ell(x, y)$. □

Thus, the topology defined by the length metric d_ℓ is finer than the topology defined by the metric d (that is, the open sets for d are also open sets for d_ℓ). In general, the two topologies do not coincide, as we can see from the following examples.

Examples 2.1.8 (The topology induced from a length metric).

(i) Consider the Euclidean plane $\mathbb{E}^2$ and let X be the subset of $\mathbb{E}^2$ defined as

$$X = \big([0, 1] \times \{1\}\big) \cup \big(\{0\} \times [0, 1]\big) \bigcup_{n \geq 1} \big(\{1/n\} \times [0, 1]\big)$$

(see Figure 2.2). We equip X with the metric induced from that of the plane. Consider the sequence of points $(p_n)_{n\geq 1}$ in X where for each $n \geq 1$, $p_n = (1/n, 0)$, and let $p = (0, 0)$. We have $d(p_n, p) \to 0$ when $n \to \infty$, whereas $d_\ell(p_n, p) \geq 2$ for all n. Thus, the identity map $(X, d) \to (X, d_\ell)$ is not continuous.

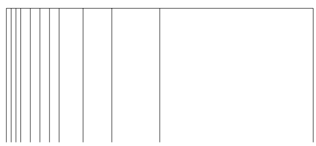


Figure 2.2. A comb (Example 2.1.8 (i)).

(ii) Consider the Koch curve of Example 1.1.11. (More generally, we can take, instead of the Koch curve, any path γ in the Euclidean plane $\mathbb{E}^2$ that is everywhere locally non-rectifiable.) Let $K \subset \mathbb{E}^2$ be the image of that curve and consider the natural embedding of $\mathbb{E}^2$ in 3-dimensional Euclidean space $\mathbb{E}^3$. We choose an arbitrary point p in $\mathbb{E}^3 \setminus \mathbb{E}^2$ and we let $D \subset \mathbb{E}^3$ be the union of the segments $[p, z]$ for all z in K. We equip D with the metric d induced from its inclusion in $\mathbb{E}^3$. This makes D homeomorphic to a closed disk. Since no subpath of the Koch curve is rectifiable, the only rectifiable paths in D are those that can be obtained as a concatenation of paths whose images are contained in segments of the form $[p, z]$ with z in K and where the concatenation takes place at the point p. Therefore, if d_ℓ is the length metric associated to d, then the distance $d_\ell(x, y)$, for any x and y in D, is equal to $d(x, y)$ if x and y

belong to the same segment of the form $[x, z]$ with $z \in K$, and to $d(p, x) + d(p, y)$ if the pair x, y is not contained in such a segment. The space D equipped with the topology defined by the metric d_ℓ is not homeomorphic to a closed disk since for any point x in D that is distinct from p, a small enough open ball centered at x is homeomorphic to an open interval of $\mathbb{R}$ and not to a 2-dimensional disk.

Proposition 2.1.9. *Let (X, d) be a metric space and let $\gamma : [a, b] \to (X, d)$ be a rectifiable path. Then γ is continuous for the metric d_ℓ. (In other words, the map $\gamma : [a, b] \to (X, d_\ell)$ is also a path.)*

Proof. For all t_0 and t in $[a, b]$, we have, by the additivity of length (Proposition 1.1.12),

$$d_\ell\big(\gamma(t_0), \gamma(t)\big) \leq L(\gamma_{|[t_0,t]}) = |L(\gamma_{|[a,t]}) - L(\gamma_{|[a,t_0]})|.$$

Thus, by Proposition 1.1.13 , we obtain $d_\ell\big(\gamma(t_0), \gamma(t)\big) \to 0$ as $t \to t_0$. This proves that γ is continuous with respect to the metric d_ℓ. □

Proposition 2.1.10. *Let (X, d) be a metric space that is connected by rectifiable paths. Then (X, d) is a length space if and only if $d_\ell = d$.*

Proof. By definition, (X, d) is a length space if and only if for all x and y in X, we have $d(x, y) = \inf_\gamma L(\gamma)$, where the infimum is taken over the paths γ joining x and y, that is, if and only if $d(x, y) = d_\ell(x, y)$. □

Proposition 2.1.11. *Let (X, d) be a metric space, let d_ℓ be the associated length metric and let $\gamma : [a, b] \to X$ be a path in (X, d_ℓ). Then γ is also a path in (X, d) and we have $L_d(\gamma) = L_{d_\ell}(\gamma)$.*

Proof. The fact that γ is a path for the metric d follows from the continuity of the identity map $(X, d_\ell) \to (X, d)$ (Proposition 2.1.7). To show the equality of the two lengths, let $\sigma = (t_i)_{i=0,\dots,n}$ be a subdivision of $[a, b]$. The inequality $d \leq d_\ell$ (Proposition 2.1.5) implies $V_\sigma^d(\gamma) \leq V_\sigma^{d_\ell}(\gamma)$. By taking the supremum over subdivisions σ, we obtain $L_d(\gamma) \leq L_{d_\ell}(\gamma)$. On the other hand, we have

$$V_\sigma^{d_\ell}(\gamma) = \sum_{i=0}^{n-1} d_\ell|\gamma(t_i) - \gamma(t_{i+1})| \leq \sum_{i=0}^{n-1} L_d(\gamma_{|[t_i,t_{i+1}]}) = L_d(\gamma).$$

Thus, we obtain $L_{d_\ell}(\gamma) \leq L_d(\gamma)$. We conclude that $L_{d_\ell}(\gamma) = L_d(\gamma)$, which completes the proof of Proposition 2.1.11. □

We deduce the following

Corollary 2.1.12. *Let (X, d) be a metric space that is connected by rectifiable paths. Then (X, d_ℓ) is a length space.*

Proof. For every x and y in X, we have, by definition,

$$d_\ell(x, y) = \inf_\gamma L_d(\gamma),$$

where the infimum is taken over the set of paths γ (for the metric d) joining x and y. Without loss of generality, we can restrict to rectifiables paths. By Proposition 2.1.9, such a path is also a path for d_ℓ.

Now if γ is any rectifiable path for d_ℓ, then, by Proposition 2.1.11, γ is a path for d and $L_{d_\ell}(\gamma) = L_d(\gamma)$. We deduce that

$$d_\ell(x, y) = \inf_\gamma L_{d_\ell}(\gamma),$$

where the infimum is taken over the set of paths γ joining x and y that are continuous and rectifiable for d_ℓ. This proves Corollary 2.1.12. □

Corollary 2.1.13. *Let (X, d) be a metric space that is connected by rectifiable paths. Then the length metric associated to d_ℓ is the metric d_ℓ itself.*

Proof. By Corollary 2.1.12, (X, d_ℓ) is a length space. Therefore, by Proposition 2.1.10, we obtain $(d_\ell)_\ell = d_\ell$.

Definition 2.1.14 (The intrinsic metric of a subspace). Let X' be a subspace of a metric space (X, d), with X' equipped with the metric induced from the metric d. We suppose that the metric space X' is connected by rectifiable paths. We shall call the *intrinsic metric* of X' the length metric on X' associated to this induced metric.

Corollary 2.1.12 implies that X', equipped with its intrinsic metric, is a length space. With respect to this metric, the distance between two points in X' can be defined using the lengths of paths in this space, with no reference to an ambient space; this is the reason for which this metric is called intrinsic. In particular, if the space (X, d) is itself connected by rectifiable paths, then the *intrinsic metric of* X is the length metric d_ℓ associated to (X, d).

To see an example, let $X = \mathbb{E}^n$ or $\mathbb{H}^n$ for some $n \geq 2$ and let P be a finite subset of X. Then, the subspace metric on $X \setminus P$ is a length metric, and therefore it is the intrinsic metric of that subspace. However, if we consider a (closed or open) ball B^n of positive radius in X, then the subspace metric on $X \setminus B^n$ is not a length metric, and therefore it is not the intrinsic metric.

We recall that the *diameter* of a metric space X is defined as

$$\operatorname{diam}(X) = \sup_{x,y \in X} d(x, y).$$

A *bounded* subset of a metric space is a subset of finite diameter, with respect to the metric induced on that subset.

The following theorem is a general version of a theorem in differential geometry that is attributed to H. Hopf and W. Rinow. It gives a characterization of compact subsets of complete locally compact length spaces that is analogous to the well-known characterization of compact subsets of $\mathbb{R}$: the compact subsets are the closed and bounded subsets. The proof that we give here follows that of Gromov in [56] p. 9.

Theorem 2.1.15 (Hopf–Rinow). *Let X be a complete and locally compact length space. Then the compact subsets of X are the closed and bounded subsets of this space.*

Proof. We prove the non-trivial part of the statement, that is, the fact that any closed and bounded subset of X is compact. For that it suffices to prove that any closed ball in X is compact. Without loss of generality, we shall suppose that $\mathrm{diam}(X) = \infty$.

Let x be a point in X and let us consider a closed ball $B(x, r)$ of center x and radius $r > 0$.[1] Since X is locally compact, $B(x, r)$ is compact if r is small enough. We claim that this ball is compact for all $r > 0$. To see this, we first prove the following

Lemma 2.1.16. *Let X be a complete length space, let x be a point in X and let ρ be a positive real number such that for all $r < \rho$, the closed ball $B(x, r)$ is compact. Then the closed ball $B(x, \rho)$ is also compact.*

Proof. We recall that a metric space is said to be *precompact* if for every $\epsilon > 0$, we can cover this set with a finite number of balls of radii $\leq \epsilon$ and that in a complete metric space, a subset is compact if and only if it is closed and precompact. Therefore, it suffices to prove that $B(x, \rho)$ is precompact.

Let ϵ be a positive real number. Without loss of generality, we take ϵ small enough so that $\rho - (\epsilon/3) > 0$ and we take r satisfying $\rho - \epsilon/3 < r < \rho$. By compactness of $B(x, r)$, we can find a finite sequence of points $x_1, \ldots, x_n$ in X such that $B(x, r)$ is contained in the union of the closed balls $B(x_i, \epsilon/3)$. Now let y be an element in $B(x, \rho)$. Since X is a length space, we can find, by Lemma 2.1.4, a point z in X such that $|x - z| \leq \rho - \epsilon/3$ and $|z - y| \leq 2\epsilon/3$. Therefore, $z \in B(x, r)$ and there exists an integer i in $\{1, \ldots, n\}$ such that $z \in B(x_i, \epsilon/3)$. The inequality $|y - z| \leq 2\epsilon/3$ implies $|y - x_i| \leq |y - z| + |z - x_i| \leq \epsilon$, which shows that $B(x, \rho)$ is contained in the union of the balls $B(x_i, \epsilon)$. Thus, $B(x, \rho)$ is compact. This proves Lemma 2.1.16. □

Now we continue the proof of Theorem 2.1.15.

Again, let us consider an arbitrary point x in X and let

$$I_x = \{r \geq 0 \text{ such that the closed ball } B(x, r) \text{ is compact}\}.$$

[1] We warn the reader that in this book, $B(x, r)$ sometimes denote a closed ball of radius x and center r, and sometimes it denotes an open ball of center x and radius r; of course, each time we specify whether the ball is closed or open.

We have $0 \in I_x$. Furthermore, if I_x contains some point t, then it contains the whole interval $[0, t]$ (we use the fact that a closed subset of a compact set is compact). Thus, I_x is an interval of $[0, \infty[$ and 0 is an endpoint of that interval. By Lemma 2.1.16, the interval I_x is closed and therefore either it is of the form $[0, \rho]$, with $\rho < \infty$, or it is equal to $[0, \infty[$. We show that the first case cannot occur, and this will complete the proof of Theorem 2.1.15.

We reason by contradiction. We assume that $I_x = [0, \rho]$ with $\rho < \infty$. For each y in X, let r_y be a positive real number such that the closed ball $B(y, r_y)$ is compact. Such an r_y exists since X is locally compact. For all $r \geq 0$, we denote by $\overset{\circ}{B}(y, r)$ the open ball with center y and radius r. We have $B(x, \rho) \subset \bigcup_{y \in B(x,\rho)} \overset{\circ}{B}(y, r_y/2)$. Since the closed ball $B(x, \rho)$ is compact, there exists a finite set $F \subset B(x, \rho)$ such that $B(x, \rho) \subset \bigcup_{y \in F} \overset{\circ}{B}(y, r_y/2)$.

We set $r_0 = \min_{y \in F} r_y/2$. We have $r_0 > 0$. We show that the ball $B(x, \rho + r_0/2)$ is compact, and this will give the desired contradiction.

Let z be an arbitrary point in $B(x, \rho + r_0/2)$. Thus, $|x - z| \leq \rho + r_0/2$. Since X is a length space, by Lemma 2.1.4, there exists a point u in X satisfying $|x - u| \leq \rho$ and $|u - z| \leq r_0$. Thus, we have $u \in B(x, \rho)$, and therefore there exists an element y in F such that $u \in \overset{\circ}{B}(y, r_y/2)$. On the other hand, we have $|z - u| \leq r_0 \leq r_y/2$, which gives

$$|z - y| \leq |z - u| + |u - y| \leq r_y/2 + r_y/2 = r_y,$$

which implies that $z \in \overset{\circ}{B}(y, r_y)$. Thus, we obtain $B(x, \rho + r_0) \subset \bigcup_{y \in F} B(y, r_y)$, which implies that the ball $B(x, \rho + r_0)$ is compact, which is a contradiction. This proves Theorem 2.1.15. □

Example 2.1.17. The following three examples show that in Theorem 2.1.15, we cannot discard any of the hypotheses.

(i) Let $\mathbb{R}$ be equipped with the metric δ defined by $\delta(x, y) = \min(1, |x - y|)$. It is easy to see that the topology induced from δ on $\mathbb{R}$ is the usual topology (in fact, for any metric space (X, d), the identity map $(X, d) \to (X, \min(1, d))$ is a homeomorphism since the two metrics coincide locally) and that $(\mathbb{R}, \delta)$ is complete and locally compact. This space is not a length space since for all x and y in $\mathbb{R}$ satisfying $|x - y| > 1$, we have $\delta(x, y) = 1$ whereas $\delta_\ell(x, y) = |x - y| > 1$. The space $(\mathbb{R}, \delta)$ is a closed and bounded subset of itself and it is not compact.

(ii) The open interval $]0, 1[$ equipped with its usual metric is a length space that is locally compact and not complete. Here also, the space itself is a closed and bounded subset of itself and it is not compact.

(iii) Let X be an infinite-dimensional Banach space. As any normed space, X is a length space (see Example 2.1.3 (ii) above). This space is not locally compact since its unit ball $B(0, 1) \subset X$ is a closed bounded subset of X which is not compact.

The following simple fact about the existence of paths of bounded length will be useful in the next chapter.

Proposition 2.1.18. *Let X be a length space, let x be a point in X and let r be a positive real number. For each y and z in the open ball $B(x, r)$ of radius r and center x, there exists a path $\gamma\colon [a, b] \to X$ of length $< 2r$ joining y and z. Furthermore, the image of any such path is contained in the open ball $B(x, 2r)$ of center x and radius $2r$.*

Proof. By the triangle inequality, we have $|y-z| \leq |y-x|+|z-x| < 2r$. Since X is a length space, there exists a path $\gamma\colon [a, b] \to X$ of length $< 2r$ joining y and z. Let us show that the image of such a path is necessarily contained in $B(x, 2r)$. We reason by contradiction. Suppose that there exists t in $[a, b]$ such that $\gamma(t)$ is not contained in $B(x, 2r)$. Then we would have

$$|y - \gamma(t)| \geq |x - \gamma(t)| - |x - y| > r$$

and

$$|z - \gamma(t)| \geq |x - \gamma(t)| - |x - z| > r.$$

Therefore, we obtain

$$L(\gamma) = L(\gamma_{|[0,t]}) + L(\gamma_{|[t,b]}) \geq |y - \gamma(t)| + |z - \gamma(t)| > 2r,$$

which is a contradiction. □

2.2 Geodesics

Definition 2.2.1 (Geodesic path, geodesic line and geodesic ray). Let X be a metric space. A *geodesic path* (or, simply, a *geodesic*) in X is a path $\gamma\colon [a, b] \to X$ that is distance-preserving, that is, such that $|\gamma(t_1) - \gamma(t_2)| = |t_1 - t_2|$ for all t_1 and t_2 in X. A *geodesic ray* in X is a distance-preserving map $\gamma\colon [0, \infty[\to X$, and a *geodesic line* in X is a distance-preserving map $\gamma\colon \mathbb{R} \to X$.[2]

It follows easily from the definition that geodesic paths, geodesic rays and geodesic lines are injective and that the restriction of a geodesic path to a closed sub-interval of its domain is again a geodesic path.

Definition 2.2.2 (Geodesic segment and straight line). Let X be a metric space. A *geodesic segment* in X is the image of a geodesic path in X. A *straight line* in X is the image of a geodesic line in this space.

[2] We warn the reader that the definition of geodesic that we use here is more restrictive than the usual definition of geodesic in Riemannian geometry (which is also the definition used by Busemann, see [28] p. 32), where a geodesic is a *locally* distance-preserving map.

If a path $\gamma: [a, b] \to X$ joins two points x and y, then we say that the geodesic segment $\gamma([a, b])$ joins these two points. We note that in general, if x and y are two points in a metric space X, there might exist zero, one or more than one geodesic segment joining them. We denote by $[x, y]$ the geodesic segment $\gamma([a, b])$ provided there is no possible ambiguity.

Example 2.2.3. L be a nonempty Euclidean open segment in $\mathbb{E}^2$ and let $X = \mathbb{E}^2 \setminus L$, equipped with the induced (subspace) metric. Then, it is clear that at the two endpoints of the segment L, there is not uniqueness of prolongation of geodesics and that geodesic segments in X are concatenations of Euclidean segments. It is also easy to see that if instead of removing an open Euclidean segment L we remove a closed segment, then such a phenomenon does not occur (see Figure 2.3).

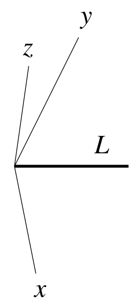


Figure 2.3. L is an open segment and $X = \mathbb{E}^2 \setminus L$. The geodesic segments joining x to y and z are concatenations of Euclidean geodesic segments. Prolongation of geodesics is not unique (Example 2.2.3).

Natural parametrization of a geodesic segment. The points on a geodesic segment $[x_0, x_1]$ are naturally parametrized by the interval $[0, 1]$. In this parametrization, we shall denote by x_t, or by $(1 - t)x_0 + tx_1$, the point on $[x_0, x_1]$ situated at distance $t|x_0 - x_1|$ from the point x_0.

Lemma 2.2.4. *Let $[x, y]$ be a geodesic segment in a metric space X and let $\gamma_1: [a_1, b_1] \to X$ and $\gamma_2: [a_2, b_2] \to X$ be two geodesics whose images are $[x, y]$. Then the two intervals $[a_1, b_1]$ and $[a_2, b_2]$ of $\mathbb{R}$ have the same length and there exists a unique real number α such that $\gamma_2(t) = \gamma_1(t + \alpha)$ for all t in $[a_2, b_2]$. In particular, the paths γ_1 and γ_2 have the same length.*

Proof. Since a geodesic map is injective, there exists a map, that we denote by $\gamma_1': [x, y] \to [a, b]$ and that is a "left inverse" of γ_1, that is, γ_1' satisfies the relation $\gamma_1' \circ \gamma_1 = \mathrm{Id}_{[x,y]}$. Since geodesic paths are distance-preserving, the map $\gamma_1' \circ \gamma_2$ is a translation between intervals of $\mathbb{R}$, and the result follows. □

Of course, there is a similar result for geodesic lines: two geodesic lines $\gamma_1 \colon \mathbb{R} \to X$ and $\gamma_2 \colon \mathbb{R} \to X$ have the same image if and only if there exists a real number α such that $\gamma_2(t) = \gamma_1(t + \alpha)$ for all t in $\mathbb{R}$.

From Lemma 2.2.4, we can make the following definition that will be useful for us later on:

Definition 2.2.5 (The length of a geodesic segment). The *length of a geodesic segment* $[x, y]$ in a metric space X is the length of an arbitrary geodesic path in X whose image is $[x, y]$.

Remark (Length pseudo-metric spaces). Given a space equipped with a pseudo-distance, one can compute the "lengths" of paths with respect to this pseudo-distance, by considering subdivisions of the domain of the path and defining the length as the supremum over all total variations of pseudo-distances with respect to subdivisions of this domain, imitating the definition of the length metric d_ℓ associated to a metric d. One says that the pseudo-metric space is a *length pseudo-metric space* if the pseudo-distance between any two points is equal to the infimum of the pseudo-lengths of paths joining them. It is known that the Kobayashi pseudo-distance is always a length pseudo-metric whereas the Carathéodory pseudo-distance is not (see [91] p. 54). Theodore Barth and then Jean-Pierre Vigué gave several interesting examples in which the Carathéodory pseudo-distance is not a length pseudo-metric (see [10] and [140]).

Proposition 2.2.6. *Let X be a metric space and let $\gamma \colon [a, b] \to X$ be a geodesic path. Then γ is parametrized by arclength.*

Proof. Let u and v be two real numbers satisfying $a \leq u < v \leq b$. For any subdivision $\sigma = (t_i)_{i=1,\dots,n}$ of $[u, v]$, we have

$$V_\sigma(\gamma_{|[u,v]}) = \sum_{i=0}^{n-1} |\gamma(t_i) - \gamma(t_{i+1})| = \sum_{i=0}^{n-1} (t_{i+1} - t_i) = v - u.$$

Therefore,

$$L(\gamma_{|[u,v]}) = \sup_\sigma V_\sigma(\gamma_{|[u,v]}) = v - u,$$

which shows that γ is parametrized by arclength. □

Proposition 2.2.7. *Let X be a metric space and let $\gamma \colon [a, b] \to X$ be a path which is parametrized by arclength. The following three properties are equivalent:*

(i) *γ is a geodesic;*

(ii) *for all real numbers u and v satisfying $a \leq u \leq v \leq b$, we have*

$$|\gamma(a) - \gamma(v)| = |\gamma(a) - \gamma(u)| + |\gamma(u) - \gamma(v)|;$$

(iii) $L(\gamma) = |\gamma(a) - \gamma(b)|$.

Proof. Let us first show that (i) $\Rightarrow$ (ii). If γ is a geodesic, then, for all u and v satisfying $a \leq u \leq v \leq b$, we have

$$|\gamma(a) - \gamma(v)| = v - a = v - u + u - a = |\gamma(u) - \gamma(v)| + |\gamma(a) - \gamma(u)|.$$

Now let us show that (ii) $\Rightarrow$ (iii). Let $\sigma = (t_i)_{i=0,\dots,n}$ be a subdivision of $[a, b]$. By applying $(n - 2)$ times Property (ii), we have

$$V_\sigma(\gamma) = \sum_{i=0}^{n-1} |\gamma(t_i) - \gamma(t_{i+1})| = |\gamma(a) - \gamma(b)|.$$

Taking the supremum over all subdivisions σ, we obtain $L(\gamma) = |\gamma(a) - \gamma(b)|$.

Finally, let us prove that (iii) $\Rightarrow$ (i). We have, for all $a \leq u \leq v \leq b$,

$$\begin{aligned} L(\gamma) &= |\gamma(a) - \gamma(b)| \\ &\leq |\gamma(a) - \gamma(u)| + |\gamma(u) - \gamma(v)| + |\gamma(v) - \gamma(b)| \\ &\leq |\gamma(a) - \gamma(u)| + L(\gamma|_{[u,v]}) + |\gamma(v) - \gamma(b)| \\ &\leq L(\gamma|_{[a,u]}) + L(\gamma|_{[u,v]}) + L(\gamma|_{[v,b]}) \\ &= L(\gamma). \end{aligned}$$

Therefore, all the inequalities in the last sequence are equalities and we have, for every u and v in $[a, b]$, $|\gamma(u) - \gamma(v)| = L(\gamma|_{[u,v]})$. Since γ is parametrized by arclength, we have $L(\gamma|_{[u,v]}) = |\gamma(u) - \gamma(v)|$, which implies $|\gamma(u) - \gamma(v)| = |u - v|$, which shows that γ is a geodesic. □

Definition 2.2.8 (Affinely reparametrized geodesic). Let $\gamma : [a, b] \to X$ be a path in a metric space X. We say that γ is an *affinely reparametrized geodesic*[3] if either γ is a constant path or there exists a geodesic path $\gamma' : [c, d] \to X$ such that $\gamma = \gamma' \circ \psi$ where $\psi : [a, b] \to [c, d]$ is the unique affine homeomorphism between the intervals $[a, b]$ and $[c, d]$, that is, the map defined by

$$\psi(x) = \big((d - c)x + (bc - ad)\big)/(b - a).$$

Proposition 2.2.9. *If $\gamma : [0, 1] \to X$ is an affinely reparametrized geodesic, then for all real numbers u and v satisfying $0 \leq u \leq v \leq 1$, we have $|\gamma(u) - \gamma(v)| = L(\gamma)|u - v|$.*

Proof. Using the arguments of Proposition 1.2.9 about paths parametrized proportionally to arclength, we obtain $|\gamma(u) - \gamma(v)| \leq L(\gamma)|u - v|$. Writing $\gamma =$

[3] Some authors use here the term *affine geodesic*. We prefer the term *affinely reparametrized geodesic* to avoid confusion with affine geodesic paths in normed vector spaces (Chapter 5 below).

$\gamma' \circ \psi$, with $\gamma' \colon [c, d] \to X$ a geodesic path in X and with $\psi \colon [a, b] \to [c, d]$ an affine homeomorphism, we have, for all u and v satisfying $0 \leq u \leq v \leq 1$, $|\gamma'(\psi(u)) - \gamma'(\psi(v))| = L(\gamma_{|[\psi(u),\psi(v)]})$. From that, we see that under our hypotheses, the large inequality in the proof of Proposition 1.2.9 is an equality, which implies that $|\gamma(u) - \gamma(v)| = L(\gamma)|u - v|$. This proves Proposition 2.2.9. □

The next proposition is a partial converse to the preceding one:

Proposition 2.2.10. *Let X be a metric space, let $\gamma \colon [0, 1] \to X$ be a path and let K be a nonnegative real number satisfying $|\gamma(u) - \gamma(v)| = K|u - v|$ for all u and v in* $[0, 1]$. *Then γ is an affinely reparametrized geodesic and $K = L(\gamma)$.*

Proof. If $K = 0$, then γ is a constant path, and therefore, by definition, it is an affinely reparametrized geodesic. Now, suppose that K is positive. Let $\psi \colon [0, K] \to [0, 1]$ be the affine map defined by $\psi(u) = u/K$ for all u in $[0, K]$ and let $\gamma' \colon [0, 1] \to X$ be the composed map $\gamma \circ \psi$. Then we have, for all u and v in $[0, K]$,

$$|\gamma'(u) - \gamma'(v)| = |\gamma \circ \psi(u) - \gamma \circ \psi(v)| = |\gamma(u/K) - \gamma(v/K)| = |u - v|,$$

which shows that γ' is geodesic. Thus, γ is an affinely reparametrized geodesic. The fact that the domain of γ' is the interval $[0, K]$ implies that $L(\gamma') = K$, which gives $L(\gamma) = K$. □

Lemma 2.2.11. *Let X be a metric space and let $[x, y]$ and $[y, z]$ be two geodesic segments in X (having y as a common point). Then the union $[x, y] \cup [y, z]$ is a geodesic segment if and only if $|x - z| = |x - y| + |y - z|$.*

Proof. Let γ_1 and γ_2 be two geodesic paths in X whose images are respectively $[x, y]$ and $[y, z]$ and let $\gamma = \gamma_1 * \gamma_2$ be their concatenation . If $|x - z| = |x - y| + |y - z|$, then $L(\gamma) = |x - z|$, which shows that γ is geodesic. Then the union $[x, y] \cup [y, z]$, which is the image of γ, is a geodesic segment in X.

Conversely, if $[x, y] \cup [y, z]$ is a geodesic segment, then this segment is an embedded image of a closed interval, which shows that the intersection of the two segments $[x, y]$ and $[y, z]$ is reduced to the point y. This implies that the union $[x, y] \cup [y, z]$ is the image of a geodesic path γ which is the concatenation of two geodesic paths γ_1 and γ_2 whose images are respectively $[x, y]$ and $[y, z]$. Thus, we have $|x - y| = L(\gamma_1)$, $|y - z| = L(\gamma_2)$ and

$$|x - y| + |y - z| = L(\gamma_1) + L(\gamma_2) = L(\gamma) = |x - z|.$$

This completes the proof of Lemma 2.2.11. □

The following notion was introduced by Menger.

Definition 2.2.12 (Betweenness). Given three points x, y and z in a metric space, we say that y *lies between* x *and* z if these three points are pairwise distinct and if we have $|x - z| = |x - y| + |y - z|$.

It is clear that the relation of betweenness is symmetric in the following sense: if y lies between x and z, then y lies between z and x. The following proposition expresses a transitivity property for this relation.

Proposition 2.2.13 (Transitivity of betweenness). *Let X be a metric space and let x, y, z and t be pairwise distinct points of X. Then we have the following equivalence:*

y lies between x and z and z lies between x and t

$$\Updownarrow$$

y lies between x and t and z lies between y and t.

Proof. If y lies between x and z and if z lies between x and t, then

$$|x - y| + |y - z| = |x - z|$$

and

$$|x - z| + |z - t| = |x - t|.$$

Then we have

$$|x - t| = |x - z| + |z - t| = |x - y| + |y - z| + |z - t|$$

which gives

$$|x - t| \geq |x - y| + |y - z| \geq |x - t|,$$

and therefore the two inequalities in the previous line are equalities. Thus, we obtain

$$|x - y| + |y - t| = |x - t|$$

and

$$|y - z| + |z - t| = |y - t|,$$

which says that y lies between x and t and that z lies between y and t. The converse implication follows by symmetry. □

Warning. Some other forms of "transitivity" of betweenness are not true in general. For instance, by considering points on a great circle in the sphere S^2, it is easy to construct examples of points x, y, z and t with y lying between x and z, with z lying between y and t, but where z does not lie between x and t.

Proposition 2.2.14 (Betweenness and existence of geodesics). *Let X be a geodesic space and let x, y and z be three points in X. Then y lies between x and z if and only if these three points are pairwise distinct and if there exists a geodesic segment $[x, z]$ containing y.*

Proof. This follows from Lemma 2.2.11. □

We also note the following fact that will be useful for us:

Proposition 2.2.15. *Let X be a metric space, let x be a point in X and let r be a positive real number. If y and z are points in the open* (*respectively closed*) *ball of center x and radius r and if $\gamma : [a, b] \to X$ is a geodesic path joining y and z, then the image of γ is contained in the open* (*respectively closed*) *ball of center x and radius $2r$.*

Proof. The proof consists of a slight variation of the proof of Proposition 2.1.18, and we omit it. □

Let us note that in general, one cannot expect a better conclusion than that of Proposition 2.2.15, as the following example shows. Let X be the unit sphere S^2 equipped with its intrinsic length metric. We recall that in this metric, the geodesic segments are subarcs of great circles and that the total length of a great circle is equal to 2π. Consider a point x in X, let r be a real number that is slightly greater than $\pi/2$ and let y and z be two points in $B(x, r)$ that are situated on a great circle $\mathcal{C}$ containing the point x and satisfying $|x - y| > \pi/2$ and $|x - z| > \pi/2$, with x lying between y and z. Finally, let γ be a geodesic path in X joining y and z. The image of γ is the arc of the great circle that joins y and z and that contains the point that is antipodal to x. We have $L(\gamma) < 2r$ and the image of γ is almost entirely contained in the complement of the ball $B(x, r)$.

2.3 Limits of geodesics

Proposition 2.3.1 (Limits of geodesic paths). *Let X be a metric space and for every integer $n \geq 0$, let $\gamma_n : [a, b] \to X$ be a geodesic* (*respectively an affinely reparametrized geodesic*). *If the sequence (γ_n) converges pointwise to a map $\gamma : [a, b] \to X$, then γ is a geodesic* (*respectively an affinely reparametrized geodesic*).

Proof. We prove the proposition in the case where each γ is a geodesic. The proof for affinely reparametrized geodesics can be done by composing with appropriate affine maps. For every t_1 and t_2 in $[a, b]$ and for every $n = 1, 2, \ldots$, we have $|\gamma_n(t_1) - \gamma_n(t_2)| = |t_1 - t_2|$. By the continuity of the distance function, we obtain

$$\lim_{n\to\infty} |\gamma_n(t_1) - \gamma_n(t_2)| = |\gamma(t_1) - \gamma(t_2)| = |t_1 - t_2|,$$

which shows that γ is a geodesic. □

Proposition 2.3.2. *Let X be a compact metric space and for every $n \geq 0$ let $\gamma_n : [a, b] \to X$ be an affinely reparametrized geodesic. Then the sequence $(\gamma_n)_{n\geq 0}$ has a subsequence which converges uniformly to an affinely reparametrized geodesic.*

Proof. For all $n \geq 0$, we have $L(\gamma_n) = |\gamma_n(a) - \gamma_n(b)|$. Since X is compact, its diameter is finite. Therefore, the sequence $\big(L(\gamma_n)\big)_{n\geq 0}$ is bounded above by a constant that is independent of n. Then Proposition 1.4.11 implies that $(\gamma_n)_{n\geq 0}$ has a subsequence that converges uniformly to a path γ. By Proposition 2.3.1, γ is an affinely reparametrized geodesic path. □

Proposition 2.3.3. *Let X be a metric space and for every integer $n \geq 0$, let $\gamma_n : [0, 1] \to X$ be an affinely reparametrized geodesic. If the sequence $(\gamma_n)_{n\geq 0}$ converges pointwise to an affinely reparametrized geodesic $\gamma : [0, 1] \to X$, then this convergence is uniform.*

Proof. We reason by contradiction. Suppose that the convergence of (γ_n) to γ is not uniform. Then we can find a real number $\epsilon > 0$ such that for every integer $n_0 \geq 0$, there exists an integer $n \geq n_0$ and a real number t_n in $[a, b]$ such that $|\gamma_n(t_n) - \gamma(t_n)| \geq \epsilon$. Therefore, we can find a sequence of integers $(n_i)_{i\geq 0}$ that tends to infinity as $i \to \infty$ and a sequence $(t_{n_i})_{i\geq 0}$ of real numbers in $[a, b]$ satisfying $|\gamma_{n_i}(t_{n_i}) - \gamma(t_{n_i})| \geq \epsilon$ for all $n_i \geq 0$. Since the interval $[a, b]$ is compact, we can assume, up to passing to a subsequence, that the sequence (t_{n_i}) converges to some point t in $[a, b]$ as $i \to \infty$. For all $i \geq 0$, the path γ_{n_i} is, up to a change of parameter, a geodesic path. Therefore, we have $L(\gamma_{n_i}) = |\gamma_{n_i}(0) - \gamma_{n_i}(1)|$ and since $(\gamma_{n_i})_{i\geq 0}$ converges pointwise to γ as i tends to infinity, the sequence $\big(L(\gamma_{n_i})\big)_{i\geq 0}$ converges to $L(\gamma) = |\gamma(0) - \gamma(1)|$. Let us set $M = L(\gamma) + 1$. Then, under our hypotheses, there exists an integer $N > 0$ such that for all $n_i \geq N$, the following three conditions are satisfied:

$$L(\gamma_{n_i}) \leq M,$$

$$|\gamma_{n_i}(t) - \gamma(t)| \leq \frac{\epsilon}{4M},$$

and

$$|t_{n_i} - t| \leq \frac{\epsilon}{4}.$$

Thus we have, for all $n_i \geq N$,

$$\begin{aligned}
\epsilon &\leq |\gamma_{n_i}(t_{n_i}) - \gamma(t_{n_i})| \\
&\leq |\gamma_{n_i}(t_{n_i}) - \gamma_{n_i}(t)| + |\gamma_{n_i}(t) - \gamma(t)| + |\gamma(t) - \gamma(t_{n_i})| \\
&= L(\gamma_{n_i})|t_{n_i} - t| + |\gamma_{n_i}(t) - \gamma(t)| + L(\gamma_{n_i})|t_{n_i} - t| \\
&\leq M|t_{n_i} - t| + |\gamma_{n_i}(t) - \gamma(t)| + M|t_{n_i} - t| \\
&= \frac{\epsilon}{4} + \frac{\epsilon}{4} + \frac{\epsilon}{4} = \frac{3\epsilon}{4},
\end{aligned}$$

which is a contradiction. Therefore, the sequence (γ_n) converges uniformly to γ. □

2.4 Geodesic spaces

Definition 2.4.1 (Geodesic space). A metric space X is said to be *geodesic* if given two arbitrary points in X there exists a geodesic path that joins them.

Proposition 2.4.2. *A geodesic space is a length space.*

Proof. Let X be a geodesic space, let x and y be two arbitrary points in X and let $\gamma\colon [a,b] \to X$ be a geodesic joining them. We have $|a-b| = |x-y|$ and since γ is parametrized by arclength, we have $|a-b| = L(\gamma)$. Hence, we obtain $|x-y| = L(\gamma)$, which implies that X is a length space. □

Examples 2.4.3 (Geodesic spaces).

(i) *Euclidean space.* For any $n \geq 0$, Euclidean space $\mathbb{E}^n$ is a geodesic space; a geodesic segment joining any two points in $\mathbb{E}^n$ is the affine segment joining them.

(ii) *Spheres.* For any $n \geq 0$, the sphere S^n equipped with its intrinsic metric associated to the metric induced by its inclusion as the unit sphere in $\mathbb{R}^{n+1}$ is a length space; we already recalled that for any two points in S^n, a geodesic segment joining them is a subset of the great circle passing through these points.

(iii) *Hyperbolic space.* For every $n \geq 2$, n-dimensional hyperbolic space $\mathbb{H}^n$ is a geodesic space. In this space, any geodesic segment is contained in a straight line. In the conformal ball model B^n of this space (see Example 2.1.3 (vii) above), straight lines are of two sorts (see Figure 2.4 (a)):

• the intersection of the unit ball B^n with a Euclidean circle of $\mathbb{R}^n$ that meets perpendicularly the boundary of B^n, or

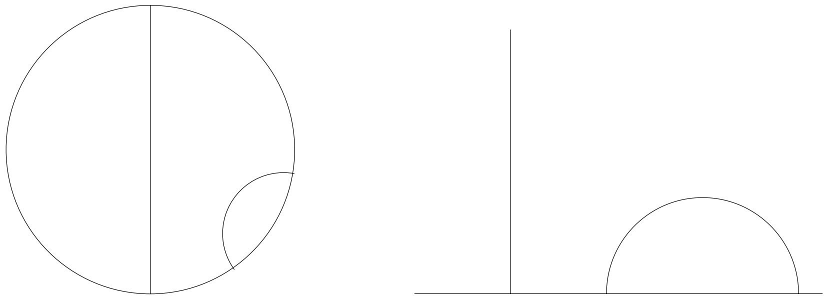
• a diameter of B^n.

In the upper half-space model H^n of $\mathbb{H}^n$, straight lines are also (from the Euclidean point of view) of two sorts (see Figure 2.4 (b)):

• the intersection of H^n with a vertical Euclidean straight lines of $\mathbb{R}^n$, or

• the intersection of H^n with a circle of $\mathbb{R}^n$ that meets perpendicularly the boundary of H^n.

By examining any one of these models, one can easily see that any two points in $\mathbb{H}^n$ are joined by a geodesic. We also see from the description of the geodesics in the conformal ball model that if $r\colon [0,\infty[\to B^n$ is a geodesic ray, then, $r(t)$ converges (in the Euclidean metric of the closed ball) to a point in the sphere S^{n-1}, considered as the boundary of B^n. We shall denote by $r(\infty)$ the limit point of the geodesic ray r. Likewise, any geodesic line $\gamma\colon \mathbb{R} \to X$ has two distinct limit points on the sphere



(a) (b)

Figure 2.4. Geodesics in the ball model and in the upper half-plane model of hyperbolic plane.

S^{n-1}, the point $\gamma(-\infty) = \lim_{t\to-\infty}\gamma(t)$ and the point $\gamma(\infty) = \lim_{t\to\infty}\gamma(t)$ (again, the limits are taken with respect to the Euclidean metric of the closed ball).

(iv) *Convex subsets of normed vector spaces.* This example is more general than example (i) above. Any convex subset of a normed vector space, equipped with the induced metric, is a geodesic space (see Proposition 5.3.7 below).

(v) *Teichmüller space.* Let $S = S_g$ be a closed oriented surface of genus $g \geq 1$. The *Teichmüller space* $\mathcal{T}_g$ of S is the space of isotopy classes of hyperbolic structures on that surface. We recall that a *hyperbolic structure* on S is a collection of local charts $\{(U_i, \varphi_i)\}_{i\in\mathcal{I}}$ where $\{U_i\}_{i\in\mathcal{I}}$ is a collection of open subsets of S whose union covers S and where for each $i \in \mathcal{I}$, φ_i is a homeomorphism from U_i onto an open subset of hyperbolic space $\mathbb{H}^2$ such that any map of the form $\varphi_i \circ \varphi_j^{-1}$ is a local isometry of $\mathbb{H}^2$, whenever it is defined.

Equivalently, $\mathcal{T}_g$ is the space of isotopy classes of conformal structures on S.[4] We also recall that a *conformal structure* on S is a collection of local charts $\{(U_i, \varphi_i)\}_{i\in\mathcal{I}}$ where $\{U_i\}_{i\in\mathcal{I}}$ is a set of open subsets of S whose union covers this surface and where for each $i \in \mathcal{I}$, φ_i is a homeomorphism from U_i onto an open subset of the complex plane $\mathbb{C}$ such that any map of the form $\varphi_i \circ \varphi_j^{-1}$ is holomorphic, wherever

[4] We note that in the case of a non-compact surface, the definition of Teichmüller space is more involved. For instance, given two integers n and g that are both ≥ 1, if $S_{n,g}$ is the surface obtained from the closed surface S_g by removing n points (which we shall call the *punctures*), then, the Teichmüller space $\mathcal{T}_{g,n}$ of $S_{n,g}$ is the space of equivalence classes of *complete and finite volume* hyperbolic metrics on $S_{g,n}$. Equivalently, the hyperbolic metrics considered have the property that the neighborhood of each puncture of $S_{g,n}$ is isometric to a *cusp*, that is, a neighborhood of infinity in the quotient of a region in the upper half-plane model of $\mathbb{H}^2$ of the form $\{(x, y) \mid y > k\}$ by a map of the form $z \mapsto z + c$. If one uses the point of view of conformal structures, then, the conformal structures considered on $S_{g,n}$ are such that the neighborhood of each puncture is conformally equivalent to a punctured disk in the complex plane.

it is defined. A surface equipped with a conformal structure is also called a *Riemann surface*. The space $\mathcal{T}_g$ has a natural topology that makes it homeomorphic to $\mathbb{R}^{6g-6}$ if $g \geq 2$ and to $\mathbb{R}^2$ if $g = 1$. This topology is induced by several interesting metrics, the most famous one being the Teichmüller metric. Let us recall its definition. First, one defines the Teichmüller distance between two hyperbolic metrics on S as the infimum of the dilatations of diffeomorphisms of S that are isotopic to the identity, where the domain space is S equipped with the first metric and the target space is S equipped with the second metric. More precisely, if g_1 and g_2 are hyperbolic metrics on S and if $f : (S, g_1) \to (S, g_2)$ is a diffeomorphism, then, the *dilatation* of f is defined as

$$K(f) = \sup \left(\frac{\sup\{\|df_x(u)\| \text{ such that } u \in T_x S, \|u\| = 1\}}{\inf\{\|df_x(u)\| \text{ such that } u \in T_x S, \|u\| = 1\}} \right).$$

In this formula, the norm of the tangent vector $df_x(u)$ is measured with respect to the metric g_2 and the norm of the tangent vector u is measured with respect to the metric g_1. The *Teichmüller distance* between g_1 and g_2 is equal to

$$d(g_1, g_2) = \frac{1}{2} \inf\{\log K(f)\}$$

where the infimum is taken over the set of diffeomorphisms of S that are isotopic to the identity. The value $d(g_1, g_2)$ is invariant if we replace g_1 or g_2 by an isotopic metric. Therefore, d defines a map on $\mathcal{T}_g \times \mathcal{T}_g$, which is the Teichmüller metric.

The Teichmüller metric is a complete geodesic metric and each pair of distinct points in $\mathcal{T}_g$ is contained in a unique straight line. For $g = 1$, Teichmüller space equipped with the Teichmüller metric is isometric to the hyperbolic plane $\mathbb{H}^2$. All these results are due to Teichmüller (see [131]; see also Bers's paper [15] for a more modern treatment). There is a nice description of the geodesic lines of the Teichmüller metric (which are called *Teichmüller geodesics*) in terms of pairs of transverse measured foliations on S. In this description, the local coordinates of the conformal structure varying along a geodesic path, are deformed in a very simple manner, which is similar to the most natural affine deformation of a rectangle. Let us describe this deformation in a few lines. A *measured foliation* on a surface is a foliation equipped with a measure on transverse arcs which is invariant by isotopies of the transverse arcs in which each point of stays on the same leaf. The foliations considered are allowed to have singular points of the types described in Figure 2.5 (i.e. k-prong singularities with k being any integer ≥ 3). At the singular points, the transversality of two foliations is described in Figure 2.6.

A pair of transverse measured foliations F_1 and F_2 on S defines a conformal structure, and in fact, a privileged class of local parameters, in the following manner. In the neighborhood of each non-singular point, there is a natural parameter $z = x + iy$, where x is a parameter along the leaves of F_1 and y a parameter along the leaves of F_2, and where distances along the leaves of each of these foliations are measured by using the transverse measure of the other foliation. In this way, the leaves of F_1 (respectively F_2) are locally defined by the equation $y =$ constant (respectively $x =$

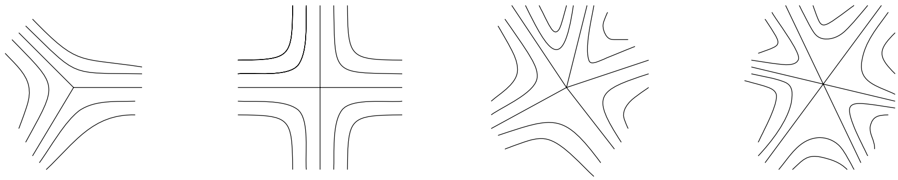


Figure 2.5. k-prong singular points with $k = 1, 2, 3$.

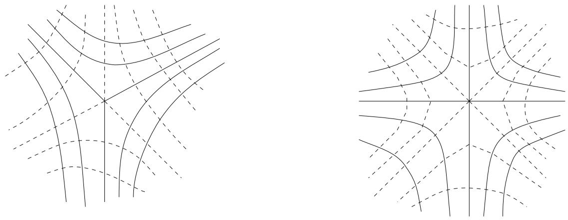


Figure 2.6. Two transverse foliations at a singular point.

constant). The local parameters $z = x + iy$ in the neighborhoods of the various nonsingular points are compatible with each other and they define a conformal structure on the complement of the singular points. This conformal structure extends in a unique way to a conformal structure on the whole surface S. Thus, the pair (F_1, F_2) of measured foliations determines a conformal structure on S. The leaves of the foliations F_1 and F_2 are mutually orthogonal with respect to this structure at each nonsingular point. A Teichmüller geodesic is then a map $\ell\colon \mathbb{R} \to \mathcal{T}_g$ of the form $\ell(t) = (e^{-t}F_1, e^tF_2)$, where F_1 and F_2 are measured foliations and where for any $\lambda > 0$, λF denotes the measured foliation obtained from F by multiplying its transverse measure by the factor λ. The pair $(e^{-t}F_1, e^tF_2)$ is identified here with the conformal structure determined by the pair of transverse measured foliations $(e^{-t}F_1, e^tF_2)$. In terms of the local coordinates $z = x + iy$, the variation of the conformal structure is defined by $(x, y) \mapsto (e^tx, e^{-t}y)$. We refer the reader to the paper [84] by S. Kerckhoff for a description of the Teichmüller metric and for interesting results on the behaviour of geodesic rays in Teichmüller space.

Let us finally note that the Teichmüller metric is a Finsler metric (this is a result of Earle and Eells, cf. [44]), that Teichmüller space has a natural structure of a complex manifold and that (by a result of Royden, cf. [125]), the Teichmüller metric is equal to the Kobayashi pseudo-metric of this complex manifold (and therefore, this pseudo-metric is a metric).

Example 2.4.4 (A non-geodesic length space). For all $n \geq 2$, Euclidean space $\mathbb{E}^n$ with a point p removed is a length space which is not a geodesic space. Indeed, if x

and y are two distinct points in $\mathbb{E}^n \setminus \{p\}$ such that the segment $[x, y]$ of $\mathbb{E}^n$ contains p, then there is no geodesic path in $\mathbb{E}^n \setminus \{p\}$ joining x and y.

Proposition 2.4.5. *Let X be a locally compact length space. Then every point x in X has a neighborhood $U = U(x)$ such that for every y and z in U, there exists a geodesic path in X joining them.*

Proof. Let x be a point in X. Since X is locally compact, we can find a positive real number r such that the open ball $B(x, 2r)$ of center x and radius $2r$ has compact closure in X. Now let U be the open ball $B(x, r)$ and let y and z be two arbitrary points in U. Since X is a length space, there exists a sequence $(\gamma_n)_{n\geq 0}$ of paths in X joining y and z such that $L(\gamma_n)$ converges to $|y - z|$ as $n \to \infty$. Thus, for n large enough, we have $L(\gamma_n) < 2r$ and by Proposition 2.1.18, the image of γ_n is contained in the closure of the ball $B(x, 2r)$. Since this closure is compact, it is complete, and Ascoli's Theorem (Theorem 1.4.9) shows that there exists a subsequence $(\gamma_{n_i})_{i\geq 0}$ of $(\gamma_n)_{n\geq 0}$ that converges to a path γ. The path γ joins y and z, its length is equal to $|y - z|$ and without loss of generality, we can assume that γ is parametrized by arclength. Then Proposition 2.2.7 shows that γ is a geodesic path and Proposition 2.1.8 shows that the image of γ is contained in $B(x, 2r)$. □

The next result is also attributed to H. Hopf and W. Rinow :

Theorem 2.4.6 (Hopf–Rinow: proper length spaces are geodesic). *Let X be a proper length space. Then for every x and y in X, there exists a geodesic path that joins these points.*

Proof. Proposition 1.4.12 shows that there exists a path of length $|x - y|$ joining x and y. We may assume without loss of generality that this path is parametrized by arclength. Therefore, by Proposition 2.2.7, this path is a geodesic. □

Remark. There are other useful forms of the theorem of Hopf–Rinow that combine Theorems 2.1.15 and 2.4.6, and we mention as an example the following statement, contained in [30], that Busemann considers as a generalization of the theorem of Hopf–Rinow due to Cohn-Vossen:

Let X be a locally compact length space. Then, the following three conditions are equivalent:

- the closed balls in X are compact;
- X is complete;
- every half-open geodesic path can be completed to a geodesic path; in other words, every distance-preserving map from a half-open interval $[a, b[\subset \mathbb{R}$ into X can be extended to a distance-preserving map $[a, b] \to X$.

By the Theorem of Hopf and Rinow (Theorems 2.4.6 and 2.1.15), a length space is geodesic if and only if it is complete and locally compact. It should be pointed out that these two hypotheses are necessary in order to have the conclusion of that theorem. Indeed, the space $\mathbb{E}^n$ ($n \geq 2$) with a point removed is locally compact, and it is not geodesic. The hypothesis that X is locally compact is also necessary, as the following example shows:

Example 2.4.7. Let X be a metric graph having exactly two vertices x and y and an infinite number of edges e_n ($n = 1, 2, \dots$), each such edges joining the vertices x and y and such that for every $n \geq 1$, the length of e_n is equal to $1 + 1/n$. The graph X is equipped with the associated length metric d (see Example 2.1.3 (iv)). The length space (X, d) is complete and it is not locally compact. There is no geodesic path joining the vertices x and y, since the length of any path joining these two points is > 1, whereas $|x - y| = 1$.

We can push this point further and prove the existence of local geodesics in homotopy classes with fixed points, using Ascoli's theorem. We first need to make the following definition:

Definition 2.4.8 (Local geodesic). Let X be a metric space and let $\gamma\colon [a, b] \to X$ be a path. Then γ is said to be a *local geodesic* if for all t in $[a, b]$ we can find a closed interval $I(t)$ containing t in its interior such that the restriction of γ to $I(t) \cap [a, b]$ is geodesic.

Proposition 2.4.9. *If $\gamma\colon [a, b] \to X$ is a local geodesic, then γ is parametrized by arclength.*

Proof. By compactness of $[a, b]$, there exists a finite number of open intervals $I_1, \dots, I_k$ in $\mathbb{R}$ that cover $[a, b]$ and such that for all $i = 1, \dots, k$, the restriction of γ to $I_i \cap [a, b]$ is geodesic, and therefore parametrized by arclength. Let $t_0, \dots, t_n$ be the sequence of points in $[a, b]$, ordered increasingly, that is obtained by taking the union of the set $\{a, b\}$ with the set of endpoints of the intervals I_i intersected with $[a, b]$. This sequence is a subdivision $(t_i)_{i=0,\dots,n}$ of $[a, b]$ and the restriction of γ to $[t_i, t_{i+1}]$ is parametrized by arclength, for all $i = 0, \dots, n - 1$. By Proposition 1.2.7, γ is parametrized by arclength. □

In the same way as for geodesic paths, it is often useful to deal with local geodesics pre-composed with affine maps, and we make the following definition:

Definition 2.4.10 (Affinely reparametrized local geodesic). Let $\gamma\colon [a, b] \to X$ be a path. We say that γ is an *affinely reparametrized local geodesic* if either γ is a constant map or there exists a local geodesic $\gamma'\colon [c, d] \to X$ such that $\gamma = \gamma' \circ \psi$, where $\psi\colon [a, b] \to [c, d]$ is the unique affine homeomorphism between the intervals $[a, b]$ and $[c, d]$.

Proposition 2.4.11. *Let X be a proper length space such that for each point p in X there exists a positive real number r such that the open ball $B(p,r)$ is simply connected, let x and y be two points in X and suppose that there exists a rectifiable path g joining x and y. Then, there exists a local geodesic joining x and y that is homotopic with fixed endpoints to g.*

Proof. By Proposition 1.4.13, there exists a path $\gamma : [a,b] \to X$ joining x and y whose length is shortest in the homotopy class with fixed endpoints of g. We may assume without lost of generality that γ is parametrized by arclength. Let us show that γ is a local geodesic. Given t in $[a,b]$, let r be a positive real number such that the open ball $B(\gamma(t), r)$ of center $\gamma(t)$ and radius r is simply connected and let $[t_1, t_2]$ be a closed interval containing t in its interior and satisfying $\gamma([t_1,t_2]) \subset B(\gamma(t), r/2)$. We claim that the restriction of γ to $[t_1,t_2]$ is a geodesic path. We prove this by contradiction. Suppose that $\gamma_{|[t_1,t_2]}$ is not geodesic. Since X is proper, there exists a path γ' joining $\gamma(t_1)$ to $\gamma(t_2)$ whose length is strictly less than the length of $\gamma_{|[t_1,t_2]}$. By Proposition 2.2.15, the image of γ' is contained in $B(\gamma(t), r)$ and since this ball is simply connected, the path γ' is homotopic with fixed endpoints to $\gamma_{|[t_1,t_2]}$. This implies that the path obtained from γ by replacing $\gamma_{|[t_1,t_2]}$ by γ' is homotopic to γ with endpoints fixed, and its length is strictly less than the length of γ, which is a contradiction. Thus, $\gamma_{|[t_1,t_2]}$ is geodesic. This completes the proof of the fact that γ is a local geodesic. □

We introduce two more notions that will be useful for us later on in these notes:

Definition 2.4.12 (Straight metric space). A metric space X is said to be *straight* if X is a geodesic metric space and if any geodesic segment in X is contained in a straight line.

Example 2.4.13 (Straight metric spaces). From the descriptions we gave in Examples 2.4.3, Euclidean space $\mathbb{E}^n$, hyperbolic space $\mathbb{H}^n$ and Teichmüller space $\mathcal{T}_g$ are straight.

Definition 2.4.14 (Uniquely geodesic space). A metric space X is said to be *uniquely geodesic* if for any x and y in X, there exists a unique geodesic segment joining them.

The study of geodesic paths in uniquely geodesic spaces is more convenient than in general metric spaces. For instance, in a proper uniquely geodesic space X, a sequence of affinely parametrized geodesic paths $\gamma_n : [0,1] \to X$ converges to a geodesic path $\gamma : [0,1] \to X$ if and only if the two sequences of endpoints $\gamma_n(0)$ and $\gamma_n(1)$ converge respectively to the points $\gamma(0)$ and $\gamma(1)$. (This follows for instance from Ascoli's theorem.) Furthermore, in a uniquely geodesic metric space X, one can talk about convex combinations of points: if x_0 and x_1 are any points in X, then, for any t in $[0,1]$, x_t is the unique point on the geodesic segment $[x_0, x_1]$ satisfying $|x_0 - x_1| = t$. We start with a few examples of uniquely geodesic metric spaces.

Examples 2.4.15 (Uniquely geodesic spaces).

(i) *Euclidean and hyperbolic spaces.* For every $n \geq 0$, Euclidean space $\mathbb{E}^n$ and hyperbolic space $\mathbb{H}^n$ are uniquely geodesic spaces. For hyperbolic space, this follows from the description of the geodesics in that space that we recalled in Examples 2.4.3 above.

(ii) $\mathbb{R}$-*trees.* An $\mathbb{R}$-*tree* is a metric space X that is characterized by the following two properties:

- for all x and y in X, there exists a unique topological segment that joins them (a topological segment being, by definition, a subset of X that is the homeomorphic image of a closed interval of $\mathbb{R}$);
- any topological segment in X is a geodesic segment.

It follows from this definition that an $\mathbb{R}$-tree is a uniquely geodesic space.

The notion of $\mathbb{R}$-tree was introduced by J. Tits in [138], as a generalization of the notion of local Bruhat–Tits building for rank-one groups, which itself generalizes the notion of simplicial tree. $\mathbb{R}$-trees made their appearance as an essential tool is the study of groups acting on hyperbolic manifolds in the work of J. Morgan and P. Shalen [112].

A particular class of $\mathbb{R}$-trees is the class of *simplicial metric trees*, that is, of metric graphs that are simply connected, equipped with one of the metrics defined in Example 2.1.3 (iv) above. An example of an $\mathbb{R}$-tree that is not homeomorphic to a simplicial tree is the two-dimensional plane $\mathbb{R}^2$ equipped with the metric for which the distance between two points x and y is equal to the Euclidean norm $\|x - y\|$ if x and y are situated on a Euclidean straight line passing through the origin, and to $\|x\| + \|y\|$ otherwise.

A special class of simplicial trees is that of homogeneous simplicial trees, to which we shall refer later on. We first recall that the *valency* of an edge in a simplicial graph is equal to the number of edges, counted with multiplicity, to which that edge belongs. In visual terms, the valency of a vertex is the number of edges that locally abut on that vertex. It is easy to see that in a simplicial tree, the two vertices of any edge are distinct. Now for any integer ≥ 2, a *homogeneous tree of degree* n is a metric simplicial tree that is nonempty and in which each vertex has valency n. The simplicial tree of degree n is unique up to homeomorphism, but of course, as a metric spaces, it depends on the choice of the lengths of its edges.

(iii) *Teichmüller space.* For $g \geq 1$, the Teichmüller space $\mathcal{T}_g$ of a closed Riemann surface of genus $g \geq 1$, equipped with its Teichmüller metric, is a uniquely geodesic space. This is one of the results of Teichmüller; cf. [132], which is also contained in [133]. For a more recent exposition, we refer the reader to Kerckhoff's paper [84].

(iv) *Normed vector spaces.* Normed vector spaces, that we study in Chapters 5 and 7, are examples of geodesic spaces, but not all of them are uniquely geodesic. We shall see that any normed vector spaces whose norm is associated to an inner product is an example of a uniquely geodesic space.

There exist length spaces in which there are points that have no neighborhoods that are uniquely geodesic (with respect to the metric induced on these neighborhoods). Let us give two examples of such spaces. In the first example, no point in the space has a neighborhood that is uniquely geodesic, and in the second example, every point except an isolated point has a neighborhood that is uniquely geodesic.

Examples 2.4.16 (Non locally uniquely geodesic spaces).

(i) ℓ^1 *norm.* Consider the vector space $\mathbb{R}^2$ equipped with the metric associated to the ℓ^1 norm, that is, the norm defined by $\|(x_1, y_1) - (x_2, y_2)\| = |x_1 - y_1| + |x_2 - y_2|$. Any path in this space whose image is an "escalator" (see Figure 2.7) or, more generally, any path whose image is the graph of a monotonic function is a geodesic segment. Therefore there are infinitely many geodesic segments joining any two points in that space, provided these points have distinct first and second coordinates.

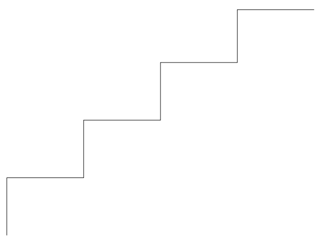


Figure 2.7. An escalator (Example 2.4.16 (i)).

(ii) *Euclidean cone.* Consider a metric space that contains a point having a neighborhood isometric to a Euclidean cone whose angle at the vertex is $< 2\pi$. To obtain such a space, we can take a standard Euclidean cone embedded in Euclidean 3-space, equipped with the length metric induced from its embedding, but we prefer the following cut-and-paste construction, which permits for later use the construction of Euclidean cones with an arbitrary angle in $]0, \infty[$. We start with a disk centered at the origin O of the Euclidean plane, and we take in that disk a region $\mathcal{R}$ situated between two radii OA and OB making an angle α (the region $\mathcal{R}$ will be a "fundamental domain" for the cone; see Figure 2.8). We then identify the two radii OA and OB by a length-preserving map. The quotient space of the region $\mathcal{R}$ by this identification is a Euclidean cone with vertex angle α. Now let us consider a third radius OC in the fundamental domain making equal angles with the two radii OA and OB. Any point on the image of the radius OA (or, equivalently, on the image of the radius OB) in the quotient space X can be joined by two distinct geodesics to a point on the image of the radius OC except if this point is the image of O.

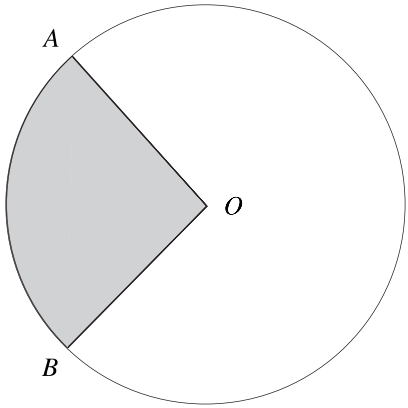


Figure 2.8. The shaded region is a fundamental domain for the Euclidean cone of Example 2.4.16 (ii).

2.5 Geodesic convexity

Definition 2.5.1 (Geodesically convex subspace). Let X be a uniquely geodesic space. A subspace A of X is said to be *geodesically convex* if for every x and y in A, the geodesic segment $[x, y]$ is contained in A.

It is clear that if A is a geodesically convex subset of X and if $f : X \to X$ is an isometry, then $f(A)$ is also a geodesically convex subset.

In hyperbolic space $\mathbb{H}^n$, open and closed balls are geodesically convex subsets. We note that such sets are also geodesically convex with respect to the Euclidean metric, in either the upper half-plane model $\mathbb{H}^n$ or the ball model B^n. However, it is easy to exhibit examples of subsets in these two models that are geodesically convex with respect to the hyperbolic metric but not with respect to the Euclidean metric, and others that are geodesically convex with respect to the Euclidean metric but not with respect to the hyperbolic metric.

Let us note that for subsets of metric spaces that are not uniquely geodesic, there is more than one notion of geodesic convexity which could be useful, and none of these notions is more natural than the others. Indeed, we can ask that for any pair of points in the subset, *at least one* geodesic segment joining them is contained in the subset, or we can ask that for any pair of points in the given subset, *any* geodesic segment joining them is contained in the subset. For uniquely geodesic spaces, these two notions of convexity coincide, which makes the definition of geodesic convexity in such a metric space natural, and the convexity results simpler (or at least simpler to state).

We now study some properties of geodesic convexity in uniquely geodesic metric spaces.

Proposition 2.5.2 (Intersection and union). *Let X be a uniquely geodesic space. Then the intersection of any family of geodesically convex subsets of X is geodesically convex and the union of any increasing family of geodesically convex subsets is geodesically convex.*

Proof. The proof is clear. □

Proposition 2.5.3 (Closure). *Let X be a proper uniquely geodesic space and let A be a geodesically convex subset of X. Then its closure $\bar{A}$ is geodesically convex.*

Proof. Given x and y in $\bar{A}$, we take two sequences $(x_n)_{n\geq 0}$ and $(y_n)_{n\geq 0}$ in A converging to x and y respectively and for every $n \geq 0$ we let $\gamma_n : [0, 1] \to X$ be an affinely reparametrized geodesic joining x_n to y_n. By Ascoli's Theorem (Theorem 1.4.9), the sequence $(\gamma_n)_{n\geq 0}$ has a subsequence $(\gamma_{n_i})_{i\geq 0}$ that converges to a path $\gamma : [0, 1] \to X$ joining x and y and by an easy limiting argument, γ is an affinely reparametrized geodesic. For all t in $[0, 1]$, $\gamma(t)$, being the limit of the sequence $(\gamma_{n_i}(t))$, is in $\bar{A}$

The image of γ is the unique geodesic segment in X joining x and y. This proves that $\bar{A}$ is geodesically convex. □

It is not true that in an arbitrary uniquely geodesic space, the interior of a geodesically convex subset is geodesically convex. In fact, the interior of a geodesically convex subset need not even be connected, as we can see by taking a simplicial metric tree T that contains a vertex v of valency ≥ 3 and taking as a subset A of T a geodesic segment containing v in its interior. Then the interior of A does not contain v, therefore it is not connected.

Definition 2.5.4 (Geodesic convex hull). Let X be a uniquely geodesic space and let A be a subset of X. Then, the *geodesic convex hull of* A is the intersection of all the geodesically convex subsets of X that contain A.

Since the space X is geodesically convex, the geodesic convex hull of any nonempty subset A of X exists and is nonempty. Furthermore, as an intersection of geodesically convex subsets, the geodesic convex hull is geodesically convex.

The following proposition gives a step-by-step construction of the geodesic convex hull:

Proposition 2.5.5. *Let X be a uniquely geodesic space and let A be a subset of X. We set $C_0(A) = A$ and for every integer $n \geq 0$, we let $C_{n+1}(A)$ be the union of all the geodesic segments in X that join pairs of points in $C_n(A)$. Then, the geodesic convex hull $C(A)$ of A is given by*

$$C(A) = \bigcup_{n\geq 0} C_n(A).$$

Proof. The union $\bigcup_{n\geq 0} C_n(A)$ is an increasing union. Using the fact that the convex hull $C(A)$ is geodesically convex, it is easy to see by induction that for each $n \geq 0$, the set $C_n(A)$ is contained in $C(A)$. Thus, we have $\bigcup_{n\geq 0} C_n(A) \subset C(A)$. Conversely, if x and y are two points in $\bigcup_{n\geq 0} C_n(A)$, then they belong to a set $C_n(A)$ for some integer $n \geq 0$. Hence, the segment $[x, y]$ is contained in $C_{n+1}(A)$ and therefore in $\bigcup_{n\geq 0} C_n(A)$. Thus, the set $\bigcup_{n\geq 0} C_n(A)$ is geodesically convex and therefore it contains $C(A)$. This proves that $C(A) = \bigcup_{n\geq 0} C_n(A)$. □

We note that this construction is analogous to a classical construction of the affine convex hull of a subset A of a vector space, and that (in the context of vector spaces) the least element k in $\mathbb{N} \cup \infty$ satisfying

$$C(X) = \bigcup_{0\leq n\leq k} C_n(X)$$

is called the *Brunn number* of A (see Definition 5.1.24 below).

We close this section with some considerations on strict convexity in uniquely geodesic metric spaces.

Definition 2.5.6 (Strictly geodesically convex subset). Let X be a uniquely geodesic space. We say that a subset A of X is *strictly geodesically convex* if for every x and y in A, any point on the geodesic segment $[x, y]$ that is distinct from x and y is contained in the interior of A.

In particular, if A contains at least two elements, then the interior of A has to be nonempty. Any open geodesically convex subset of X is strictly geodesically convex.

It is easy to produce examples of families of strictly geodesically convex subsets whose intersection is not strictly geodesically convex. For instance, take the family of open ϵ-neighborhoods of a closed square in $\mathbb{E}^2$; each element in this family is an open geodesically convex set, therefore it is strictly geodesically convex. But the intersection of this family is the closed square, that is geodesically convex but not strictly. In fact, the same argument works if we take, instead of the open ϵ-neighborhoods of a closed square, closed ϵ-neighborhoods of a closed square. These closed ϵ-neighborhoods are also strictly geodesically convex.

2.6 Menger convexity

The following notion was introduced by Menger in [103].

Definition 2.6.1 (Menger convexity).[5] A metric space X is said to be *Menger convex*, or *convex in the sense of Menger*, if for all distinct points x and y in X, there exists a

[5] Menger calls such a space a "convex space" and Busemann in [28] calls it *M-convex*.

point z that lies between them. We recall from Definition 2.11 that this means that z is distinct from x and from y and that we have

$$|x + z| + |z + y| = |x + y|.$$

Theorem 2.6.2 (Menger convexity and geodesic metric). *Let X a proper metric space. The four following properties are equivalent:*

(i) *the space X is convex in the sense of Menger;*

(ii) *for all x and y in X, there exists a point z in X such that*

$$|x - z| = |y - z| = (1/2)|x - y|;$$

(iii) *for all x and y in X and for all positive real numbers d_1 and d_2 satisfying $d_1 + d_2 = |x - y|$, there exists a point z in X such that $|x - z| = d_1$ and $|y - z| = d_2$;*

(iv) *X is geodesic.*

Proof. Implications (ii) ⇒ (i) and (iii) ⇒ (ii) are clear. We prove (iv) ⇒ (iii), (ii) ⇒ (iv) and (i) ⇒ (ii) in that order.

Suppose that Condition (iv) is satisfied. For every t in $[a, b]$, let $\gamma_t = \gamma_{|[a,t]}$. Since γ is parametrized by arclength, we have $L(\gamma_t) = t - a$. In particular, the map $t \mapsto L(\gamma_t)$ is continuous. Its value is 0 for $t = a$ and $|x - y|$ for $t = b$. Thus, if d_1 and d_2 are as in (iii), then, by the mean value theorem, there exists c in $[a, b]$ such that $L(\gamma_c) = d_1$. By the additivity of length (Proposition 1.1.12), we obtain $L(\gamma_{|[c,b]}) = d_2$. Now let $z = \gamma(c)$. We have $|x - z| \leq L(\gamma_c) = d_1$ and $|y - z| \leq L(\gamma_{|[c,b]}) = d_2$. The two inequalities $|x - z| \leq d_1$ and $|y - z| \leq d_2$, combined with $|x - z| + |y - z| = d_1 + d_2$, imply $|x - z| = d_1$ and $|y - z| = d_2$. Thus, we have (iv) ⇒ (iii).

Now, we prove (ii) ⇒ (iv). Suppose that condition (ii) is satisfied. Let x_0 and x_1 be two distinct points in X, let $\alpha = |x_0 - x_1|$ and let us define a geodesic path $\gamma\colon [0, \alpha] \to X$ satisfying $\gamma(0) = x_0$ and $\gamma(\alpha) = x_1$. Let D be the subset of $[0, \alpha]$ consisting of the points of the form $k\alpha/2^n$ where k and n are natural numbers satisfying $k \leq 2^n$.

We first define the values of γ on the elements in D. We do this by applying infinitely many times Condition (ii). We start by applying this condition to the points x_0 and x_1. We obtain a point $x_{1/2}$ in X satisfying

$$|x_0 - x_{1/2}| = |x_1 - x_{1/2}| = (1/2)|x_0 - x_1|,$$

and we define $\gamma(\alpha/2) = x_{1/2}$. Then, applying the same condition to the pair of points x_0 and $x_{1/2}$ and to the pair of points $x_{1/2}$ and x_1, we obtain two points $x_{1/4}$ and $x_{3/4}$ in X satisfying respectively

$$|x_0 - x_{1/4}| = |x_{1/2} - x_{1/4}| = (1/2)|x_0 - x_{1/2}|$$

and

$$|x_{1/2} - x_{3/4}| = |x_1 - x_{3/4}| = (1/2)|x_{1/2} - x_1|,$$

and we set $\gamma(\alpha/4) = x_{1/4}$ and $\gamma(3\alpha/4) = x_{3/4}$.

Now it is clear how, continuing in the same manner, we can define γ on the whole subset D: for each $n \geq 1$, at step n, we choose n points in X, each of which is the midpoint of a pair of points obtained at the previous step, and we define γ on these points. Thus, after n steps, the map γ is defined on the points in $[0, \alpha]$ that are of the form $k/2^n$ with $0 \leq k \leq 2^n$. This defines γ on D.

Now we must prove that we can extend this map on D to a map whose domain is $[0, \alpha]$, and that the extended map (which we also call γ) is a geodesic path.

We claim that for every pair of distinct real numbers t_1 and t_2 in D, we have $|\gamma(t_1) - \gamma(t_2)| = |t_1 - t_2|$. Indeed, let us write t_1 and t_2 as $k_1\alpha/2^n$ and $k_2\alpha/2^n$ respectively, for some nonpositive integer n, and with $0 \leq k_1 < k_2 \leq 2^n$. Then we have, by construction,

$$t_1 = \sum_{j=0}^{k_1-1} \frac{(j+1-j)\alpha}{2^n} = \sum_{j=0}^{k_1-1} |x_j - x_{j+1}| \geq |x_0 - x_{k_1}|,$$

$$t_2 - t_1 = \sum_{j=k_1}^{k_2-1} \frac{(j+1-j)\alpha}{2^n} = \sum_{j=k_1}^{k_2-1} |x_j - x_{j+1}| \geq |x_{k_1} - x_{k_2}|$$

and

$$\alpha - t_2 = \sum_{j=k_2}^{2^n-1} \frac{(j+1-j)\alpha}{2^n} = \sum_{j=k_2}^{2^n-1} |x_j - x_{j+1}| \geq |x_{k_2} - x_1|.$$

The sum of the left hand sides of the three formulas is equal to α, which is equal to $|x_0 - x_1|$, while the sum of the right hand sides is equal to $|x_0 - x_{k_1}| + |x_{k_1} - x_{k_2}| + |x_{k_2} - x_1|$, which is $\geq |x_0 - x_1|$. We conclude that the three inequalities are equalities. In particular, the second inequality (being an equality) implies

$$|t_1 - t_2| = |x_{k_1} - x_{k_2}| = |\gamma(t_1) - \gamma(t_2)|,$$

which proves the claim.

Now, let t be an arbitrary real number in $[0, \alpha]$. Since D is dense in $[0, \alpha]$, there exists a sequence of points $(t_n)_{n\geq 0}$ in D such that $t_n \to t$ as $n \to \infty$. The sequence $\big(\gamma(t_n)\big)_{n\geq 0}$ is a Cauchy sequence in X. Indeed, we have, for all i and $j \geq 1$, $|\gamma(t_i) - \gamma(t_j)| = |t_i - t_j|$, that tends to 0 as i and j tend to infinity. Furthermore, the points of the sequence $\big(\gamma(t_n)\big)_{n\geq 0}$ are contained in a closed ball in X, and since X is proper, such a ball is compact and therefore complete. Thus, the sequence $\big(\gamma(t_n)\big)_{n\geq 0}$ is convergent. Now we set $\gamma(t) = \lim_{n\to\infty} \gamma(t_n)$, and we prove that the point $\gamma(t)$ does not depend on the choice of the sequence $(t_n)_{n\geq 0}$. If $(t'_n)_{n\geq 0}$ is another sequence in D that converges to t, then the sequence $(t''_n)_{n\geq 0}$ defined by $t''_{2n} = t_n$ and $t''_{2n+1} = t'_n$ also converges to t, the sequence of images $\gamma(t''_{2n})_{n\geq 0}$ is a Cauchy sequence which is therefore convergent, and the two sequences $\gamma(t_{2n})_{n\geq 0}$ and $\gamma(t'_{2n})_{n\geq 0}$ have the same limit.

Thus, we have a map $\gamma : [0, \alpha] \to X$ that satisfies $\gamma(0) = x$ and $\gamma(\alpha) = y$. Let us show that this map is distance-preserving. For t and t' in $[0, \alpha]$, let $(t_n)_{n\geq 0}$ and $(t'_n)_{n\geq 0}$ be two sequences in D satisfying $t_n \to t$ and $t'_n \to t'$ as $n \to \infty$. We have $\gamma(t) = \lim_{n\to\infty} \gamma(t_n)$ and $\gamma(t') = \lim_{n\to\infty} \gamma(t'_n)$, which gives

$$|\gamma(t) - \gamma(t')| = \lim_{n\to\infty} |\gamma(t_n) - \gamma(t'_n)| = \lim_{n\to\infty} |t_n - t'_n|.$$

Thus, γ is distance-preserving. This completes the proof of (ii) ⇒ (iv).

Finally, let us prove (i) ⇒ (ii). Suppose that Condition (i) is satisfied, let x and y be two points in X and let us consider the set

$$B = \{z \in X \text{ such that } |x - z| + |z - y| = |x - y|\}.$$

By continuity of the distance function, B is a closed subset of X. This set is also bounded, since $|x - z| \leq |x - y|$ implies that B is contained in the ball of center x and radius $|x - y|$. Since X is proper, B is compact.

The map from B to $\mathbb{R}$ defined by $z \mapsto \min\{|z - x|, |z - y|\}$ is continuous, and therefore it attains a maximum. Let β be a maximum value, let m be a point in B where this maximum is attained and let us show that we have

$$\beta = |m - x| = |m - y| = (1/2)|x - y|.$$

By definition, we have

$$\beta = \min\{|m - x|, |m - y|\}$$

and

$$|m - x| + |m - y| = |x - y|,$$

which implies

$$\beta \leq (1/2)|x - y|.$$

To show that $|m - x| = (1/2)|x - y|$, we reason by contradiction. Suppose that we had $\beta = |m - x| < |m - y|$. Then we would have $\beta < (1/2)|x - y|$. By Condition (i), there would exist z in X, that is distinct from m and y, such that $|m - z| + |z - y| = |m - y|$. By Proposition 2.2.13, we would have $|x - z| + |z - y| = |x - y|$ and $|x - m| + |m - z| = |x - z|$.

We claim that $|z - y| \leq \beta$. Indeed, suppose that we had $|z - y| > \beta$. As $|x - z| +$ $|z - y|$, this would contradict the fact that $|z - y|$ is a minimum of $\{|z - x|, |z - y|\}$. Thus, we obtain

$$|m - z| = |x - z| - |x - m| = |x - y| - |z - y| - |x - m|,$$

which implies

$$|m - z| \geq |x - y| - 2\beta > 0.$$

The set

$$Z = \{z \in X \text{ such that } |m - z| + |z - y| = |m - y|\} \cup \{y\}$$

is closed and bounded in X, and therefore it is compact. The map

$$z \mapsto d_m(z) = |m - z|$$

defined on Z attains its minimum at some point z_0 of Z. By Property (i), we find a point z' in X that is distinct from m and from z_0 and that satisfies $|m-z'|+|z'-z_0| = |m-z_0|$. Thus, we have $|m - z'| < |m - z_0|$. Applying again Proposition 2.2.13, we obtain $|m - z'| + |z' - y| = |m - y|$. This contradicts the fact that z_0 is a point where the minimum of d_m is attained. This shows that $|m-x| = |x-y|$, which proves (i) ⇒ (ii). This completes the proof of Theorem 2.6.2. □

We note (following Busemann) that Condition (iii) in Theorem 2.6.2 does not imply Condition (iv) in case the space X is not complete, as one can see by taking X to be the set of rational numbers equipped with its usual metric.

Let us note a few corollaries of Theorem 2.6.2.

The first corollary says in some sense that in a geodesic metric space, we can approach any geodesic segment by a finite sequence of points forming a "discrete geodesic" whose points are arbitrarily close together.

Corollary 2.6.3. *Let X be a geodesic metric space, let x and y be two arbitrary points in X and let $d_1, \ldots, d_n$ be a sequence of positive real numbers satisfying $d_1 + \cdots + d_n = |x - y|$. Then there exists a sequence $x_0, \ldots, x_n$ of points in X satisfying $x_0 = x$, $x_n = y$, and $|x_i - x_{i+1}| = d_i$ for all $i = 0, \ldots, n - 1$.*

Proof. We apply n times Implication (iv) ⇒ (iii) of Theorem 2.6.2. □

We can use Theorem 2.6.2 to study products of metric spaces.

We recall that if X_1 and X_2 are two metric spaces, there is no privileged metric on the product $X_1 \times X_2$,[6] and in fact, there are several metrics on this product, none of which imposes itself on the others, since each such metric is useful in some particular context. For instance, for every real number $p \in [1, \infty[$, one defines a metric d_p on $X_1 \times X_2$ by setting, for all (x_1, x_2) and $(y_1, y_2) \in X_1 \times X_2$,

$$d_p\big((x_1, x_2), (y_1, y_2)\big) = \Big(\big(d_{X_1}(x_1, y_1)\big)^p + \big(d_{X_2}(x_2, y_2)\big)^p\Big)^{\frac{1}{p}}.$$

[6] Let us note however that there is a natural topology on the product space. It is defined as the coarser topology for which the canonical projections on the two factors are continuous. All the metrics that we consider here induce this topology on the product.

Another useful metric on $X_1 \times X_2$ is the metric d_∞ defined by

$$d_\infty\big((x_1, x_2), (y_1, y_2)\big) = \max\{d_X(x_1, y_1), d_Y(x_2, y_2)\}.$$

For $p = 2$, the metric d_p is the "Euclidean metric" on $X_1 \times X_2$. It is called so because in the case where X_1 and X_2 are respectively the Euclidean spaces $\mathbb{E}^n$ and $\mathbb{E}^m$, then the product space $\mathbb{E}^n \times \mathbb{E}^m$, equipped with the metric d_2, is isometric to the Euclidean space $\mathbb{E}^{n+m}$.

The proof of the fact that the map d_p, for $p \in]1, \infty[$, defines a metric on $X_1 \times X_2$ is based on the following result of Minkowski which is contained in [110]:

Proposition 2.6.4 (Minkowski's inequality). *For every integer $n \geq 1$, for any pair of vectors $(a_1, \dots a_n)$ and $(b_1, \dots b_n)$ of $\mathbb{R}^n$ whose coordinates are nonnegative and for every real number p in $[1, \infty[$, we have*

$$\Big(\sum_{i=1}^{n}(a_i + b_i)^p\Big)^{\frac{1}{p}} \leq \Big(\sum_{i=1}^{n} a_i^p\Big)^{\frac{1}{p}} + \Big(\sum_{i=1}^{n} b_i^p\Big)^{\frac{1}{p}}.$$

Furthermore, for any $p > 1$, this large inequality is an equality if and only if the vectors $(a_1, \dots a_n)$ and $(b_1, \dots b_n)$ are collinear (the proportionality constant can be equal to 0). Of course, for $p = 1$, the large inequality is always an equality.

For a proof of this result, we refer to [63] p. 30. □

Proposition 2.6.5 (Product metric). *For any p in $[0, \infty[\cup\{\infty\}$, the map d_p defines a metric on $X_1 \times X_2$.*

Proof. We prove the proposition in the case where p is finite. The case $p = \infty$ is simpler, and it can be dealt with using the same outline of proof.

It is clear from the definition that for all x and y in $X_1 \times X_2$, we have $d_p(x, y) \geq 0$, $d_p(x, y) = d_p(y, x)$, and that $d_p(x, y) = 0 \iff x = y$. It remains to prove that d_p satisfies the triangle inequality. For that, we use Minkowski's inequality of Proposition 2.6.4. Let $x = (x_1, x_2)$, $y = (y_1, y_2)$, and $z = (z_1, z_2)$ be three points in $X_1 \times X_2$. Applying Minkowski's inequality with $n = 2$, $a_1 = |x_1 - y_1|_{X_1}$, $a_2 = |x_2 - y_2|_{X_2}$, $b_1 = |y_1 - z_1|_{X_1}$ and $b_2 = |y_2 - z_2|_{X_2}$, we obtain

$$\begin{aligned} d_p(x, z) &= (|x_1 - z_1|_{X_1}^p + |x_2 - z_2|_{X_2}^p)^{\frac{1}{p}} \\ &\leq \big((|x_1 - y_1|_{X_1} + |y_1 - z_1|_{X_2})^p \\ &\quad + (|x_2 - y_2|_{X_1} + |y_2 - z_2|_{X_2})^p\big)^{\frac{1}{p}} \\ &= \big((a_1 + b_1)^p + (a_2 + b_2)^p\big)^{\frac{1}{p}} \end{aligned}$$

$$\begin{aligned}
&\leq (a_1^p + a_2^p)^{\frac{1}{p}} + (b_1^p + b_2^p)^{\frac{1}{p}} \quad \text{(by Minkowski's inequality)}\\
&= (|x_1 - y_1|_{X_1}^p + |x_2 - y_2|_{X_2}^p)^{\frac{1}{p}}\\
&\quad + (|y_1 - z_1|_{X_1}^p + |y_2 - z_2|_{X_2}^p)^{\frac{1}{p}}\\
&= d_p(x, y) + d_p(y, z).
\end{aligned}$$

□

We note that for all p in $[0, \infty[\cup\{\infty\}$, the projection of $X_1 \times X_2$ (equipped with the metric d_p) on each of its two factors is a 1-Lipschitz map, and therefore it is continuous. It is also easy to see that a sequence $(x_n)_{n\geq 0}$ in $X_1 \times X_2$ converges to a point x in that space if and only if $d_{X_1}(y_n, y) \to 0$ and $d_{X_2}(z_n, z) \to 0$, where y and z are the projections of x on X_1 and X_2 respectively and where for each $n \geq 0$, y_n (respectively z_n) is the projection of x_n on X_1 (respectively X_2).

Proposition 2.6.6. *If X_1 and X_2 are two proper (respectively geodesic) spaces, then the product $X_1 \times X_2$, equipped with any of the metrics d_p for $p \in [0, \infty[\cup\{\infty\}$, is proper (respectively geodesic).*

Proof. We give the proof for the metrics d_p for $p \in [1, \infty[$; the proof for d_∞ can be done in the same manner.

It is easy to see that if X_1 and X_2 are proper, then $X_1 \times X_2$ is also proper. Indeed, let K be a closed and bounded subset of $X_1 \times X_2$ and let K_1 and K_2 be its projections on X_1 and X_2 respectively. Since K_1 and K_2 are bounded, their closures $\overline{K}_1$ and $\overline{K}_2$ are compact, since X_1 and X_2 are proper. The product $\overline{K}_1 \times \overline{K}_2$ is a compact subset of $X_1 \times X_2$, and the set K, which is closed in this compact set, is itself compact. This shows that $X_1 \times X_2$ is proper.

Now let us prove that if X_1 and X_2 are geodesic spaces, then $X_1 \times X_2$ is also geodesic. We use the equivalence (ii)⇔(iv) of Theorem 2.6.2. Let $x = (x_1, x_2)$ and $y = (y_1, y_2)$ be two points in $X_1 \times X_2$. Since X_1 and X_2 are geodesic spaces, then, by Theorem 2.6.2, there exists a point z_1 in X_1 and a point z_2 in X_2 satisfying

$$|x_1 - z_1| = |y_1 - z_1| = \frac{1}{2}|x_1 - y_1|$$

and

$$|x_2 - z_2| = |y_2 - z_2| = \frac{1}{2}|x_2 - y_2|.$$

Thus, we have

$$\begin{aligned}
d_p\big((z_1, z_2), (x_1, x_2)\big)^p &= \big(d_{X_1}(z_1, x_1)\big)^p + \big(d_{X_2}(z_2, y_2)\big)^p\\
&= \frac{1}{2^p}\big(d_{X_1}(x_1, y_1)\big)^p + \frac{1}{2^p}\big(d_{X_2}(x_2, y_2)\big)^p\\
&= \frac{1}{2^p} d_p\big((x_1, x_2), (y_1, y_2)\big)^p,
\end{aligned}$$

which gives

$$d_p\big((z_1, z_2), (x_1, x_2)\big) = \frac{1}{2} d_p\big((x_1, x_2), (y_1, y_2)\big),$$

that is,

$$d_p(z, x) = \frac{1}{2} d_p(x, y),$$

where z is the point (z_1, z_2) of $X_1 \times X_2$. In the same way (or by symmetry), we obtain

$$d_p(z, y) = \frac{1}{2} d_p(x, y).$$

Therefore, Condition (ii) of Theorem 2.6.2 is satisfied for the metric d_p on $X_1 \times X_2$. This shows that the metric space $(X_1 \times X_2, d_p)$ is geodesic. □

We have presented the proof of Proposition 2.6.6 as an application of Theorem 2.6.2, but one can also obtain this result as a consequence of the following proposition:

Proposition 2.6.7. *Let X_1 and X_2 be two geodesic spaces and let us equip the product space $X = X_1 \times X_2$ with one of the metrics d_p, for some p in $[1, \infty[\cup \{\infty\}$. Let x_1 and y_1 be two arbitrary points in X_1, let x_2 and y_2 be two arbitrary points in X_2, let $\gamma_1 \colon [0, 1] \to X_1$ be an affinely reparametrized geodesic joining x_1 and y_1 and let $\gamma_2 \colon [0, 1] \to X_2$ be an affinely reparametrized geodesic joining x_2 and y_2. Then the product path $\gamma \colon [0, 1] \to X$, defined by $\gamma(t) = (\gamma_1(t), \gamma_2(t))$ is an affinely reparametrized geodesic joining (x_1, y_1) and (x_2, y_2). Furthermore, in the case where $p \in [1, \infty[$, we have*

$$L(\gamma) = \big(L(\gamma_1)^p + \big(L(\gamma_2)^p\big)^{\frac{1}{p}},$$

and in the case $p = \infty$, we have

$$L(\gamma) = \max\big(L(\gamma_1), L(\gamma_2)\big).$$

Proof. The map γ is continuous and it is a path joining the points (x_1, y_1) and (x_2, y_2) in X. Let us show that this map is an affinely reparametrized geodesic. We first consider the case where $p \in [1, \infty[$. For all u and v in $[0, 1]$, we have

$$\begin{aligned} d_p\big(\gamma(u), \gamma(v)\big) &= \big(|\gamma_1(u) - \gamma_1(v)|^p + |\gamma_2(u) - \gamma_2(v)|^p\big)^{\frac{1}{p}} \\ &= \big(L(\gamma_1)^p |u - v|^p + L(\gamma_2)^p |u - v|^p\big)^{\frac{1}{p}} \\ &= \big(L(\gamma_1)^p + L(\gamma_2)^p\big)^{\frac{1}{p}} |u - v|. \end{aligned}$$

By Proposition 2.2.10, γ is therefore an affinely reparametrized geodesic whose length is $\big(L(\gamma_1)^p + \big(L(\gamma_2)^p\big)^{\frac{1}{p}}$.

In the case where $p = \infty$, we have

$$\begin{aligned} d_\infty\big(\gamma(u), \gamma(v)\big) &= \max\big(|\gamma_1(u) - \gamma_1(v)|, |\gamma_2(u) - \gamma_2(v)|\big) \\ &= \max\big(L(\gamma_1)|u - v|, L(\gamma_2)|u - v|\big) \\ &= \max\big(L(\gamma_1), L(\gamma_2)\big)|u - v|. \end{aligned}$$

We again apply Proposition 2.2.10, which shows that γ is an affinely reparametrized geodesic of length $\max\big(L(\gamma_1), L(\gamma_2)\big)$. This completes the proof of Proposition 2.6.7. □

Notes on Chapter 2

The theorems of Hopf–Rinow. Theorem 2.4.6, which states that in a proper length space, any two points can be joined by a geodesic segment, is also contained in Elie Cartan's book [36] p. 260, for the special case of Riemannian manifolds. Theorems 2.1.5 and 2.4.6 are due to H. Hopf and W. Rinow [73] in the case where the space X is a surface equipped with a Riemannian metric, with some local uniqueness condition on local geodesics. S. Myers, in [115], showed that Theorem 2.4.6 is valid (with almost the same proof) for any manifold of dimension ≥ 2. In [41], S. Cohn-Vossen gave general versions of the theorems of Hopf–Rinow (without the local uniqueness condition) that are valid in arbitrary metric spaces and even in more general spaces (spaces that satisfy the axioms of a metric space except the axiom that says that the distance function is symmetric). For several useful versions of the result of Hopf–Rinow, we refer the reader to the book by Ballmann [7] p. 29. The arguments used in the proof of (ii) $\Rightarrow$ (iv) are classical (see [36] p. 360). Implication (i) $\Rightarrow$ (iv) is due to Menger [103].

Existence and uniqueness of geodesics in Riemannian manifolds. In [36], E. Cartan proves that for any two points in a Riemannian manifold that is proper (Cartan calls such a space a *normal* Riemannian space), there exists at least one geodesic joining them. Thus, a proper Riemannian manifold is a geodesic metric space. Cartan proves also that for any point x in such a space, there is a positive real number r such that for any y in the closed ball $B(x, r)$, there exists a unique geodesic joining x to y, and that the image of such a geodesic is contained in $B(x, r)$ (see [36], Note IV, Theorem I, p. 356 and Theorem III, p. 360).

Busemann's G-spaces. Busemann's theory of geodesics in metric spaces is developed in [28] in the setting of metric spaces which he calls G-spaces. The definition is as follows: a G-space in the sense of Busemann is a metric space X that satisfies the following properties:

(i) X is a proper metric space;

(ii) for every x and y in X, there exists a point z that lies between x and y;

(iii) for every point in X, there is an open ball B of positive radius centered at this point and for every x and y in B, there exists a point z in B such that y lies between x and z;

(iv) for every quadruple of points x, y, z_1 and z_2 in X satisfying $|x-y|+|y-z_1| = |x-z_1|$ and $|x-y|+|y-z_2| = |x-z_2|$, if $|y-z_1| = |y-z_2|$, then we have $z_1 = z_2$.

Property (ii) is equivalent to a property of existence of geodesics (cf. Theorem 2.6.2), Property (iii) is a property of local existence of prolongation of geodesics and Property (iv) is a property of uniqueness of the prolongation.

Chapter 3

Maps between metric spaces

Introduction

In this chapter, we study properties of maps between metric spaces. Of course, the maps that most naturally come to mind in this context are isometries (that is, distance-preserving surjective maps). We shall study isometries in more detail in Chapter 11 below. Besides isometries, there are several classes of maps between metric spaces that are important, and we mention the following classes: distance-preserving maps, length-preserving maps[1] (which may be considered as the analogs of distance-preserving maps in the setting of length spaces), Lipschitz maps, K-Lipschitz maps for some fixed K, bi-Lipschitz homeomorphisms, Hölder maps, distance-non-increasing maps, length-non-increasing maps, contractions, distance-decreasing maps, local homeomorphisms, locally K-Lipschitz maps, local isometries, quasi-conformal maps, quasi-isometries, covering maps that are local isometries, and there are many others. Each such class of maps is important in some particular context. In this chapter, we study a few general properties of maps belonging to some of these classes. We shall use some of the results in later chapters.

We mention by the way, since we are talking about the relation between maps and metrics, that there are several interesting instances where a metric on a given space is defined in terms of properties of self-maps of that space. For example, the Poincaré metric of the unit disk D in the complex plane, was defined by Poincaré in the course of his study of automorphic maps of that disk (see [122]). This metric is the unique metric that is invariant by the conformal self-maps of D. Likewise, we already saw that the Carathéodory and the Kobayashi pseudo-metrics of a complex manifold M are defined by their behaviour with respect to some maps from M to D and from D to M respectively. In fact, there are beautiful characterizations of these two pseudo-metrics as extremal metrics with respect to a property of non-expanding maps: the Carathéodory (respectively the Kobayashi) pseudo-metric of a complex manifold M is the smallest (respectively the largest) pseudo-metric on M such that any holomorphic map from the unit disk D, equipped with its Poincaré metric, to M (respectively any holomorphic maps from M to the unit disk) is non-expanding.[2]

This chapter can be divided into two parts. The first part concerns maps defined by global properties (K-Lipschitz maps, non-expanding maps, contractions and so

[1] Busemann sometimes calls these maps "equilong maps", a word already used by Elie Cartan (see [35]).

[2] Of course, we are talking here about pseudo-metrics, but in most cases of interests these pseudo-metrics are genuine metrics.

on). The second part deals with maps defined by local properties (local isometries and coverings).

The outline of this chapter is as follows.

In Section 1, we study the classes of K-Lipschitz maps and K-length-non-increasing maps. The latter are, in some sense, analogs of K-Lipschitz maps in the setting of length spaces.

In Section 2, we consider the class of non-expanding (that is, 1-Lipschitz) maps, and in particular the subclasses of contractions and of distance-decreasing maps.

In Section 3, we study non-decreasing maps. We present in particular a result of Freudenthal and Hurewicz that states that if X is a compact metric space and $f: X \to X$ is a distance non-decreasing map, then f is an isometry (Theorem 3.3.4).

In Section 4, we study local isometries. Local isometries are non-expanding (Corollary 3.4.5). If Y is a length space and $f: X \to Y$ a local homeomorphism, then we give conditions for the existence and uniqueness of a metric on X with respect to which f is a local isometry (Proposition 3.4.7). We then establish existence and uniqueness results for the lifts of geodesics and of local geodesics by local isometries (Proposition 3.4.11 and 3.4.12).

In Section 5, we study covering maps that are local isometries. In particular, we give sufficient conditions under which a local isometry between length spaces is a covering map (Theorem 3.5.4).

3.1 K-Lipschitz maps and K-length-non-increasing maps

We start by recalling the following

Definition 3.1.1 (K-Lipschitz map). Let X and Y be two metric spaces and let K be a nonnegative real number. A map $f: X \to Y$ is said to be *K-Lipschitz* if for all x and y in X we have $|f(x) - f(y)| \leq K|x - y|$. The map f is said to be *Lipschitz* if it is K-Lipschitz for some K.

The composition of two Lipschitz maps is Lipschitz. Smooth maps between compact Riemannian manifolds are examples of Lipschitz maps. In any metric space X, for every $x_0 \in X$, the map defined by $x \mapsto |x - x_0|$ is 1-Lipschitz (a consequence of the triangle inequality). Composing this map with Lipschitz maps of the real line, we obtain a lot of examples of real-valued Lipschitz maps defined on any metric space.

It is easy to see that a Lipschitz map is uniformly continuous. Therefore we have the following

Proposition 3.1.2. *The image of a Cauchy sequence by a K-Lipschitz map is a Cauchy sequence.* □

Now we recall the definition of the displacement function. This is certainly one of the most important functions associated to a self-map of a metric space and we shall use it thoroughly in Chapter 11 below, in the study of isometries. Right now, we mention it an example of a Lipschitz map.

Definition 3.1.3 (Displacement function). Let X be a metric space and let $f: X \to X$ be a map. The *displacement function* of f is the map $d_f: X \to [0, \infty[$ defined by

$$d_f(x) = |x - f(x)|$$

for every x in X.

The displacement function will be useful in the study of distance-decreasing maps (Section 3 below), in the study of distance-non-decreasing maps (Section 4 below) and in the classification of the isometries of metric spaces (see Definition 11.1.6, p. 244).

Proposition 3.1.4. *Let X be a metric space and let $f: X \to X$ be a K-Lipschitz map. Then the displacement function of f is a $(K + 1)$-Lipschitz map.*

Proof. Let x and y be two points in X. From the triangle inequality, we obtain

$$\big||x - f(x)| - |y - f(y)|\big| \leq |x - y| + |f(x) - f(y)| \leq (K + 1)|x - y|. \qquad \square$$

The following notion is closely related to the notion of K-Lipschitz map; in some sense, it describes an analogous property in the setting of length spaces.

Definition 3.1.5 (K-length-non-increasing map). Let K be a nonnegative real number and let X and Y be two metric spaces. A map $f: X \to Y$ is said to be *K-length-non-increasing* if for any rectifiable path $\gamma: [a, b] \to X$, we have $L_Y\big(f(\gamma)\big) \leq K L_X(\gamma)$.

Proposition 3.1.6 (K-Lipschitz implies K-length non-increasing). *Let X and Y be metric spaces and let $f: X \to Y$ be a K-Lipschitz map. Then f is K-length-non-increasing.*

Proof. Let $\gamma: [a, b] \to X$ be a rectifiable path. For each subdivision $\sigma = (t_i)_{i=0,\ldots,n}$ of $[a, b]$, we have

$$V_\sigma(f \circ \gamma) = \sum_{i=0}^{n-1} |f \circ \gamma(t_i) - f \circ \gamma(t_{i+1})| \leq K \sum_{i=0}^{n-1} |\gamma(t_i) - \gamma(t_{i+1})| = K V_\sigma(\gamma).$$

Taking the supremum over all subdivisions σ of $[a, b]$, we obtain $L_Y(f(\gamma)) \leq K L_X(\gamma)$. $\square$

Proposition 3.1.6 has the following partial converse:

Proposition 3.1.7 (Continuous and K-length non-increasing implies K-Lipschitz). *Let X be a length space, let Y be a metric space and let $f: X \to Y$ be a continuous map. If f is K-length-non-increasing , then it is K-Lipschitz.*

Proof. For every x and y in X, we have $|f(x) - f(y)|_Y \leq \inf L_Y(\gamma')$, where the infimum is taken over the set of paths γ' in Y joining $f(x)$ to $f(y)$. Thus, $|f(x) - f(y)|_Y \leq \inf L_Y(f(\gamma))$, where the infimum is taken over the paths γ in X joining x to y. (Here, we use the fact that f is continuous.) Since f is K-length-non-increasing, we obtain $|f(x) - f(y)|_Y \leq K \inf L_X(\gamma)$, where the infimum is again taken over the paths γ in X joining x to y. Since X is a length space, the last term is equal to $K|x - y|$. This shows that f is K-Lipschitz. □

We note that a K-length-non-increasing map is not necessarily continuous. For instance, if Y is a discrete metric space (that is, if each point in Y is isolated), then the length of any path in Y is equal to zero and therefore any map from any metric space X into Y is K-length-non-increasing, for all $K \geq 0$. On the other hand, of course, there exist non-continuous maps from X to Y, provided X is not discrete.

3.2 Non-expanding maps

Definition 3.2.1 (Non-expanding map).[3] Let X and Y be two metric spaces. We say that a map $f: X \to Y$ is *distance-non-expanding*, or simply, *non-expanding*, if f is a 1-Lipschitz map, that is, if it satisfies

$$|f(x) - f(y)| \leq |x - y|$$

for all x and y in X.

Important examples of non-expanding maps include isometries, contractions (that we shall study below) and holomorphic maps between complex spaces equipped with their Carathéodory and their Kobayashi pseudo-metrics. We start by presenting these beautiful examples:

Examples 3.2.2 (Non-expanding map).

(i) *The Carathéodory and the Kobayashi pseudo-metrics.* Again, we consider here pseudo-metrics which in general are not metrics; this is justified by the importance of these pseudo-metrics and, of course, by the fact that in some cases they are metrics. If X and Y are two complex manifolds equipped with their Carathéodory (respectively Kobayashi) pseudo-metric (see Example 2.1.3 (ix) and (x)) and if $f: X \to Y$ is a holomorphic map, then it follows easily from the definition that f is non-expanding, *i.e.* it satisfies Definition 3.2.1 where instead of distances we have pseudo-distances.

[3] Busemann, in [29], calls a non-expanding map a "shrinkage".

In particular, if $X \subset Y$ is an inclusion between complex manifolds, then the inclusion map is non-expanding with respect to the Carathéodory (respectively the Kobayashi) pseudo-metrics of X and Y. Another consequence is that the Carathéodory (respectively the Kobayashi) pseudo-metric is invariant under biholomorphic maps.

(ii) *Holomorphic maps of the disk.* If D is the 2-dimensional disk equipped with its hyperbolic metric and if $f : D \to D$ is a holomorphic map, then f is non-expanding. This is a particular case of Example (i) above, since, as we already recalled in Chapter 1, the Carathéodory and the Kobayashi pseudo-distances of the disk both coincide with the hyperbolic metric of that disk. (This amounts to the Schwarz–Pick lemma).

Let us now return to general metric spaces. Busemann made in [29] a detailed study of non-expanding maps. We present here some of his results.

The following lemma will be used several times in the rest of this section. It is due to Busemann.

Lemma 3.2.3. *Let X be a metric space, let $f : X \to X$ be a non-expanding map and let x and y be two points in X. Then,*

(i) *if x lies between y and $f(y)$, then $d_f(x) \leq d_f(y)$, and if equality holds, then $f(y)$ lies between x and $f(x)$ and $|x - y| = |f(x) - f(y)|$;*

(ii) *if y lies between x and $f(y)$, then $d_f(x) \geq d_f(y)$, and if equality holds, then $f(x)$ lies between x and $f(y)$ and $|x - y| = |f(x) - f(y)|$.*

Proof. We start by proving (i). Suppose that x lies between y and $f(y)$. Then we have

$$\begin{aligned} d_f(x) = |x - f(x)| &\leq |x - f(y)| + |f(y) - f(x)| \\ &\leq |x - f(y)| + |y - x| \\ &= |y - f(y)| = d_f(y). \end{aligned}$$

If $d_f(x) = d_f(y)$, then the two large inequalities in the last sequence are equalities. The second of these inequalities gives $|f(y) - f(x)| = |y - x|$, which is > 0 since x lies between y and $f(y)$. We conclude that $f(y)$ is distinct from x and from $f(x)$ and the first inequality (which now is an equality) implies that $f(y)$ lies between x and $f(x)$. This proves (i).

Now we prove (ii). Suppose that y lies between x and $f(y)$. In particular, we have $d_f(y) > 0$, since y is distinct from $f(y)$. Then we have

$$\begin{aligned} d_f(x) = |x - f(x)| &\geq |x - f(y)| - |f(y) - f(x)| \\ &\geq |x - f(y)| - |x - y| \\ &= |y - f(y)| = d_f(y) > 0. \end{aligned}$$

Again, if $d_f(x) = d_f(y)$, then the two large inequalities in the last sequence are equalities. The second inequality gives $|f(y) - f(x)| = |y - x|$ which is > 0 since y

lies between x and $f(y)$. Thus, $f(x)$ is distinct from $f(y)$ and from x, and the first inequality (which now is an equality) implies that $f(x)$ lies between x and $f(y)$. This proves (ii). □

Proposition 3.2.4. *Let X and Y be metric spaces and let $f: X \to Y$ be a non-expanding map. Suppose that x and y are points in X satisfying $|x-y| = |f(x) - f(y)|$. Then f maps any geodesic segment $[x, y]$ isometrically onto a geodesic segment $[f(x), f(y)]$.*

Proof. Let z be a point on $[x, y]$. Then we have

$$\begin{aligned} |x-y| &= |x-z| + |z-y| \\ &\geq |f(x) - f(z)| + |f(z) - f(y)| \\ &\geq |f(x) - f(y)| \\ &= |x-y|. \end{aligned}$$

The two large inequalities in the last sequence are therefore equalities. From the first of these equalities, we obtain $|x-z| = |f(x)-f(z)|$ and $|z-y| = |f(z)-f(y)|$. Now let w be any point on $[z, y]$. By the same argument, we obtain $|z-w| = |f(z)-f(w)|$. Thus, the restriction of f to the geodesic segment $[x, y]$ is an isometry from this segment onto its image. This implies that $f([x, y])$ is a geodesic segment joining $f(x)$ and $f(y)$. □

From Proposition 3.2.4, we deduce the following two corollaries.

Corollary 3.2.5. *Let X be a metric space, let $f: X \to X$ be a non-expanding map, let x and y be two fixed points of f and suppose that there exists a unique geodesic segment $[x, y]$ joining x and y. Then every point on $[x, y]$ is a fixed point of f.*

Proof. Since f sends isometrically the segment $[x, y]$ to itself preserving the orientation, f fixes any point of $[x, y]$. □

Corollary 3.2.6. *Suppose that the space X is uniquely geodesic and let $f: X \to X$ be a non-expanding map. Then the fixed point set of f is a closed convex subset of X.*

Proof. The fact that $Fix(f)$ is closed follows simply from the continuity of f. Corollary 3.2.5 implies that for any x and y in Fix(f), the geodesic segment $[x, y]$ is contained in Fix(f). Thus, Fix(f) is convex. □

The following notion is closely related to the notion of non-expanding maps.

Definition 3.2.7 (Length-non-expanding map). Let X and Y be two metric spaces. A map $f: X \to Y$ is said to be *length-non-expanding* if f is 1-length-non-increasing in the sense of Definition 3.1.5 above, that is, if for every path $\gamma : [a, b] \to X$, we have $L_Y\big(f(\gamma)\big) \leq L_X(\gamma)$.

Proposition 3.2.8. *Let X and Y be two metric spaces. Then,*

(i) *every distance-non-expanding map $f : X \to Y$ is length-non-expanding;*

(ii) *if X is a length space and if $f : X \to Y$ is continuous and length-non-expanding, then f is distance-non-expanding.*

Proof. The two statements are particular cases of Propositions 3.1.6 and 3.1.7 above. □

The following special class of distance-non-expanding maps has been the object of particular attention:

Definition 3.2.9 (Contraction). Let X be a metric space. A map $f : X \to X$ is said to be a *contraction* (of factor K) if f is K-Lipschitz for some real number K in $[0, 1[$.

The most classical result on contractions is certainly the following

Proposition 3.2.10 (Banach). *Let X be a complete metric space and let $f : X \to X$ be a contraction. Then there exists a unique fixed point x for f. Furthermore, for any point y in X, the sequence $(f^n(y))_{n\geq 0}$ converges to x as $n \to \infty$.*

Proof. We start with an arbitrary point x_0 in X and we set, for every integer $n \geq 1$, $x_n = f^n(x_0)$. Let K be the contraction factor of f. Then, for all nonnegative integers m and n satisfying $m \geq n$, we have

$$\begin{aligned}
|x_m - x_n| &\leq |x_m - x_{m-1}| + \cdots + |x_{n+1} - x_n| \\
&\leq K^{m-1}|x_0 - x_1| + \cdots + K^n|x_0 - x_1| \\
&= \frac{K^n}{1-K}|x_1 - x_0|,
\end{aligned}$$

and the last term tends to 0 as n tends to ∞. We deduce that (x_n) is a Cauchy sequence. Since X is complete, this sequence converges. Let x be its limit. Since f is continuous, we obtain, by letting $n \to \infty$ in the equation $x_n = f(x_{n-1})$, the equality $f(x) = x$. This shows that x is a fixed point of f.

Now let us prove that x is the unique fixed point of f. Suppose that some point y in X is also a fixed point of f. Then we would have

$$|x - y| = |f(x) - f(y)| \leq K|x - y| < |x - y|,$$

which is impossible unless $x = y$.

To prove the last statement in the proposition, we consider an arbitrary point y in X and we set, for all $n \geq 0$, $y_n = f^n(y)$. Then we have, for $n \geq 1$,

$$\begin{aligned}|y_n - x| &= |f(y) - f(x)| \\ &\leq K|y_{n-1} - x| \\ &\leq \cdots \\ &\leq K^n|y_0 - x|.\end{aligned}$$

Since $K < 1$, we obtain $|y_n - x| \to 0$ as $n \to \infty$, which shows that $f^n(y)$ converges to x. □

It is well-known that in the case where X is compact, the hypothesis of Proposition 3.2.10 can be weakened, and Proposition 3.2.12 below is also classical. Before stating it, we introduce a class of maps that lies between the class of contractions and the class of non-expanding maps.

Definition 3.2.11 (Distance-decreasing map). Let X be a metric space. A map $f : X \to X$ is said to be *distance-decreasing* if for all distinct points x and y in X we have $|f(x) - f(y)| < |x - y|$.

Proposition 3.2.12. *Let X be a compact metric space and let $f : X \to X$ be a distance-decreasing map. Then f has a unique fixed point x in X. Furthermore, for every y in X, the sequence $(f^n(y))$ converges to x as $n \to \infty$.*

Proof. Consider the displacement function $d_f : X \to [0, \infty[$. Since X is compact, we can find a point x in X where the minimum of d_f is attained. If x were different from $f(x)$, then we would have

$$d_f(f(x)) = |f(x) - f^2(x)| < |x - f(x)| = d_f(x),$$

which contradicts the fact that the minimum of d_f is attained at x. Thus, x is a fixed point of f. The uniqueness of this fixed point follows as in the proof of Proposition 3.2.10.

Now let y be an arbitrary point in X and, for all $n \geq 0$, let $y_n = f^n(y)$. Then either $y_n = x$ for some $n \geq 0$, and in this case, we have $f^m(y) = x$ for every $m \geq n$, or $|y_{n+1} - x| = |f(y_n) - f(x)| < |y_n - x|$ for all $n \geq 0$. Thus, either $(y_n)_{n\geq 0}$ is eventually the constant sequence $= x$, or the sequence $(|y_n - x|)_{n\geq 0}$ of positive real numbers is strictly decreasing. Suppose the latter holds and let $l = \lim_{n\to\infty} |y_n - x|$. Since X is compact, we can extract from (y_n) a convergent subsequence $(y_{n_i})_{i\geq 1}$. Let z be the limit of this subsequence. We have

$$\begin{aligned}l &= \lim_{i\to\infty} |y_{n_i} - x| = |z - x| = \lim_{i\to\infty} |y_{n_i+1} - x| \\ &= \lim_{i\to\infty} |f(y_{n_i+1}) - x| = |f(z) - x| \\ &= l.\end{aligned}$$

This implies that $z = x$, since otherwise we would have

$$|f(z) - x| = f(z) - f(x) < |z - x|$$

which implies $l < l$, which is a contradiction. Thus, every subsequence of (y_n) converges to x. This implies that the whole sequence (y_n) converges to x. This proves Proposition 3.2.12. □

Let us mention another result on distance-decreasing maps, which is also due to Busemann [29]:

Proposition 3.2.13. *Let X be a geodesic metric space and let $f : X \to X$ be a distance-decreasing map. If x is a point in X at which the displacement function d_f has a local minimum, then x is a fixed point of f (and therefore, d_f has a global minimum at x). Furthermore, x is the unique fixed point of f.*

Proof. The proof is by contradiction. Suppose that under the hypothesis of the proposition we have $x \neq f(x)$. Consider a geodesic segment $[x, f(x)]$ and let y be an arbitrary point in the interior of that segment. By Lemma 3.2.3(i), we have either $d_f(y) < d_f(x)$ or $|x - y| = |f(x) - f(y)|$. The first case cannot occur for y close enough to x, since x is a local minimum of d_f, and the second case contradicts the fact that f is distance-decreasing. Thus, by taking y close enough to x, we conclude that x is a fixed point for f. Now suppose that there is a fixed point x' for f that is distinct from x. Then we would have $|x - x'| = |f(x) - f(x')|$, which again contradicts the fact that f is distance-decreasing. Thus, x is the unique fixed point of f. □

3.3 Distance non-decreasing maps

Definition 3.3.1 (Distance non-decreasing map). Let X and Y be two metric spaces. A map $f : X \to Y$ is said to be *distance non-decreasing* if for all x and y in X, we have $|f(x) - f(y)| \geq |x - y|$.[4]

It is clear that a distance-non-decreasing map is injective and that it is not necessarily continuous. However, the left-inverse of such a map, which is defined on its image set, being a non-expanding map, is continuous.

The following technical lemma is the analog of Lemma 3.2.3 that was useful in the setting of non-expanding maps. It is also due to Busemann.

Lemma 3.3.2. *Let X and Y be metric spaces, let $f : X \to X$ be a distance non-decreasing map and let x and y be two points in X. If $f(y)$ lies between x and y, then $d_f(x) \geq d_f(y)$, and if equality holds, then $|x - y| = |f(x) - f(y)|$ and x lies between $f(x)$ and $f(y)$.*

[4] Busemann, in [29], calls such a map an "expansion".

Proof. We have

$$\begin{aligned}|x - f(x)| &\geq |f(y) - f(x)| - |f(y) - x| \\ &\geq |y - x| - |f(y) - x| \\ &= |y - f(y)| = d_f(y) > 0.\end{aligned}$$

If $d_f(x) = d_f(y)$, then all the large inequalities in the last sequence are equalities. The second of these inequalities then gives

$$|y - x| = |f(y) - f(x)| > 0,$$

and the first inequality being an inequality implies that x lies between $f(x)$ and $f(y)$. □

The following theorem is due to Busemann ([29]).

Theorem 3.3.3. *Let X be a compact metric space and let $f : X \to X$ be a distance non-decreasing map. Then f is surjective.*

Proof. Suppose that the set $X \setminus f(X)$ is nonempty and let x_0 be a point in this set. For any integer $n \geq 0$, let $x_n = f^n(x)$. We prove by induction that $x_n \notin f^{n+1}(X)$ for all $n \geq 0$. For $n = 0$, the property is true since $x_0 \notin f(X)$. If $x_n \in f^{n+1}(X)$ for some $n \geq 1$, then we would have $x_n = f^n(x_0) = f^{n+1}(y)$ for some y in X. Since f is injective, this implies that $x_{n-1} = f^{n-1}(x_0) = f^n(y)$, hence x_{n-1} is in $f^n(X)$. This completes the induction.

Therefore we have, for all $n \geq 0$, $x_n \in f^n(X) \setminus f^{n+1}(X)$.

Let r be the distance from x_0 to the set $f(X)$, that is,

$$r = \inf_{x \in X} |x_0 - f(x)|.$$

Since $x_0 \notin f(X)$, we have $|x_0 - f(x)| > 0$ for all x in X. Since X is compact, $f(X)$ is also compact and therefore $r > 0$.

For all nonnegative integers n and k, we have

$$|x_n - x_{n+k}| = |f^n(x_0) - f^{n+k}(x_0)| \geq \inf_{x \in X} |f^n(x_0) - f^{n+k}(x)|.$$

Since f is distance non-decreasing, all the iterates of f are also distance non-decreasing. Therefore, for every x in X and for all nonnegative integers n and k, we have $|f^n(x_0) - f^{n+k}(x)| \geq |x_0 - f^k(x)|$, which implies

$$|x_n - x_{n+k}| \geq \inf_{x \in X} |x_0 - f^k(x)| \geq \inf_{x \in X} |x_0 - f(x)| = r > 0.$$

This shows that no subsequence of (x_i) is convergent, contradicting the fact that X is compact. □

The following theorem is due to Freudenthal and Hurewicz [50]; cf. also Busemann [30] p. 42.

Theorem 3.3.4 (Freudenthal–Hurewicz). *Let X be a compact metric space and let $f: X \to X$ be a distance non-decreasing map. Then f is an isometry.*

Proof. Let us prove that f is distance-preserving. Let x_0 and y_0 be two points in X. For every integer $n \geq 0$, we set $x_n = f^n(x_0)$ and $y_n = f^n(y_0)$. Since X is compact, we can find subsequences $(x_{n_i})_{i\geq 0}$ and $(y_{n_i})_{i\geq 0}$ of $(x_n)_{n\geq 0}$ and $(y_n)_{n\geq 0}$ respectively that are convergent. Let ϵ be a positive real number and let i be a nonnegative integer such that $|x_{n_i} - x_{n_{i+1}}| < \epsilon/2$ and $|y_{n_i} - y_{n_{i+1}}| < \epsilon/2$. Since f is distance-non-decreasing, we can write

$$|x_{n_i} - x_{n_{i+1}}| \geq |x_{n_{i-1}} - x_{n_i}| \geq \cdots \geq |x_0 - x_k|,$$

where $k = n_{i+1} - n_i$. Thus, we have $|x_0 - x_k| < \epsilon/2$. In the same manner, we obtain $|y_0 - y_k| < \epsilon/2$. Therefore we have

$$\begin{aligned} |x - y| = |x_0 - y_0| &\leq |x_1 - y_1| \leq |x_k - y_k| \\ &\leq |x_k - x_0| + |x_0 - y_0| + |y_k - y_0| \\ &\leq |x_0 - y_0| + \epsilon. \end{aligned}$$

Since this is true for every $\epsilon > 0$, we conclude that all the large inequalities in the last sequence are equalities, and in particular, that $|x_0 - y_0| = |x_1 - y_1|$, or, equivalently, $|x_0 - y_0| = |f(x) - f(y)|$, which shows that f is distance-preserving. By Theorem 3.3.3, f is surjective. Therefore f is an isometry. This proves Theorem 3.3.4. □

We close this section with the following consequence of Theorem 3.3.4:

Corollary 3.3.5. *Let X be a compact metric space and let $f: X \to X$ be a distance-preserving map. Then f is an isometry.* □

3.4 Local isometries

Definition 3.4.1 (Local isometry). Let X and Y be two metric space. A map $f: X \to Y$ is said to be a *local isometry* if f is surjective and if every point x in X has a neighborhood $V(x)$ such that $f(V(x))$ is a neighborhood of $f(x)$ in Y such the restriction of f to $V(x)$ is an isometry between $V(x)$ and $f(V(x))$.

Proposition 3.4.2 (Inclusion and local isometry). *Let X be a geodesic space and let $X' \subset X$ be an open subset of X. Let $d_{X'}$ be the induced metric on X' and suppose that the metric space $(X', d_{X'})$ is connected by rectifiable paths. If d_ℓ is the intrinsic metric of $(X', d_{X'})$, then (X', d_ℓ) is a geodesic metric space and the inclusion map $(X', d_\ell) \to X$ is a local isometry.*

Proof. Let x be a point in X', let us choose a positive real number r such that the open ball $B(x, 2r)$ is contained in X'. Let B be the open ball $B(x, r)$. For every y and z in B, there exists a geodesic path in X joining y and z, whose image, by Proposition 2.2.15, is contained in the ball $B(x, 2r)$, and therefore in X'. Let us prove that γ is a geodesic path for the metric d_ℓ on X'. We have

$$L(\gamma) \geq d_\ell(y, z)$$

and, on the other hand,

$$L(\gamma) = d_X(y, z) = d_\ell(y, z) = d_{X'}(y, z) \leq d_\ell(y, z).$$

We conclude that $L(\gamma) = d_\ell(y, z)$, which proves that γ is a geodesic. This proves that the space (X', d_ℓ) is geodesic and that the inclusion map from $B(x, r)$ to X is distance-preserving. Finally, the fact that X' is open implies then that this inclusion is a local isometry. □

Lemma 3.4.3. *Let X be a metric space and let $f : X \to Y$ be a local isometry. Then, for each point x in X, there exists an open ball $B(x, r)$ of center x and radius $r > 0$ such that the restriction of f to $B(x, r)$ is an isometry between $B(x, r)$ and the open ball $B(f(x), r)$ of center $f(x)$ and radius r. Furthermore, if r_x denotes the supremum of the real numbers r such that $f_{|B(x,r)}$ is an isometry between $B(x, r)$ and $B(f(x), r)$, then the map $X \to]0, \infty[\cup \{\infty\}$ defined by $x \mapsto r_x$ is continuous.*

Proof. Let $V(x)$ be a neighborhood of x in X such that f induces an isometry between this neighborhood and its image. Choose $r > 0$ small enough so that the open ball $B(x, r)$ is contained in $V(x)$. Then it is clear that $f_{|B(x,r)}$ is a distance-preserving map. Let us prove that $f_{|B(x,r)}$ sends $B(x, r)$ surjectively on $B(f(x), r)$. We have $B(f(x), r) \subset f(V(x))$, and since $f_{|V(x)} : V(x) \to f\big(V(x)\big)$ is surjective, for any q in $B(f(x), r)$, we can find a point p in $V(x)$ such that $f(p) = q$. Since $|x - p|_X = |f(x) - q|_Y$, p is in $B(x, r)$. Thus, $f_{|B(x,r)} : B(x, r) \to B(f(x), r)$ is an isometry.

Now let us prove the continuity of $x \mapsto r_x$. Suppose first that there exists a point x in X such that $r_x = \infty$. Then $f : X \to Y$ is an isometry and $r_x = \infty$ for all x in X. Now suppose that $r_x < \infty$ for some x in X (or, equivalently, for all x in X). For every $r > 0$ and for every x' in $B(x, r)$, we have $B(x', r - |x - x'|) \subset B(x, r)$, which implies $r_x - r_{x'} \leq |x - x'|$. By symmetry, we also have $r_{x'} - r_x \leq |x - x'|$, which implies $|r_x - r_{x'}| \leq |x - x'|$, which shows that the map $r \mapsto r_x$ is 1-Lipschitz and therefore continuous. □

Proposition 3.4.4 (A local isometry is length-preserving). *Let X and Y be metric spaces and let $f : X \to Y$ be a local isometry. Then f is length-preserving.*

Proof. Let $\gamma : [a, b] \to X$ be a path. In the particular case where f is a isometry, we have, for any subdivision σ of $[a, b]$, $V_\sigma(\gamma) = V_\sigma(f \circ \gamma)$, whence $L_X(\gamma) = L_Y(f \circ \gamma)$.

Now let us suppose that f is only a local isometry. Consider the map $x \mapsto r_x$ of Lemma 3.4.3. By compactness of the interval $[a, b]$, there exists a positive real number r such that for all t in $[a, b]$, we have $r_{\gamma(t)} > r$. Since γ is uniformly continuous, there exists a positive real number η such that for all t and t' in $[a, b]$ satisfying $|t - t'| < \eta$, we have $|\gamma(t) - \gamma(t')| < r$. Now let $\sigma = (t_i)_{i=0,\ldots,n-1}$ be a subdivision of $[a, b]$ satisfying $|\sigma| < \eta$. Then we have $\gamma([t_i, t_{i+1}]) \subset B(\gamma(t_i), r)$ for all $i = 0, \ldots, n-1$, which implies, by the particular case considered at the beginning of the proof, $L_X(\gamma_{|[t_i,t_{i+1}]}) = L_Y(f \circ \gamma_{|[t_i,t_{i+1}]})$. By the additivity of the length function, we then obtain $L_X(\gamma) = L_Y(f \circ \gamma)$. □

Corollary 3.4.5 (A local isometry is non-expanding). *Let X and Y be two length spaces and let $f : X \to Y$ be a local isometry. Then f is non-expanding.*

Proof. Let x and y be in X. Then, for every path $\gamma : [0, 1] \to X$ joining these points, we have, using Proposition 3.4.4, $|f(x) - f(y)| \leq L_Y(f \circ \gamma) = L_X(\gamma)$. Taking the infimum over all paths γ joining x and y, we obtain $|f(x) - f(y)| \leq |x - y|$. □

Corollary 3.4.6 (Homeomorphism and local isometry implies isometry). *Let X and Y be two length spaces and let f be a homeomorphism that is a local isometry. Then f is an isometry.*

Proof. By Corollary 3.4.5, f and f^{-1} are non-expanding, which proves that f is an isometry. □

The next result will be particularly useful in the case where $f : X \to Y$ is a covering map.

Proposition 3.4.7 (The pull-back length metric induced by a local homeomorphism). *Let X be a Hausdorff topological space that is arcwise connected, let Y be a length space and let $f : X \to Y$ be a surjective local homeomorphism. Then there exists a unique length metric on X that makes $f : X \to Y$ a local isometry.*

Proof. For every x in X, we choose an open neighborhood $V(x)$ of x such that $f(V(x))$ is an open subset of Y containing $f(x)$ and such that f induces a homeomorphism between $V(x)$ and $f(V(x))$. To define the metric of X, we define first a map that assigns to each path γ in X a quantity $L^*(\gamma)$ that we call its "length", by setting, for each path $\gamma : [a, b] \to X$, $L^*(\gamma) = L_Y\big(f(\gamma)\big)$. (We note that since f is continuous, $f \circ \gamma$ is a path in Y.) Then we define the map $d : X \times X \to \mathbb{R} \cup \{\infty\}$ by setting, for all x and y in X,

$$d(x, y) = \inf_{\gamma} L^*(\gamma) \tag{3.4.7.1}$$

where the infimum is taken over the set of paths γ joining x to y.

Let us prove that d is a metric on X.

It is clear that $d(x, y) \geq 0$ for all x and y in X. It is also clear that $d(x, y) = d(y, x)$ since to every path $\gamma : [a, b] \to X$ joining x to y, we can associate a path $\overline{\gamma}$ joining y to x that has the same length. For instance, we take the path $\overline{\gamma} : [a, b] \to X$ defined by $\overline{\gamma}(t) = \gamma(a+b-t)$ for t in $[a, b]$. Let us prove that d satisfies the triangle inequality. Let x, y and z be three points in X. If γ is a path joining x to y and if γ' is a path joining y to z, then the concatenation of γ and γ' is a path joining x to z, and therefore we have $d(x, z) \leq L^*(\gamma * \gamma')$. From the definition, we can see that the "length" map L^* is additive with respect to concatenation of paths. This gives $d(x, z) \leq L^*(\gamma) + L^*(\gamma')$. Taking the infimum of the two members of this inequality over all paths γ and γ' joining the pairs x, y and y, z respectively, we obtain $d(x, z) \leq d(x, y) + d(y, z)$.

Now let us prove that if x and y are distinct points in X, we have $d(x, y) > 0$. Let $V(x)$ be a neighborhood of x and let $V(y)$ be a neighborhood of y such that f induces a homeomorphism between each of these neighborhoods and its image. Since X is Hausdorff, we can assume that $V(x) \cap V(y) = \emptyset$. The image set $f\big(V(x)\big)$ contains an open ball $B\big(f(x), r(x)\big)$ of radius $r(x) > 0$. Likewise, the image set $f\big(V(y)\big)$ contains an open ball $B\big(f(y), r(y)\big)$ of radius $r(y) > 0$. Then, for every path γ joining x to y, we have $L^*(\gamma) = L_Y(f(\gamma)) > r(x) + r(y)$, which implies $d(x, y) > 0$.

To show that d is a metric, it remains to prove that for all x and y in X, we have $d(x, y) < \infty$. For that, we must prove that for every x and y in X, there exists a path joining them and satisfying $L^*(\gamma) < \infty$. We start with an arbitrary path γ joining x and y, and we shall modify it to obtain a path with the desired property. Let $\gamma' = f \circ \gamma$ and let s be a point in $[a, b]$. Let $V(\gamma(s))$ be a neighborhood of $\gamma(s)$ such that f induces a homeomorphism between $V(\gamma(s))$ and its image $f\big(V(\gamma(s))\big)$. Let $r(s)$ be a positive real number such that the open ball $B(\gamma'(s), r(s))$ is contained in $f\big(V(\gamma(s))\big)$. We set

$$V'(\gamma(s)) = V(\gamma(s)) \cap f^{-1}\big(B(\gamma'(s), r(s))\big).$$

Then f induces a homeomorphism between the set $V'(\gamma(s))$, which is a neighborhood of $\gamma(s)$, and the open ball $B(\gamma'(s), r(s))$. By compactness of the subset $\gamma'([a, b])$ of Y, we can find a finite collection of open balls $\{B(\gamma'(s_i), r(s_i))\}_{i=0,\dots,k}$ covering the set $\gamma'([a, b])$, with $s_i \in \gamma'([a, b])$ for each $i = 0, 1, \dots, k$. By compactness of $[a, b]$, we can find a subdivision $\sigma = (t_i)_{i=0,\dots,n}$ of $[a, b]$ such that for each $i = 0, \dots, n-1$, there exists an integer j in $[0, k]$ such that the set $\gamma'([t_i, t_{i+1}])$ is contained in the ball $B(\gamma'(s_j), r(s_j)/2)$. By Proposition 2.1.19, we can find, for each $i = 0, \dots, n-1$, a path γ_i' that joins $\gamma'(t_i)$ and $\gamma'(t_{i+1})$ whose length is $< 2r$ and whose image is contained in the open ball $B(\gamma'(s_j), r(s_j))$. Using the inverse map of the restriction of f to the set $V'(\gamma(s_j))$, we can lift the path γ_i' to a path γ_i that joins $\gamma(t_i)$ to $\gamma(t_{i+1})$ and that satisfies $L^*(\gamma_i) = L_Y(\gamma_i') < 2r$. Concatenating the n paths γ_i ($i = 0, \dots, n-1$), we obtain a path γ^* that joins x to y and that satisfies $L(\gamma^*) < \infty$. This proves that $d(x, y) < \infty$.

Now let us prove that if we equip X with the metric d, then the map f becomes a local isometry. Again, let x be a point in X and let $V(x)$ be a neighborhood of x such that f induces a homeomorphism between $V(x)$ and an open ball $B\big(f(x), r(x)\big)$ of center $f(x)$ and of radius $r(x) > 0$. Let

$$V'(x) = f^{-1}\big(B(f(x), r(x)/4)\big) \cap V(x),$$

let x_1 and x_2 be two points in $V'(x)$ and let y_1 and y_2 be their images in $B\big(f(x), r(x)/4\big)$. We have $|y_1 - y_2| = \inf L_Y(\gamma)$, where the infimum is taken over the set of paths γ joining y_1 and y_2. Using Proposition 2.1.19, we can find a path γ joining y_1 and y_2 whose image is contained in the open ball $B\big(f(x), r(x)\big)$ and whose length satisfies $L(\gamma) < 2r(x)/4 = r(x)/2$. Such a path γ can be lifted to a path γ' that joins x_1 and x_2 and satisfies $L^*(\gamma') = L_Y(\gamma)$. Therefore we have $d(x_1, x_2) \leq |y_1 - y_2|$. To prove the inverse inequality, let $(\gamma_n)_{n\geq 0}$ be a sequence of paths in X joining x_1 and x_2 and satisfying $L^*(\gamma_n) \to d(x_1, x_2)$ as $n \to \infty$. The image paths $f(\gamma_n)$ join y_1 and y_2, and therefore we have $|y_1 - y_2| \leq d(x_1, x_2)$. Thus, we obtain $|y_1 = y_2| = d(x_1, x_2)$, which shows f is an isometry between $V'(x)$ and its image.

Now we prove that d is a length metric. Consider two points x and y in X. Since f is length-preserving (Proposition 3.4.4), for every path γ joining x and y, we have $L_X(\gamma) = L_Y(f(\gamma))$. Thus, $L_X(\gamma) = L^*(\gamma)$, and from the definition of d, we have $d(x, y) = \inf_\gamma L_X(\gamma)$ where the infimum is taken over the set of paths γ joining x and y. This proves that d is a length metric.

It remains to see the uniqueness of the metric d on X for which $f\colon X \to Y$ is a local isometry. Let d' be a metric on X for which the map f is a local isometry. Since f is length-preserving, we have $d'(x, y) = \inf_\gamma L_X(\gamma) = \inf_\gamma L_Y(f \circ \gamma)$, where the infimum is taken over the set of paths γ joining x and y, which implies $d'(x, y) = d(x, y)$. This completes the proof of Proposition 3.4.7. □

The metric on X provided by Proposition 3.4.7 is called the *pull-back length metric on X induced by the local homeomorphism f.*

Proposition 3.4.8 (The image of a local geodesic by a local isometry). *Let X and Y be length spaces, let $f\colon X \to Y$ be a local isometry and let $\gamma\colon [a, b] \to X$ be a path. If γ is a local geodesic, then $f \circ \gamma\colon [a, b] \to Y$ is also a local geodesic.*

Proof. As in the proof of Proposition 3.4.4, we take a positive real number r such that for all t in $[a, b]$, the restriction of f to the open ball $B(\gamma(t), r)$ is an isometry between this ball and its image $B(f \circ \gamma(t), r)$ and we take $\eta > 0$ such that for all t and t' in $[a, b]$, satisfying $|t - t'| < \eta$, we have $|\gamma(t) - \gamma(t')| < r$. Since the map $\gamma\colon [a, b] \to X$ is a local geodesic, there exists, for each t in $[a, b]$, a neighborhood $V(t)$ of t in $[a, b]$ such that $\gamma_{|V(t)}$ is distance-preserving. Now let t be a point in $[a, b]$ and let $W(t) = V(t) \cap]t - \eta, t + \eta[$. Then $W(t)$ is a neighborhood of t in $[a, b]$ and for all t' and t'' in $W(t)$, we have

$$|f \circ \gamma(t') - f \circ \gamma(t'')| = |\gamma(t') - \gamma(t'')| = |t' - t''|,$$

which shows that $f \circ \gamma$ is a local geodesic. □

Definition 3.4.9 (The lift of a path). Let X and Y be two metric spaces and let $f \colon X \to Y$ be a continuous map. Let $\gamma \colon [a,b] \to Y$ be a path and let x be a point in $f^{-1}(\gamma(a))$. We say that γ' is a *lift of γ starting at x* (*by the map f*) if $\gamma' \colon [a,b] \to X$ is a path satisfying $f \circ \gamma' = \gamma$ and $\gamma'(a) = x$.

Proposition 3.4.10 (The lift of a geodesic by a local isometry). *Let X and Y be two length spaces and let $f \colon X \to Y$ be a local isometry. Let I be a compact interval of $\mathbb{R}$* (*respectively let $I = \mathbb{R}$, $I = [0, \infty[$*) *and let $\gamma \colon I \to X$ be a continuous map. If the map $f \circ \gamma \colon I \to Y$ is a geodesic path* (*respectively a geodesic line, a geodesic ray*)*, then γ is also a geodesic path* (*respectively a geodesic line, a geodesic ray*).

Proof. We first consider the case where I is a compact interval $[a,b]$. We have:

$$\begin{aligned} |f \circ \gamma(a) - f \circ \gamma(b)| &= L_Y(f \circ \gamma) \quad \text{(since } f \circ \gamma \text{ is geodesic)} \\ &= L_X(\gamma) \quad \text{(by Proposition 3.4.4)} \\ &\geq |\gamma(a) - \gamma(b)| \\ &\geq |f \circ \gamma(a) - f \circ \gamma(b)| \quad \text{(by Proposition 3.4.5).} \end{aligned}$$

We conclude that the last two inequalities are equalities, which implies $L(\gamma) = |\gamma(a) - \gamma(b)|$. This proves that the path γ is geodesic. In the case where I is $\mathbb{R}$ or $[0, \infty[$, we use the preceding case to show that the restriction of f to any compact sub-interval of I is distance-preserving, which shows that $\gamma \colon I \to X$ is distance-preserving, and therefore that it is a geodesic line (respectively a geodesic ray). □

Proposition 3.4.11 (Existence and uniqueness of lifts of geodesics and of local geodesics). *Let X and Y be two length spaces such that X is complete, let $f \colon X \to Y$ be a local isometry, let $\gamma \colon [a,b] \to Y$ be a local geodesic and let x be a point in $f^{-1}(\gamma(a))$. Then there exists a unique local geodesic $\gamma' \colon [a,b] \to X$ that is a lift of γ starting at x. In the case where γ is geodesic, γ' is also geodesic*

Proof. Let us first prove that if $\gamma' \colon [a,b] \to X$ is a lift of γ, then γ' is a local geodesic. Let t a point in $[a,b]$ and let $y = \gamma'(t)$. The map f being a local isometry, there exists a neighborhood $V(y)$ of y in X such that $f(V(y))$ is a neighborhood of $f(y) = \gamma(t)$ in Y and such that the restriction of f to $V(y)$ is an isometry between $V(y)$ and $f(V(y))$. Let ϵ_1 be a positive real number satisfying $\gamma'([a,b] \cap]t - \epsilon_1, t + \epsilon_1[) \subset V(y)$. Since γ is a local geodesic, there exists a positive real number ϵ_2 such that the restriction of γ to $[a,b] \cap]t - \epsilon_2, t + \epsilon_2[$ is distance-preserving. Let $\epsilon = \min\{\epsilon_1, \epsilon_2\}$ and let $V(t) = [a,b] \cap]t - \epsilon, t + \epsilon[$. Then $V(t)$ is a neighborhood of t in $[a,b]$ and we have, for all t' and t'' in $V(t)$,

$$|\gamma'(t') - \gamma'(t'')| = |f(\gamma'(t')) - f(\gamma'(t''))| = |\gamma(t') - \gamma(t'')| = |t' - t''|.$$

This shows that the restriction of γ' to $V(t)$ is a geodesic. We conclude that γ' is a local geodesic.

Next, we prove uniqueness. As before, for all s satisfying $a \leq s \leq b$, γ_s denotes the path $\gamma_{|[a,s]}$.

Let J be the set of real numbers s in $[a, b]$ such that γ_s can be lifted in a unique way to a path starting at x.

The set J is nonempty since it contains the point a. It is also clear that J is an interval having a as an initial point. We prove that J is open and closed in $[a, b]$; this will imply that $J = [a, b]$.

Let $(s_n)_{n\geq 0}$ be a sequence of real numbers in J converging to a point s, and let us prove that s is in J. Without loss of generality, we can suppose that for every integer $n \geq 0$, we have $s_n < s$ (otherwise, s is trivially in J). For each $n \geq 0$, the path γ_{s_n} can be lifted in a unique way to a local geodesic $\gamma'_{s_n} : [a, b] \to X$ satisfying $\gamma'_{s_n}(a) = x$. Furthermore, for all $j \leq n$, the path $\gamma'_{s_n} : [a, s_n] \to X$ is an extension of the local geodesic $\gamma'_{s_j} : [a, s_j] \to X$ (we are using the uniqueness of lifts). The sequence of points $(\gamma'_{s_n}(s_n))$ is then a Cauchy sequence in X and since X is complete, this sequence converges to a point x_s in X. We then define $\gamma'_s : [a, s] \to X$ by setting $\gamma'_s(t) = \gamma'_{s_n}(t)$ for any t in $[a, s_n]$ and $\gamma'_s(s) = x_s$. The local geodesic γ'_s is a lift of the path γ_s and, again by the uniqueness of lifts, γ'_s is an extension of any path γ_{s_n}, for every integer $n \geq 0$. Likewise, the uniqueness of the lifts of the paths γ_{s_n} shows that γ'_s is the unique lift of γ_s starting at x. Thus, J is a closed subset of $[a, b]$.

Now let us prove that J is an open subset of $[a, b]$.

Let s a real number in $[a, b[$ satisfying $[a, s] \subset J$ and let $\gamma'_s : [a, s[\to X$ be the unique lift of γ_s satisfying $\gamma'_s(a) = x$. Let us take an open neighborhood $V\big(\gamma'_s(s)\big)$ of $\gamma'_s(s)$ such that f induces an isometry between $V\big(\gamma'_s(s)\big)$ and the open neighborhood $f\big(V\big(\gamma'_s(s)\big)\big)$ of $f\big(\gamma'_s(s)\big) = \gamma(s)$. Using the inverse image of $\gamma([s, b]) \cap f\big(V\big(\gamma'_s(s)\big)\big)$ by this isometry, we obtain a real number s' satisfying $s < s' \leq b$ such that $\gamma_{|[s,s']}$ can be lifted in a unique way in a path γ'' that satisfies $\gamma''(s) = \gamma'_s(s)$. Then the path $\gamma'_{s'}$, defined as the concatenation of γ'_s with γ'', is a path that is a lift of $\gamma_{s'}$ and that satisfies $\gamma'_{s'}(a) = x$, and it is clear by construction that this is the unique path that satisfies these properties. This shows that J is an open subset of $[a, b]$.

Thus, we have proved that $J = [a, b]$.

It remains to show that if γ is geodesic, then γ' is also geodesic.

In case γ is a geodesic path, we have $L(\gamma) = |\gamma(a) - \gamma(b)|$. Hence, by Proposition 3.4.4 (A local isometry is length-preserving), we have $L(\gamma') = |\gamma(a) - \gamma(b)|$. By Corollary 3.4.5 (A local isometry is non-expanding), we have $|\gamma(a) - \gamma(b)| \leq |\gamma'(a) - \gamma'(b)|$. Therefore we obtain $L(\gamma') \leq |\gamma'(a) - \gamma'(b)|$, which implies $L(\gamma') = |\gamma'(a) - \gamma'(b)|$. Thus, γ' is a geodesic path. This completes the proof of Proposition 3.4.11. □

Proposition 3.4.12. *Let X and Y be two uniquely geodesic metric spaces such that X is complete and let $f : X \to Y$ be a local isometry. Let $\gamma : [a, b] \to Y$ be a geodesic path and let x be a point in $f^{-1}(\gamma(a))$. Then there exists a unique geodesic path*

$\gamma'\colon [a,b] \to X$ that satisfies $\gamma'(a) = x$ and $\gamma'(b) \in f^{-1}(\gamma(b))$ and whose length is equal to $|\gamma(a) - \gamma(b)|$. Furthermore, γ' is the unique local geodesic path starting at x that is a lift of γ.

Proof. By Proposition 3.4.11, there exists a unique local geodesic $\gamma'\colon [a,b] \to X$ (which in fact is a geodesic, since X is uniquely geodesic) that is a lift of γ and satisfies $\gamma'(a) = x$. In particular, $\gamma'(b)$ is in $f^{-1}(\gamma(b))$. By Proposition 3.4.11, we have

$$L(\gamma') = L(f(\gamma')) = L(\gamma) = |\gamma(a) - \gamma(b)|.$$

Let $\gamma''\colon [a,b] \to X$ be an arbitrary geodesic satisfying $\gamma''(a) = x$ and $\gamma''(b) \in f^{-1}(\gamma(b))$ and whose length is equal to $|\gamma(a) - \gamma(b)|$. To prove that $\gamma' = \gamma''$, it suffices to prove that γ'' is a lift of γ. The path $f \circ \gamma''$ joins the points $\gamma(a)$ and $\gamma(b)$, and its length is equal to $|\gamma(a) - \gamma(b)|$. We conclude that this path is geodesic. By the uniqueness of the geodesic segment joining $\gamma(a)$ and $\gamma(b)$, we have $f \circ \gamma'' = \gamma$. Then, by the uniqueness of lifts (Proposition 3.4.11), we have $\gamma' = \gamma''$. This competes the proof of Proposition 3.4.12. □

3.5 Covering spaces

A *covering map* between two topological spaces X and Y is a continuous map $f\colon X \to Y$ such that every point y in Y has an open neighborhood V such that $f^{-1}(V)$ is a disjoint union of sets $\{V_\alpha\}_{\alpha\in\mathcal{I}}$, for some set $\mathcal{I}$ and for all α in $\mathcal{I}$, the restriction of f to V_α is a homeomorphism between V_α and V. Such an open set V is called a *distinguished neighborhood* of x. The space X is called a *covering space* of Y, and Y is called the *base space* of the covering. Every covering map is a local homeomorphism but there are standard examples of local homeomorphisms that are not coverings (for instance, the map $x \mapsto (\cos x, \sin x)$ from $]0, 3\pi[$ to the circle S^1). For the bases of the theory of coverings, we refer the reader to the books [98] and [114] by Massey and by Munkres.

There is a geometric theory of covering spaces that is particularly well adapted to the setting of metric spaces, and in that theory, it is assumed that a covering map $f\colon X \to Y$ is a local isometry. In this geometrical setting, there are useful sufficient conditions for a local homeomorphism between metric spaces to be a covering. In this section, we present some of these conditions.

The classical theory of covering spaces associates to a space X that is locally arcwise connected and semi-locally simply connected a covering space $\tilde{X}$ that is simply connected, which is called the *universal covering* of X. The universal covering is uniquely defined up to a natural equivalence relation between covering spaces. For the definition of $\tilde{X}$, one starts by choosing a basepoint x in X, and then one considers the set C_x of all paths in X starting at x. As a set, the space $\tilde{X}$ is the quotient of C_x by the equivalence relation that identifies two elements γ and γ' in C_x whenever these paths have the same endpoint and if there exists a homotopy between these two paths that leaves the endpoints fixed. In Chapter 9 we shall see that the universal covering of a space whose distance function satisfies some convexity condition that we study there is naturally realized as a set of local geodesics starting at x.

We start with the following

Proposition 3.5.1 (The length metric induced on a covering). *Let X be a Hausdorff topological space that is arcwise connected, let Y be a length space and let $f\colon X \to Y$ be a covering map. Then there exists a unique length metric on X such that $f\colon X \to Y$ is a local isometry.*

Proof. Since f is a local surjective homeomorphism, we take on X the pull-back length metric provided by Proposition 3.4.7. □

If $p\colon X \to Y$ is a covering map, then a homeomorphism of X is called a *deck transformation* if it satisfies $p \circ f = p$. A deck transformation is also called an automorphism of the covering, and the deck transformations form a group called the *automorphism group* of the covering.

Proposition 3.5.2. *Let X be a length space and let $p\colon X \to Y$ be a covering map that is a local isometry. Then each deck transformation $f\colon X \to X$ of this covering is an isometry.*

Proof. Since p is a local isometry and since $p \circ f = p$, we deduce easily that f is also a local isometry. By Proposition 3.4.6, a homeomorphism that is a local isometry is an isometry. □

By the classical theory of coverings, if $f\colon \tilde{X} \to X$ is a universal covering map, then there is a canonical isomorphism between the fundamental group of X and the group of deck transformations. We also recall that a group *acts freely* on a space if every element of the group except the identity is fixed point-free. Finally, we recall that a group acts *properly discontinuously* if every element of the space has a neighborhood K such that the set of elements g in the group satisfying $gK \cap K \neq \emptyset$ is finite.

We record the following consequence of Proposition 3.5.2:

Corollary 3.5.3. *Let X be a length space. Then the fundamental group of X is isomorphic to a group acting freely and properly discontinuously by isometries on a simply connected length space.*

Proof. We consider the universal covering space $f\colon \tilde{X} \to X$. The space X is Hausdorff, and this implies that $\tilde{X}$ is Hausdorff. Using Proposition 3.5.2, we equip $\tilde{X}$ with a length metric such that f is a local isometry. The fundamental group of X acts on $\tilde{X}$ as the group of deck transformations. By Proposition 3.5.2, each deck transformation is an isometry of $\tilde{X}$, and the group of deck transformations satisfies all the required properties. This proves Corollary 3.5.3. □

We now study some basic properties of covering spaces of metric spaces, in which the covering map is a local isometry.

Theorem 3.5.4. *Let X and Y be two complete, locally compact and locally uniquely geodesic length spaces and let $f : X \to Y$ be a local isometry. Then f is a covering map.*

Proof. Let y be a point in Y and let $B = B(y, r)$ be an open ball of center y and of positive radius r. We choose r small enough so that B is uniquely geodesic. We claim that B is a distinguished neighborhood of y, that is, that $f^{-1}(B) = \bigcup_{x \in f^{-1}(y)} B_x$, where for all x in $f^{-1}(y)$, B_x is a neighborhood of x such that $f_{|B_x} : B_x \to B$ is a homeomorphism and such that for all distinct points x and x' in $f^{-1}(y)$, we have $B_x \cap B_{x'} = \emptyset$. This will imply that f is a covering map.

To prove the claim, let x be an arbitrary point in $f^{-1}(y)$. (Such a point exists since f is surjective.) For every q in B, let $\gamma : [0, |y - q|] \to B$ be the unique geodesic starting at y and ending at q (that is, the unique such geodesic with $[0, |y - q|]$ as domain). By Propositions 3.4.11 and 3.4.12, there exists a unique geodesic $\gamma' : [0, |y - q|] \to X$ starting at x, parametrized by $[0, |y - q|]$, whose endpoint $\gamma'(|y - q|)$ is in $f^{-1}(\gamma(|y - q|))$, that is a lift of γ and that satisfies $L(\gamma') = L(\gamma) = |y - q|$. We denote by B_x the set of points $p = \gamma'(|y - q|)$ in X that can be obtained as endpoints of such geodesics γ'. In particular, we have $|x - p| = |y - q|$. Let us note that B_x contains the point x itself since this point is the endpoint of the geodesic of length 0. We denote by $\psi_x : B \to B_x$ the map that associates to any point q in B the point p provided by this construction. For all q in B, we have $f \circ \psi_x(q) = q$. Therefore the map ψ_x is injective. We conclude that the restriction of f to B_x is a bijection between B_x and B. Since Y is complete and locally compact, the open ball B is relatively compact (Theorem 2.1.16). Therefore the map induced by f between B_x and B, which is bijective and continuous, is a homeomorphism. In particular, B_x is a neighborhood of x in X.

Now let us prove that $f^{-1}(B) = \bigcup_{x \in f^{-1}(y)} B_x$. Let p be in $f^{-1}(B)$ and let $q = f(p)$. Since q is in B, there is a unique geodesic $\gamma : [0, |y - q|] \to Y$ starting at q and ending at y. By Proposition 3.4.12, there exists a unique geodesic $\gamma' : [0, |y - q|] \to X$ starting at p and ending at a point in $f^{-1}(y)$. Let x be this point. We have $|p - x| = |y - q|$, which implies that p is in B_x. Thus, $f^{-1}(B) \subset \bigcup_{x \in f^{-1}(y)} B_x$. Now let x be a point in $f^{-1}(y)$ and let us prove that $B_x \subset f^{-1}(B)$. For all p in B_x, we have, by construction, $f(p) \in B$, which implies that $f(B_x) \subset B$, that is, $B_x \subset f^{-1}(B)$. Therefore we have $f^{-1}(B) = \bigcup_{x \in f^{-1}(y)} B_x$.

To prove that f is a covering map, it remains to show that if x and x' are two distinct points of $f^{-1}(y)$, then $B_x \cap B_{x'} = \emptyset$. Thus, let us take two such points x and x' and let us suppose that there exists a point p in the intersection $B_x \cap B_{x'}$. Let $q = f(p)$ and let $\gamma : [0, |y - p|] \to B$ be the geodesic in B that joins p to y. Then there exist two geodesics $\gamma'_1 : [0, |y - p|] \to B$ and $\gamma'_2 : [0, |y - p|] \to B$ that start at p and satisfy $\gamma'_1(|y - p|) = x$, $\gamma'_2(|y - p|) = x'$ and $L(\gamma'_1) = L(\gamma'_2) = |y - p|$.

Thus, we obtain two distinct geodesics that are lifts of γ and that start at p, which is a contradiction. Therefore we have $B_x \cap B_{x'} = \emptyset$. This completes the proof of Theorem 3.5.3. □

We shall use the following result in Chapter 9.

Proposition 3.5.5. *Let X and Y be two length spaces and let $f: X \to Y$ be a covering map that is a local isometry. If Y is complete, then X is complete.*

Proof. Suppose that Y is complete. Let $(x_n)_{n\geq 1}$ be a Cauchy sequence in X. For every $n \geq 1$, let us set $y_n = f(x_n)$. The map f being non-expanding (Proposition 3.4.5), the sequence (y_n) is a Cauchy sequence in Y. Since Y is complete, this sequence converges. Let y be its limit. To prove that (x_n) converges, we first suppose that f is a homeomorphism. Let $x = f^{-1}(y)$ and let $B = B(y, r)$ be the open ball of center y and radius r, where r is small enough so that if $B' = B(x, r)$, then $f_{|B'}: B' \to B$ is an isometry. There exists an integer n_0 such that y_n is in B for all $n \geq n_0$. Then B' contains x_n for all $n \geq n_0$, and since $f_{|B'}: B' \to B$ is an isometry, (x_n) converges to x. We conclude that every Cauchy sequence in X converges, and therefore X is complete. Now let us consider the general case where f is not necessarily a homeomorphism. Let r be a positive real number such that the ball $B(y, r)$ is a distinguished neighborhood of y and let x' and x'' be two distinct points in $f^{-1}(y)$. The open balls $B(x', r)$ and $B(x'', r)$ are disjoint and therefore we have $|x' - x''| \geq 2r$.We conclude that for every z' in $B(x', r/2)$ and z'' in $B(x'', r/2)$, we have $|z' - z''| > r$. Taking n large enough, we have $y_n \in B(y, r/2)$, which implies $x_n \in f^{-1}\big(B(y, r/2)\big)$. Now we take k large enough so that $|x_n - x_m| \leq r/2$ whenever n and m are $\geq k$. The sequence $(x_n)_{n\geq k}$ is then contained in a ball $B(x, r/2)$ for some x in $f^{-1}(y)$. In the same way as in the case where f is homeomorphism, the fact that f induces an isometry between $B(x, r/2)$ and $B(y, r/2)$ implies that (x_n) converges to x. We conclude that X is complete. □

Definition 3.5.6 (Locally uniquely locally geodesic space). A metric space X is said to be *locally uniquely locally geodesic* if every point x in X has a neighborhood $V(x)$ such that for all p and q in $V(x)$, there exists a unique local geodesic $\gamma: [0, |p-q|] \to V(x)$ that joins p and q.

Examples of such spaces are the nonpositively curved Riemannian manifolds and the locally convex metric spaces that we consider in Chapter 8.

Proposition 3.5.7. *Let $f: X \to Y$ be a local isometry, where X and Y are complete locally compact length spaces such that the space X (or, equivalently, Y) is locally uniquely geodesic and locally uniquely locally geodesic. Then for all y in Y, there exists an open ball $B = B(y, r)$ of radius $r > 0$ such that $f^{-1}(B)$ is the disjoint union of open balls $B(x, r)$ with x ranging in $f^{-1}(y)$ and such that for each x in $f^{-1}(y)$ the map $f_{|B(x,r)}: B(x, r) \to B(y, r)$ is a homeomorphism.*

Proof. We use the notations of the proof of Theorem 3.5.4 and we suppose furthermore that the open ball $B(y,r)$ is uniquely geodesic and uniquely locally geodesic. Let us prove that the set B_x is then the open ball $B(x,r)$ of center x and of radius r. For every q in B, we have $|x-\psi_x(q)| = |y-q| < r$, which implies $B_x \subset B(x,r)$. Now if p is an arbitrary point in $B(x,r)$, then, since X is geodesic, there exists a geodesic path $\gamma'\colon [0,|x-p|] \to X$ that joins x to p. The image of this path is contained in $B(x,r)$. Let $\gamma = f \circ \gamma'$ and let $q = \gamma(|x-p|)$. We have $L(\gamma) = L(\gamma') < r$ and γ is a local geodesic whose image is contained in the ball $B(y,r)$. Since this ball is uniquely locally geodesic, γ is the unique geodesic that joins y to q, and therefore we have $p = \psi_x(q)$, which shows that $B(x,r) \subset B_x$. Thus, we have $B_x = B(x,r)$. This completes the proof of Proposition 3.5.7. □

Proposition 3.5.8. *Let X be a compact length space that is locally uniquely locally geodesic and let $f\colon X \to X$ be a covering map that is a local isometry. Then f is an isometry.*

Proof. Let us first prove that f is a homeomorphism. Let y be a point in X. By Proposition 3.5.7, there exists an open ball $B = B(y,r)$ of positive radius r such that $f^{-1}(B)$ is the disjoint union of open balls $B(x,r)$, with x ranging in the set $f^{-1}(y)$, and such that for every such x the map $f_{|B(x,r)}\colon B(x,r) \to B(y,r)$ is a homeomorphism. Since f is a covering map, the cardinality of the fiber $f^{-1}(y)$ does not depend on the choice of the point y. If f is not a homeomorphism, this cardinality is ≥ 2. Let x_1 and x_2 be two distinct points in $f^{-1}(y)$. Since the balls $B(x_1,r)$ and $B(x_2,r)$ are disjoint, we have $|x_1 - x_2| > 2r$.

In the same way, we can find two distinct points x_1^1 and x_1^2 in $f^{-1}(x_2)$, and two distinct points x_2^1 and x_2^2 in $f^{-1}(x_1)$. Furthermore, for all integers i and $j = 1,2$ and for all $k \neq \ell$, we have $x_i^k \neq x_j^\ell$, since the images by $f^2 = f \circ f$ of these two points are distinct. We conclude that the four points x_1^1, x_1^2, x_2^1 and x_2^2 are the centers of balls of radius r that are disjoint, and the distance between any two points among these four points is $> 2r$.

By iterating n times this construction, we obtain, for any integer $n \geq 0$, a sequence of 2^n points in X whose mutual distances are all $> 2r$. This contradicts the fact that X is compact. We deduce that f is a homeomorphism.

The inverse of the homeomorphisms f is also a local isometry. By Proposition 3.4.5, f and f^{-1} are non-expanding maps. This implies that f is a isometry. This completes the proof of Proposition 3.5.8. □

Remark. The result of Proposition 3.5.8 does not remain true if instead of a covering $f\colon X \to X$ we take a covering $f\colon X \to Y$ between two different metric spaces X and Y that satisfy the hypotheses of the proposition (that is, if X and Y are compact length spaces that are locally uniquely locally geodesic). To see this, let X and Y be the circle S^1, $f\colon S^1 \to S^1$ defined by $z \mapsto z^2$, with the metric on the range being the length metric θ induced by the inclusion of the circle S^1 in the Euclidean plane

(θ is the "angular metric") and with the metric on the domain being the length metric induced by θ by the local homeomorphism $z \mapsto z^2$ as in Proposition 3.5.2.

Notes on Chapter 3

Fixed point theory. Several results in Section 2 of this chapter concern fixed points of maps, and it is worth mentioning, since convexity is the main topic of this book, that fixed point theory is closely related to convexity theory. For instance, it was realized a long time ago that the famous fixed point theorem of Brouwer [21], stating that a continuous map of an n-dimensional simplex has a fixed point, holds for a continuous map of any closed convex set in an n-dimensional topological vector space.

The fixed point theorem for contractions (Theorem 3.2.10) is due to Banach. This is Theorem 6, p. 160 of the paper [9] which constitutes the doctoral thesis of Banach, presented to the University of Leopol (Lvov) in 1922. This result of Banach was historically a starting point for a whole series of results in functional analysis that deal with fixed points and approximation theory.

Corollaries 3.2.5 and 3.2.6 are classical instances of results in the theory of fixed points for non-expanding maps, and this subject is vast. To give a more elaborate example, we mention the following result due to E. Rakotch (see [123]). Let X be a complete metric space, let $f: X \to X$ be a map and suppose there exists α in $]0, 1[$ such that every x in X has a neighborhood $V(x)$ such that $|f(y) - f(z)| \leq \alpha|y - z|$ for every y and z in $V(x)$. (The author calls such a map α-*locally contractive*). If there exists a point x_0 in X such that x_0 and $f(x_0)$ can be joined by a rectifiable path, then f has a fixed point. In the same paper, the author gives an example of a complete and connected space X with an α-locally contractive mapping $f: X \to X$ that has no fixed point, and he proves that if any two points in X can be joined by a rectifiable path, then f has a unique fixed point.

Busemann's theory of covering spaces. Busemann worked on a theory of covering spaces for complete G-spaces, where covering maps are locally isometric maps. This theory is developed in [25], chap. IV and [28], §27, and it is used by Busemann in his study of the isometries of G-spaces. A version of Theorem 3.5.4 is contained in [28] (Theorem 27.9) with different hypotheses. The hypotheses that we put here on the space X and Y will be satisfied when we take X and Y to be locally convex spaces, in Chapters 7 and 8 below. Likewise, Proposition 3.5.8 is due to Busemann, with slightly different hypotheses (see [25], Theorem 12.15 and [28], Theorem 27.14).

The Carathéodory and the Kobayashi pseudo-metrics for coverings. If $X \to Y$ is a holomorphic covering between complex manifolds, the pull-back of the Kobayashi pseudo-metric of Y is the Kobayashi pseudo-metric of X. This means that the Kobayashi pseudo-distance satisfies a property analogous to the property described in Proposition 3.4.7 for metrics; see [126]. The Carathéodory pseudo-metric does not satisfy this property. This is due to the fact (which we mentioned in Chapter 2) that the Kobayashi pseudo-distance is a length pseudo-distance whereas the Carathéodory pseudo-distance is not.

Chapter 4

Distances

Introduction

In this chapter, we use the word "distance" in the following broad sense: a distance is a function on the product space of a set with itself, that satisfies some of the axioms of a metric, but not necessarily all of them. (In general, the axiom that will not be satisfied will be the separation axiom, that is, $d(x, y) = 0 \Rightarrow x = y$.) Of course, we shall be interested in knowing under which additional conditions on the base set such a generalized distance defines a genuine metric.

We shall deal with two kinds of such generalized distances:

• distances between subsets of metric spaces;

and

• distances between isometries of a metric space.

We study these two notions of distances in the same chapter because there are similarities in the definitions and in the developments.

We shall use some of the results in later chapters, but this chapter can also be regarded as a study of examples of metric spaces.

Let X be a metric space. Given two nonempty subsets A and B of X, their *Hausdorff distance* is defined as

$$d_{\mathcal{H}}(A, B) = \inf\{\epsilon \geq 0 \text{ such that } N(A, \epsilon) \supset B \text{ and } N(B, \epsilon) \supset A\},$$

where $N(A, \epsilon)$ denotes the ϵ-neighborhood of the set A, that is,

$$N(A, \epsilon) = \{x \in X \text{ such that } d_A(x) \leq \epsilon\}.$$

Equivalently, we have

$$d_{\mathcal{H}}(A, B) = \sup_{x \in X} |d_A(x) - d_B(x)|,$$

where $d_A(x)$ is the distance from x to the set A.

This notion of "distance" between subsets of a metric space is a useful tool in topology, but in general, it does not satisfy the axioms of a metric. For instance, for every subset A of X, we have $d_{\mathcal{H}}(A, \bar{A}) = 0$, therefore $d_{\mathcal{H}}$ does not necessarily separate points. Furthermore, If A is bounded and B unbounded, then their Hausdorff distance is infinite. This last property is certainly the major inconvenience of the function $d_{\mathcal{H}}$. To remedy to this, one considers the set $\mathcal{F}(X)$ of nonempty closed subsets of X and one chooses a point p in X and defines a map d_p on the product $\mathcal{F}(X) \times \mathcal{F}(X)$ by setting, for each A and B in $\mathcal{F}(X)$,

$$d_p(A, B) = \sup_{x \in X} |d_A(x) - d_B(x)| e^{-|p-x|}.$$

This map is a genuine metric on $\mathcal{F}(X)$. The family $(d_p)_{p \in X}$ of metrics on $\mathcal{F}(X)$ has been considered by Busemann, and in this chapter, we shall treat it in some detail. Since each metric in this family is defined by scaling the Hausdorff distance, we call it a *Busemann–Hausdorff metric*. For any p and q in X, the metrics d_p and d_q are commensurable.

There are several notions of limits of subsets in a metric space that are related to Hausdorff distance and to Busemann–Hausdorff distances, and we shall consider them here.

Given a sequence of nonempty subsets $(A_n)_{n \geq 0}$ of subsets of X, its *lower closed limit* is the set of points x in X such that every neighborhood of x contains points in all but finitely many sets A_n, and its *upper closed limit* is the set of points x in X such that every neighborhood of x contains points in infinitely many sets A_n. We say that (A_n) has a closed limit if the lower and upper closed limits of this sequence are equal, and in this case the *closed limit* is the common value of the lower and upper closed limits.

We shall see some relations between convergence with respect to the Hausdorff distance and the existence of lower and upper limits. For instance, if $d_{\mathcal{H}}(A_n, A) \to 0$ where A is some bounded subset of X, then $\lim A_n = \bar{A}$. In the case where the metric space X is compact, we have the following converse: if (A_n) has a nonempty closed limit A, then $d_{\mathcal{H}}(A_n, A) \to 0$.

There are analogous results for the Busemann–Hausdorff distance. Given a point p in X, if $d_p(A_n, A) \to 0$ for some subset A of X, then $\lim A_n = \bar{A}$. In the case where X is proper, the following converse holds: if $\lim A_n = A$, then $d_p(A_n, A) \to 0$.

In the same spirit as for the distances d and d_p (for $p \in X$) on subsets, there are distances d and d_p ($p \in X$) on the isometry group $\mathrm{Isom}(X)$ of a metric space X. The definitions are as follows.

For f and g in $\mathrm{Isom}(X)$ and for any point p in X, we set

$$d(f, g) = |f - g| = \sup_{x \in X} |f(x) - g(x)|$$

and

$$d_p(f, g) = |f - g|_p = \sup_{x \in X} |f(x) - g(x)| e^{-|p-x|}.$$

In the special case where X is compact, the map d is a metric on $\mathrm{Isom}(X)$. For any metric space X and for each p in X, the map d_p is a metric on $\mathrm{Isom}(X)$. For any p and q in X, the metrics d_p and d_q are commensurable.

The outline of this chapter is as follows.

In Section 1, we start by recalling some basic facts about the distance from a point to a subset, about ϵ-neighborhoods of subsets and about the Hausdorff distance between two subsets of a metric space X. Restricted to the set $\mathcal{B}(X)$ of nonempty closed bounded subsets of X, the Hausdorff distance is a genuine distance (that is, it satisfies the axioms of a metric space).

In Section 2, we consider the family of Busemann–Hausdorff distances $(d_p)_{p\in X}$ on the set $\mathcal{F}(X)$ of all nonempty closed subsets of X. We prove that for any p and q in X, the metrics d_p and d_q are commensurable.

In Section 3, we consider the notions of upper closed limit, lower closed limit and closed limit of a sequence of subsets of an arbitrary metric space and we make the relation between these notions and the Hausdorff and the Busemann–Hausdorff distances.

Section 4 concerns the distances d and d_p $(p \in X)$ on the space $\mathrm{Isom}(X)$.

4.1 The Hausdorff distance

Let X be a metric space. We recall that if x is a point in X and if A is a subset of X, then the distance from x to A is defined as

$$d_A(x) = \inf_{y\in A} d(x, y).$$

We start with a few elementary properties of the map $x \mapsto d_A(x)$.

Proposition 4.1.1. *If A and B are two nonempty subsets of X, then*

(i) $d_A(x) = d_{\bar{A}}(x)$ *for all* $x \in X$;

(ii) $d_A(x) = 0 \iff x \in \bar{A}$;

(iii) $d_A(x) = d_B(x)$ *for all* $x \in X \iff \bar{A} = \overline{B}$.

Proof. Properties (i) and (ii) follow trivially from the definitions. To prove (iii), suppose that $\bar{A} \neq \overline{B}$. Up to interchanging the names of A and B, we can assume that there is a point x in $\bar{A} \setminus \overline{B}$. Then, by (i), we have $d_A(x) = 0$ and $d_B(x) \neq 0$, which shows that $d_A(x) \neq d_B(x)$. The converse implication is trivial. □

Proposition 4.1.2 (The map d_A is non-expanding). *For any metric space X and for any nonempty subset A of X, the map $x \mapsto d_A(x)$ is non-expanding.*

Proof. For any x and y in X and for any z in A, we have, by the triangle inequality,

$$|d(x, z) - d(y, z)| \leq d(x, y).$$

Taking the infimum over z in A, we obtain

$$|d_A(x) - d_A(y)| \leq d(x, y),$$

which is the required result. □

Definition 4.1.3 (Closed ϵ-neighborhood of a set). Let X be a metric space, let A be a subset of X and let ϵ be in $[0, \infty[$. The closed ϵ-neighborhood of A is the subset of X defined by

$$N(A, \epsilon) = \{x \in X \text{ such that } d_A(x) \leq \epsilon\}.$$

Equivalently, we have

$$N(A, \epsilon) = \bigcup_{x \in A} B(x, \epsilon)$$

where for each x in A, $B(x, \epsilon)$ denotes the closed ball in X of center x and radius ϵ.

Notice that in the case where A consists of a single point x, the set $N(A, \epsilon)$ is the closed ball of center x and radius ϵ.

It is clear from the definition that if A and B are subsets of X satisfying $A \subset B$, then for any $\epsilon \geq 0$ we have $N(A, \epsilon) \subset N(B, \epsilon)$.

Notice also that the closed ϵ-neighborhood of any subset A is closed in X, since it is the inverse image of the closed interval $[0, \epsilon]$ by the continuous map $d_A \colon X \to \mathbb{R}$. In Figure 4.1, we have drawn the closed ϵ-neighborhood of an (open or closed) segment in $\mathbb{R}^2$. This is an example of an object that Busemann calls a "capsule" (see Definition 9.2.6 below).

Figure 4.1. A capsule.

Proposition 4.1.4. *Let X be a metric space. Then, for any subset A of X and for any nonnegative real numbers ϵ_1 and ϵ_2, we have $N\big(N(A, \epsilon_1), \epsilon_2\big) \subset N(A, \epsilon_1 + \epsilon_2)$. In the case where X is a geodesic space, we have $N\big(N(A, \epsilon_1), \epsilon_2\big) = N(A, \epsilon_1 + \epsilon_2)$.*

Proof. The proof follows easily from the definitions. □

Proposition 4.1.5. *Let X be a metric space. For any $A \subset X$ and for any $\epsilon \geq 0$, we have $N(\bar{A}, \epsilon) = N(A, \epsilon)$.*

Proof. Let x be a point in $N(\bar{A}, \epsilon)$. Since $\bar{A}$ is closed, we can find a point $\overline{x}$ in $\bar{A}$ such that $|x - \overline{x}| \leq \epsilon$. Since $\overline{x}$ is in $\bar{A}$, for every integer $n \geq 1$, there exists a point x_n in A satisfying $|\overline{x} - x_n| \leq \epsilon$. Thus, we have

$$|x - x_n| \leq |x - \overline{x}| + |\overline{x} - x_n| \leq \epsilon + 1/n,$$

which shows that $d_A(x) \leq \epsilon$. Thus, x is in $N(A, \epsilon)$. This proves that $N(\bar{A}, \epsilon) \subset N(A, \epsilon)$. The other inclusion is obvious. □

Proposition 4.1.6. *Let X be a proper metric space, let A and B be two subsets of X and let $\delta = d_{\mathcal{H}}(A, B)$. Then, we have $A \subset N(B, \delta)$ and $B \subset N(A, \delta)$.*

Proof. Without loss of generality, we can assume that δ is finite. Furthermore, since $N(A, \delta) = N(\bar{A}, \delta)$ and $N(B, \delta) = N(\overline{B}, \delta)$ (Proposition 4.1.5), we can assume without loss of generality that A and B are closed. Let x be an element of A. For any integer $n \geq 1$, we have $X \subset N(B, \delta + 1/n)$. Since B is closed, there exists $y_n \in B$ such that $|x - y_n| \leq \delta + 1/n$. Since X is proper, B is also proper, and the sequence $(y_n)_{n \geq 1}$ has a convergent subsequence. If y is the limit of such a subsequence, we have $|x - y| \leq \delta$, which shows that x is in $N(B, \delta)$. Thus, we obtain $A \subset N(B, \delta)$. By symmetry, we also have $B \subset N(A, \delta)$. □

Definition 4.1.7 (Hausdorff distance). Let X be a metric space and let A and B be two nonempty subsets of X. The *Hausdorff distance* $d_{\mathcal{H}}(A, B)$ is defined as

$$d_{\mathcal{H}}(A, B) = \inf\{\epsilon \geq 0 \text{ such that } N(A, \epsilon) \supset B \text{ and } N(B, \epsilon) \supset A\}.$$

We start with the following proposition whose proof follows easily from the definitions.

Proposition 4.1.8. *Let X be a metric space. For any subsets A and B of X we have*

(i) $0 \leq d_{\mathcal{H}}(A, B) \leq \infty$, *and if A and B are bounded,* $d_{\mathcal{H}}(A, B) < \infty$;

(ii) $d_{\mathcal{H}}(A, B) = d_{\mathcal{H}}(B, A)$;

(iii) $d_{\mathcal{H}}(A, \bar{A}) = 0$;

(iv) $d_{\mathcal{H}}(A, B) = d_{\mathcal{H}}(A, \overline{B})$;

(v) *if* $A = \{x\}$ *and* $B = \{y\}$, *then* $d_{\mathcal{H}}(A, B) = |x - y|$. □

Notice that we can measure Hausdorff distances between objects of apparently different nature. For instance, if $X = \mathbb{R}$ and if we take $A = \mathbb{R}$ and $B = \mathbb{N}$, then $d_{\mathcal{H}}(A, B) = 1$.

It is easy to see that if X is an arbitrary metric space and if A is a bounded and B an unbounded subset of X, then $d_{\mathcal{H}}(A, B) = \infty$.

We now give a set of examples of subsets of metric spaces that are at mutual finite Hausdorff distance and others at mutual infinite Hausdorff distance. In later chapters, we shall consider Hausdorff distances between images of geodesic rays and of geodesic lines, and therefore we give examples of distances between such subsets in the classical spaces.

Examples 4.1.9 (Hausdorff distance).

(i) *Disks in Euclidean space.* If A and B are open or closed disks in $\mathbb{E}^n$ of centers respectively x_1 and x_2 and of radii respectively r_1 and r_2, then $d_{\mathcal{H}}(A, B) = |x_1 - x_2| + |r_1 - r_2|$.

(ii) *Geodesic rays and geodesic lines in Euclidean space.* The Hausdorff distance between two straight lines in $\mathbb{E}^n$ is finite if and only if these straight lines coincide or if they span a 2-dimensional plane in $\mathbb{E}^n$ in which they are parallel in the sense of Euclidean plane geometry. The Hausdorff distance between the images of two geodesic rays in $\mathbb{E}^n$ is finite if and only if these images are contained in parallel straight lines and if in a plane that contains them, the geodesic rays have the same direction.

(iii) *Geodesic rays and geodesic lines in hyperbolic space.* Let B^n be the conformal ball model of hyperbolic space $\mathbb{H}^n$ and let $r_1 \colon [0, \infty[\to B^n$ and $r_2 \colon [0, \infty[\to B^n$ be two geodesic rays. Then, $d_{\mathcal{H}}(\mathrm{Im}(r_1), \mathrm{Im}(r_2))$ is finite if and only if $r_1(t)$ and $r_2(t)$ converge as $t \in \infty$ (with respect to the Euclidean metric of the ball) to the same point on the boundary sphere S^{n-1} of B^n. The proof can be done by a calculation using the formula for the distance in B^n. In fact, if $r_1 \colon [0, \infty[\to B^n$ and $r_2 \colon [0, \infty[\to B^n$ are two geodesic rays satisfying $r_1(\infty) \neq r_2(\infty)$, then the projection of $r_1(t)$ on $\mathrm{Im}(r_2)$ tends to infinity as $t \to \infty$. From this fact, we deduce that the Hausdorff distance between the two straight lines in $\mathbb{H}^n$ is finite if and only if the two lines have the same set of limit points in S^n, or, equivalently, if the images of these lines coincide.

(iv) *Geodesic rays and geodesic lines in an $\mathbb{R}$-tree.* If T is an $\mathbb{R}$-tree and if $r_1 \colon [0, \infty[\to T$ and $r_2 \colon [0, \infty[\to T$ are two geodesic rays, then $d_{\mathcal{H}}(\mathrm{Im}(r_1), \mathrm{Im}(r_2)) < \infty$ if and only if $\mathrm{Im}(r_1)$ and $\mathrm{Im}(r_2)$ coincide up to a compact set, that is, if and only if there exist two nonnegative real numbers t_1 and t_2 such that $r_1([t_1, \infty[) = r_2([t_2, \infty[)$. The case of straight lines in T is identical to the case of straight lines in $\mathbb{H}^n$: the Hausdorff distance between the images is finite if and only if these two images coincide.

We already saw that for any subset A of X, we have $d_{\mathcal{H}}(A, \bar{A}) = 0$. Thus, even if we restrict the map $d_{\mathcal{H}}$ to bounded subsets, $d_{\mathcal{H}}$ does not satisfy the axioms of a metric. We shall see below that it does satisfy them if we restrict the map to the set of *closed* bounded subsets of X.

The following characterization of the Hausdorff distance is contained in [25].

Proposition 4.1.10. *Let X be a metric space and let A and B be two nonempty subsets of X. Then,*

$$d_{\mathcal{H}}(A, B) = \sup_{x \in X} |d_A(x) - d_B(x)|. \tag{4.1.10.1}$$

Proof. Suppose first that $d_{\mathcal{H}}(A, B) = 0$. Then, for any $\epsilon > 0$, we have $A \subset N(B, \epsilon)$. This shows that $A \subset \overline{B}$ and since $\overline{B}$ is closed, we obtain $\bar{A} \subset \overline{B}$. By symmetry, we also have $\overline{B} \subset \bar{A}$. Therefore, $\bar{A} = \overline{B}$. By Proposition 4.1.1 (iii), this implies $d_A(x) = d_B(x)$ for all x in X. Thus, (4.1.10.1) is satisfied in the case where $d_{\mathcal{H}}(A, B) = 0$.

Suppose now that $d_{\mathcal{H}}(A, B) = \infty$. Then, up to interchanging the names of A and B, we can assume that for each nonnegative integer n, we can find a point x_n in A that is not contained in $N(B, n)$. Thus, $d_B(x_n) \geq n$ and since $d_A(x_n) = 0$, we obtain $\sup_{x \in X} |d_A(x) - d_B(x)| = \infty$. Thus, (1.1.10.1) is also satisfied in this case.

It remains to consider the case where $0 < d_{\mathcal{H}}(A, B) < \infty$. Let n be a positive integer satisfying $d_{\mathcal{H}}(A, B) > 1/n$. Up to interchanging the names of A and B, we can assume that $A \not\subset N\big(B, d_{\mathcal{H}}(A, B) - 1/n\big)$. Then, we can find a point z in A satisfying $d_{\mathcal{H}}(A, B) - 1/n < d_B(z)$. Since $d_A(z) = 0$, we have

$$d_{\mathcal{H}}(A, B) - 1/n < d_B(z) = |d_A(z) - d_B(z)| \leq \sup_{x \in X} |d_A(x) - d_B(x)|.$$

Letting n tend to infinity, we obtain $d_{\mathcal{H}}(A, B) \leq \sup_{x \in X} |d_A(x) - d_B(x)|$.

To prove the converse inequality, let x be an arbitrary point in X and let n be a positive integer. We can find a point z in B such that $|x - z| < d_B(x) + 1/n$. Since z is also in $N(A, d_{\mathcal{H}}(A, B) + 1/n)$, we can find a point y in A satisfying $|z - y| \leq d_{\mathcal{H}}(A, B) + 1/n$. Thus, for every x in X, we have

$$\begin{aligned} d_A(x) - d_B(x) &\leq |x - y| - d_B(x) \\ &\leq |x - z| + |z - y| - d_B(x) \\ &\leq d_B(x) + 1/n + d_{\mathcal{H}}(A, B) + 1/n - d_B(x) \\ &= d_{\mathcal{H}}(A, B) + 2/n. \end{aligned}$$

Letting n tend to infinity, we obtain $d_A(x) - d_B(x) \leq d_{\mathcal{H}}(A, B)$. By symmetry, we also have $d_B(x) - d_A(x) \leq d_{\mathcal{H}}(A, B)$. We conclude that $|d_A(x) - d_B(x)| \leq d_{\mathcal{H}}(A, B)$ for every x in X. Summing up, we have $|d_A(x) - d_B(x)| = d_{\mathcal{H}}(A, B)$, as required. This completes the proof of Proposition 4.1.10. □

For any metric space X, we denote by $\mathcal{B}(X)$ the set of nonempty closed bounded subsets of X. Of course, if X is proper, then $\mathcal{B}(X)$ is the set of nonempty compact subsets of X

Proposition 4.1.11 ($d_{\mathcal{H}}$ is a metric on $\mathcal{B}(X)$). *For any metric space X, the map $(A, B) \mapsto d_{\mathcal{H}}(A, B)$ defines a metric on $\mathcal{B}(X)$.*

Proof. It is plain that for any A and B in $\mathcal{B}(X)$ we have $0 \leq d_{\mathcal{H}}(A, B) < \infty$, $d_{\mathcal{H}}(A, B) = d_{\mathcal{H}}(B, A)$ and $A = B \Rightarrow d_{\mathcal{H}}(A, B) = 0$. Let us show that

$$d_{\mathcal{H}}(A, B) = 0 \Rightarrow A = B.$$

By Proposition 4.1.10, we have

$$d_{\mathcal{H}}(A, B) = 0 \Rightarrow d_A(x) = d_B(x) \quad \text{for all } x \in X.$$

By Proposition 4.1.1 (iii), this implies $\bar{A} = \overline{B}$ and since A and B are closed we obtain $A = B$.

It remains to prove the triangle inequality. Let x be an arbitrary point in X and let A, B and C be in $\mathcal{B}(X)$. We have, by the triangle inequality in $\mathbb{R}$,

$$|d_A(x) - d_C(x)| \leq |d_A(x) - d_B(x)| + |d_B(x) - d_C(x)|.$$

Taking the supremum over all points x in X, we obtain, using Proposition 4.1.10,

$$d_{\mathcal{H}}(A, C) \leq d_{\mathcal{H}}(A, B) + d_{\mathcal{H}}(B, C).$$

This proves Proposition 4.1.11. □

Remark. If instead of the set $\mathcal{B}(X)$ we consider the set of bounded subsets of X, then $d_{\mathcal{H}}$ defines a pseudo-metric.

4.2 The Busemann–Hausdorff distance

We already noted that if one of the two subsets A or B of a metric space X is unbounded, then the Hausdorff distance $d_{\mathcal{H}}(A, B)$ may be infinite. We need a distance function that is more adapted to unbounded subsets. This is all the more useful because some of the subsets that we shall deal with are images of geodesic rays and geodesic lines, and we already saw that in Euclidean space, if A and B are the images of two geodesic rays, then $d_{\mathcal{H}}(A, B) = \infty$ unless A and B are contained in parallel Euclidean lines. To obtain a distance function whose value is always finite, Busemann uses in [25] a modified version of the Hausdorff distance, which we call here the Busemann–Hausdorff distance. The definition of this distance is an adaptation of the Hausdorff distance in the version which is provided by Proposition 4.1.10.

Definition 4.2.1 (Busemann–Hausdorff distance). Let X be a metric space, let p be a point in X and let A and B be two nonempty subsets of X. The *Busemann–Hausdorff distance with basepoint p*, between A and B, is defined by

$$d_p(A, B) = \sup_{x \in X} |d_A(x) - d_B(x)| e^{-|p-x|}.$$

Proposition 4.2.2. *Let X be a metric space and let p be a point in X. Then for any two nonempty subsets A and B of X, the Busemann–Hausdorff distance $d_p(A, B)$ is finite.*

Proof. Since the maps d_A and d_B are non-expanding (Proposition 4.1.2), we have, for every x in X,

$$\begin{aligned}|d_A(x) - d_B(x)| &\leq d_A(x) + d_B(x) \\ &\leq d_A(p) + |p - x| + d_B(p) + |p - x|.\end{aligned}$$

Using the fact that $e^{-|p-x|} < 1$ and $|p - x|e^{-|p-x|} < 1$, we obtain

$$\begin{aligned}|d_A(x) - d_B(x)|e^{-|p-x|} &\leq \big(d_A(p) + d_B(p) + 2|p - x|\big)e^{-|p-x|} \\ &\leq d_A(p) + d_B(p) + 2.\end{aligned}$$

Thus, we obtain $d_p(A, B) < \infty$, as required. □

Given a metric space X, we denote by $\mathcal{F}(X)$ the set of nonempty closed subsets of X.

Proposition 4.2.3 (d_p is a metric on $\mathcal{F}(X)$). *For any metric space X and for any p in X, the map $(A, B) \mapsto d_p(A, B)$ is a metric on $\mathcal{F}(X)$.*

Proof. It is clear that d_p is symmetric and that $A = B \Rightarrow d_p(A, B) = 0$. Let us show that $d_p(A, B) = 0 \Rightarrow A = B$. The proof is an adaptation of the proof of Proposition 4.1.10.

From the definition, we have

$$d_p(A, B) = 0 \iff d_A(x) = d_B(x) \quad \text{for all } x \in X.$$

Assume that $A \neq B$. Up to interchanging the names of A and B, we can suppose that there exists a point x in $B \setminus A$. Then, $d_B(x) = 0$ and, since A is closed, $d_A(x) \neq 0$. Therefore we have

$$d_p(A, B) \geq |d_A(x) - d_B(x)|e^{-|p-x|} > 0.$$

Finally let us prove that d_p satisfies the triangle inequality. For every A, B and C in $\mathcal{F}(X)$ and for every positive integer n, we can find a point x in X such that

$$\begin{aligned}d_p(A, C) - 1/n &< |d_A(x) - d_C(x)|e^{-|p-x|} \\ &\leq \big(|d_A(x) - d_B(x)| + |d_B(x) - d_C(x)|\big)e^{-|p-x|} \\ &\leq |d_A(x) - d_B(x)|e^{-|p-x|} + |d_B(x) - d_C(x)|e^{-|p-x|}.\end{aligned}$$

Thus, we obtain, for all $n \geq 1$, $d_p(A, C) - 1/n \leq d_p(A, B) + d_p(B, C)$. Letting n tend to infinity, we obtain

$$d_p(A, C) \leq d_p(A, B) + d_p(B, C).$$

This completes the proof of Proposition 4.2.3. □

Two metrics d_1 and d_2 defined on a metric space X are said to be *commensurable* if the identity map $(X, d_1) \to (X, d_2)$ is bi-Lipschitz, that is, if there exist two constants C_1 and C_2 such that for every x and y in X we have $d_1(x, y) \le C_1 d_2(x, y)$ and $d_2(x, y) \le C_2 d_1(x, y)$. We recall that two commensurable metrics define the same topology.

Proposition 4.2.4. *Let X be a metric space. For any p and q in X, the two metrics d_p and d_q on $\mathcal{F}(X)$ are commensurable.*

Proof. For each x in X and for each A and B in $\mathcal{F}(X)$, we have, by the triangle inequality in $\mathbb{R}$,

$$|p - x| - |p - q| \le |q - x|$$

which implies

$$|d_A(x) - d_B(x)| e^{-|q-x|} \le |d_A(x) - d_B(x)| e^{-|p-x|+|p-q|}.$$

Taking the supremum over x in X, we obtain

$$d_q(A, B) \le d_p(A, B) e^{|p-q|}.$$

By symmetry, we also have

$$d_p(A, B) \le d_q(A, B) e^{|p-q|}.$$

This proves Proposition 4.2.4. □

4.3 Closed limits of subsets

This section concerns upper closed limits and lower closed limits of sequences of subsets in a metric space and their relation to convergence with respect to the Hausdorff and the Busemann–Hausdorff metrics. We start with the following definition, which is due to Hausdorff (see [66] p. 168):

Definition 4.3.1 (Upper and lower closed limits of subsets). Let X be a metric space and let $(A_n)_{n \ge 0}$ be a sequence of nonempty subsets of X. The *lower closed limit* of (A_n), denoted by $\liminf A_n$ is the set of points x in X such that every neighborhood of x contains points in all but finitely many sets A_n. The *upper closed limit* of (A_n), denoted by $\limsup A_n$, is the set of points x in X such that every neighborhood of x contains points in infinitely many sets A_n.

Equivalently, $\liminf A_n$ is the set of limit points of sequences $(x_n)_{n \ge 0}$ with $x_n \in A_n$ for all $n \ge 0$, and $\limsup A_n$ is the set of accumulation points of such sequences.

It is easy to see from these definitions that $\limsup A_n$ and $\liminf A_n$ are closed subsets of X and that we always have

$$\liminf A_n \subset \limsup A_n$$

Definition 4.3.2 (Closed limit of subsets). Let X be a metric space and let (A_n) be a sequence of nonempty subsets of X. If $\liminf A_n = \limsup A_n$, then we say that the sequence (A_n) *has a closed limit*, which we denote by $\lim A_n$, and which is equal to that common value:

$$\lim A_n = \liminf A_n = \limsup A_n.$$

It is easy to see that if $(A_n)_{n\geq 0}$ is an increasing sequence of subsets, *i.e.* if

$$A_0 \subset A_1 \cdots \subset A_n \subset \cdots,$$

then $\lim A_n = \bigcup_{n\geq 0} A_n$. Likewise, if $(A_n)_{n\geq 0}$ is a decreasing sequence, *i.e.* if

$$A_0 \supset A_1 \cdots \supset A_n \supset \cdots,$$

then $\lim A_n = \bigcap_{n\geq 0} A_n$.

In the sequel, given a sequence (A_n) of subsets of a metric space, then the fact of writing $\lim A_n = A$ will imply that the closed limit of (A_n) exists and is equal to A.

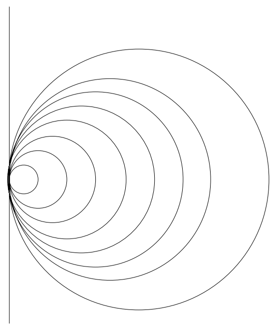


Figure 4.2. In the Euclidean plane, the vertical line is the closed limit of the sequence of circles.

Examples 4.3.3 (Limits of subsets). In the first two examples that follow, the ambient space is the real line.

(i) For every integer $n \geq 0$, let $A_n = \{n\}$. Then $\limsup A_n = \liminf A_n = \emptyset$. Therefore, $\lim A_n = \emptyset$.

(ii) For every integer $n \geq 0$, let $A_{2n} = \{1\}$ and $A_{2n+1} = \{1/(2n+1)\}$. Then $\limsup A_n = \{0, 1\}$ and $\liminf A_n = \emptyset$. Therefore, $\lim A_n$ does not exist.

(iii) If X is an arbitrary metric space, if $(x_n)_{n\geq 0}$ is a sequence of points in X and if for any $n \geq 0$ $A_n = \{x_n\}$, then $\lim A_n$ exists if and only if the sequence (x_n) converges as $n \to \infty$, and in that case $\lim A_n = \lim_{n\to\infty} x_n$.

(iv) Let A and B be two subsets of a metric space X and for every $n \geq 0$, let $A_{n+1} = A$ and $A_{n+2} = B$. Then, $\liminf A_n = A \cap B$ and $\limsup A_n = A \cup B$.

Thus, the sequence (A_n) has a closed limit if and only if $A \cup B = A \cap B$, that is, if and only if $A = B$.

(v) For any metric space X, choose a point x_0 in X and a sequence $(r_n)_{n\geq 0}$ of positive real numbers tending to infinity. For any $n \geq 0$, let B_n be the open (or closed) ball of center at x_n and radius r_n. Then $\lim B_n = X$.

(vi) In the 2-dimensional Euclidean space $\mathbb{E}^2$, let D be a straight line, let p be a point on D, let $(r_n)_{n\geq 0}$ be an increasing sequence of positive real numbers tending to infinity and let C_n be a sequence of circles which are tangent to D at p and of radius r_n (Figure 4.2). Then, $\lim C_n = D$.

(vii) Let X be a metric space, let $r\colon [0,\infty[\to X$ be a geodesic ray, let $(t_n)_{n\geq 0}$ be a sequence of positive numbers tending to infinity and let $\gamma_n = r_{|[0,t_n]}$. Then, $\lim \mathrm{Im}(\gamma_n) = \mathrm{Im}(r)$.

Example 4.3.3 (i) shows that the closed limit of a sequence of subsets can be the empty set. We shall use the term “nonempty closed limit” to denote this limit in the case where it exists and is not empty. It follows easily from the definitions that if $\limsup A_n = \emptyset$ then $\lim A_n$ exists and is equal to the empty set. The following properties of limits of subsets are also easy to prove:

For any sequence (A_n) of subsets of X, we have

$$\limsup A_n = \limsup \overline{A_n}$$

and

$$\liminf A_n = \liminf \overline{A_n}.$$

Consequently, the sequence (A_n) has a nonempty closed limit if and only if the sequence of closures $(\overline{A_n})$ has a closed limit.

If $(A_{n_i})_{i\geq 0}$ is a subsequence of $(A_n)_{n\geq 0}$, then

$$\liminf A_n \subset \liminf A_{n_i} \subset \limsup A_{n_i} \subset \limsup A_n.$$

Thus, if a sequence (A_n) has a closed limit, then any subsequence of (A_n) has the same closed limit.

Let us now give a few more examples:

Examples 4.3.4 (Limits of subsets in hyperbolic space). In the following examples, B^n is the conformal ball model of hyperbolic space $\mathbb{H}^n$ and S^{n-1} is its boundary.[1]

(i) *Closed limits of geodesic lines in* $\mathbb{H}^n$. Let $g_n\colon \mathbb{R} \to B^n$ be a sequence of geodesic lines. Then, the sequence $(\mathrm{Im}(g_n))$ has a closed limit if and only if the

[1] Convergence of sequences points in $B^n \cup S^{n-1}$ is understood here in terms of the metric induced from the inclusion of these spaces in Euclidean space $\mathbb{E}^n$.The reader probably knows that this is unnatural, and that it is possible to discuss convergence of sequences of points in hyperbolic space $\mathbb{H}^n$ to points on the boundary without making any reference to any model of $\mathbb{H}^n$, in the way we do it in Chapter 10 below. But here, in order to avoid a long digression, we stick to this (unnatural) setting of B^n and S^{n-1} included in $\mathbb{E}^n$.

sequence $\big(g_n(-\infty), g_n(\infty)\big)$ of corresponding endpoints in $S^{n-1} \times S^{n-1}$ is convergent. The closed limit of $(\mathrm{Im}(g_n))$, when it exists, is nonempty if and only if the sequence $\big(g_n(-\infty), g_n(\infty)\big)$ converges to a point (p_1, p_2) in $S^{n-1} \times S^{n-1}$ satisfying $p_1 \neq p_2$. In this case, the closed limit of $(\mathrm{Im}(g_n))$ is the straight line in $\mathbb{H}^n$ whose endpoints are p_1 and p_2.

(ii) *Closed limits of spheres in* $\mathbb{H}^n$. For any point x in a metric space X and for any nonnegative real number r, the sphere of center x and radius r is the set of points in X whose distance at x is equal to r. It is a classical fact that in B^n, a sphere (with respect to the hyperbolic metric) coincides, as a subset of B^n, with a Euclidean sphere (which in general has a different center and a different radius). Let $r\colon [0, \infty[\to B^n$ be a geodesic ray and for each integer $n \geq 0$, let S_n be the sphere of center $r(n)$ and radius n. Any such sphere contains the point $r(0)$. Then, the sequence $(S_n)_{n\geq 0}$ has a closed limit which, as a subset of B^n, is the Euclidean sphere whose diameter is the segment $[r(0), r(\infty)]$ with the point $r(\infty)$ deleted (see Figure 4.3). This limit is an unbounded set with respect to the hyperbolic metric, and it is called the *horosphere* with central ray r passing through the point $r(0)$. We shall deal with such objects in Chapter 12.

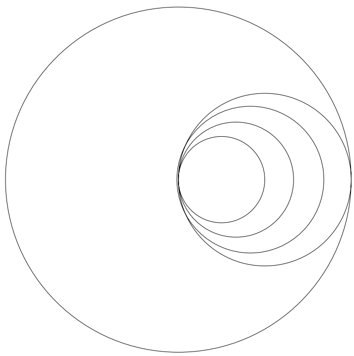


Figure 4.3. A horosphere in B^n as a limit of spheres.

The following two results are due to Hausdorff ([66] p. 171 & 172). They make the relation between convergence with respect to the Hausdorff metric and the existence of a closed limit.

Proposition 4.3.5. *Let X be a metric space, let $(A_n)_{n\geq 0}$ be a sequence of subsets of X and let A be a bounded subset of X. Then, we have*

$$d_{\mathcal{H}}(A_n, A) \to 0 \Rightarrow \lim A_n = \bar{A}.$$

Proof. Suppose that $d_{\mathcal{H}}(A_n, A) \to 0$. We prove that

$$\limsup A_n \subset \bar{A} \subset \liminf A_n.$$

Since we always have $\liminf A_n \subset \limsup A_n$, this will imply $\bar{A} = \liminf A_n = \limsup A_n$.

Let x be in $\bar{A}$ and let ϵ be an arbitrary positive real number. For n large enough, we have $d_{\mathcal{H}}(A_n, A) = d_{\mathcal{H}}(A_n, \bar{A}) < \epsilon$, which implies that $x \in N(A_n, \epsilon)$. Thus, we can find a point x_n in A_n such that $d_{\mathcal{H}}(x, x_n) \leq \epsilon$. Therefore, $x \in \liminf A_n$. This proves that $\bar{A} \subset \liminf A_n$.

Now let x be a point in $\limsup A_n$. Then there exists a sequence n_i $(i = 0, 1, \dots)$ of integers tending to infinity and for every $i \geq 0$ a point x_{n_i} in A_{n_i} such that $|x_{n_i} - x| < 1/n_i$. Since $d_{\mathcal{H}}(A_n, A) \to 0$, up to replacing the sequence A_{n_i} by a subsequence, we can find, for every $i \geq 0$, a point y_{n_i} in A such that $|x_{n_i} - y_{n_i}| < 1/n_i$. Thus, we also have $y_{n_i} \to x$ as $i \to \infty$, which implies that x is in $\bar{A}$. This proves that $\limsup A_n \subset \bar{A}$, which completes the proof of Proposition 4.3.5. □

The converse of Proposition 4.3.5 is false without further assumptions, as we can see by taking $X = \mathbb{R}$ and $A_n = [-n, n]$ for all $n \geq 0$. Here, $\lim A_n = \mathbb{R}$ and $d_{\mathcal{H}}(A_n, \mathbb{R}) = \infty$ for all n. However, we have the following partial converse:

Proposition 4.3.6. *Let X be a compact metric space and let $(A_n)_{n\geq 0}$ be a sequence of subsets of X that has a nonempty closed limit A. Then, $d_{\mathcal{H}}(A_n, A) \to 0$.*

Proof. We reason by contradiction. Suppose that $d_{\mathcal{H}}(A_n, A) \not\to 0$. Then, there exists a positive real number ρ and a subsequence $(A_{n_i})_{i\geq 0}$ of $(A_n)_{n\geq 0}$ such that $d_{\mathcal{H}}(A_{n_i}, A) > \rho$ for every $i \geq 0$. Thus, for every $i \geq 0$, we can find a point x_{n_i} in A satisfying $d_{A_{n_i}}(x_{n_i}) > \rho$. By compactness of X, up to passing to a subsequence, we can assume that the sequence x_{n_i} converges to a point x in X, and since A is closed, x is in A. Since the map $x \mapsto d_{A_{n_i}}(x)$ is 1-Lipschitz, we have

$$|d_{A_{n_i}}(x_{n_i}) - d_{A_{n_i}}(x)| \leq |x_{n_i} - x|.$$

Since $x_{n_i} \to x$, we have $d_{A_{n_i}}(x) \to 0$ as $i \to \infty$, which implies $d_{A_{n_i}}(x_{n_i}) \to 0$, which contradicts the fact that $d_{A_{n_i}}(x_{n_i}) > \rho$. This proves Proposition 4.3.6. □

The following two propositions give a relation between convergence with respect to the Busemann–Hausdorff metric and the existence of a closed limit for sequences of subsets. The results and the proofs are also due to Busemann (cf. [28], §3).

Proposition 4.3.7. *Let $(A_n)_{n\geq 0}$ be a sequence of nonempty subsets of a metric space X, let p be a point in X and let A be a subset of X. Then we have*

$$d_p(A_n, A) \to 0 \Rightarrow \lim A_n = \bar{A}.$$

Proof. Suppose that $d_p(A_n, A) \to 0$ and let us prove that

$$\limsup A_n \subset \bar{A} \subset \liminf A_n.$$

Let x be a point in $\bar{A}$. We have

$$d_{A_n}(x)e^{-|p-x|} = |d_{A_n}(x) - d_A(x)|e^{-|p-x|} \leq d_p(A_n, A).$$

Therefore, $d_{A_n}(x) \to 0$ as $n \to \infty$. Thus, for every $n \geq 0$, we can find a point y_n in A_n satisfying $|x - y_n| \to 0$ as $n \to \infty$. This implies that $x \in \liminf A_n$. Thus, we have $\bar{A} \subset \liminf A_n$.

We now prove that $\limsup A_n \subset \bar{A}$. For every y in $\limsup A_n$, we can find a subsequence $(A_{n_i})_{i\geq 0}$ of $(A_n)_{n\geq 0}$ such that for every $i \geq 0$, there exists a point y_{n_i} in A_{n_i} with $y_{n_i} \to y$ as $i \to \infty$. We have

$$d_p(A_{n_i}, A)e^{|p-y|} \geq |d_{A_{n_i}}(y) - d_{\bar{A}}(y)|.$$

As $i \to \infty$, we have $d_p(A_{n_i}, A) \to 0$, which implies $|d_{A_{n_i}}(y) - d_{\bar{A}}(y)| \to 0$. Since $|y - y_{n_i}| \to 0$, we obtain $d_{A_{n_i}}(y) \to 0$ which gives $d_{\bar{A}}(y) \to 0$ and therefore $y \in \bar{A}$. This proves that $\liminf A_n \subset \bar{A}$.

Thus, we have

$$\bar{A} \subset \liminf A_n \subset \limsup A_n \subset \bar{A},$$

and therefore all these inclusions are equalities. This proves that the closed limit of $(A_n)_{n\geq 0}$ exists and is equal to $\bar{A}$. □

The following is a partial converse to Proposition 4.3.7.

Proposition 4.3.8. *Let X be a proper metric space, let p be a point in X and let $(A_n)_{n\geq 0}$ be a sequence of nonempty subsets of X. Then*

$$\lim A_n = A \Rightarrow d_p(A_n, A) \to 0.$$

Proof. Let us first recall that since A is a closed limit of a sequence of subsets, then A is closed.

To prove the proposition, we reason by contradiction. Assume that $d_p(A_n, A) \not\to 0$. Then we can find a positive real number ϵ and a subsequence $(A_{n_i})_{i\geq 0}$ of $(A_n)_{n\geq 0}$ such that

$$d_p(A_{n_i}, A) \geq 4\epsilon > 0.$$

Thus, for each $n_i \geq 0$, we can find a point x_{n_i} in X such that for all n_i large enough, we have

$$|d_{A_{n_i}}(x_{n_i}) - d_A(x_{n_i})|e^{-|p-x_{n_i}|} > 3\epsilon. \tag{4.3.8.1}$$

We claim that the sequence $(|p - x_{n_i}|)_{i\geq 0}$ is bounded.

To prove this claim, let z be in A. Then for each n_i, we can find a point z_{n_i} in A_{n_i} such that $(z_{n_i}) \to z$ as $i \to \infty$. We have

$$\begin{aligned} |d_{A_{n_i}}(x_{n_i}) - d_A(x_{n_i})| &\leq d_{A_{n_i}}(x_{n_i}) + d_A(x_{n_i}) \\ &\leq |x_{n_i} - z_{n_i}| + |x_{n_i} - z| \\ &\leq 2|p - x_{n_i}| + |p - z_{n_i}| + |z - p|. \end{aligned}$$

Thus, we obtain

$$(2|p - x_{n_i}| + |p - z_{n_i}| + |z - p|)e^{-|p - x_{n_i}|} \geq |d_{A_{n_i}}(x_{n_i}) - d_A(x_{n_i})|e^{-|p - x_{n_i}|},$$

which implies

$$(2|p - x_{n_i}| + |p - z_{n_i}| + |z - p|)e^{-|p - x_{n_i}|} > 3\epsilon. \tag{4.3.8.2}$$

If the sequence $(|p - x_{n_i}|)$ were unbounded, then, up to passing to a subsequence, we would have $|p - x_{n_i}| \to \infty$ as $i \to \infty$, which implies

$$(2|p - x_{n_i}| + |p - z_{n_i}| + |z - p|)e^{-|p - x_{n_i}|} \to 0,$$

which contradicts (4.3.8.2).

Now Inequality (4.3.8.1) implies $|d_{A_{n_i}}(x_{n_i}) - d_A(x_{n_i})| > 3\epsilon$. Thus, up to replacing the sequence (A_{n_i}) by a subsequence, we can assume that either

$$d_{A_{n_i}}(x_{n_i}) - d_A(x_{n_i}) > 3\epsilon \tag{4.3.8.3}$$

or

$$d_A(x_{n_i}) - d_{A_{n_i}}(x_{n_i}) > 3\epsilon. \tag{4.3.8.4}$$

We first deal with the case where (4.3.8.3) occurs. For each $i \geq 0$, let y_{n_i} be a point in A satisfying

$$|x_{n_i} - y_{n_i}| < d_A(x_{n_i}) + \epsilon. \tag{4.3.8.5}$$

Since the sequence $(|p - x_{n_i}|)$ is bounded, the sequence $(d_A(x_{n_i}))$ is bounded, therefore the sequence (y_{n_i}) is also bounded. Since X is proper, up to passing to a subsequence, we can assume that (y_{n_i}) has a limit, which we denote by y. Since A is closed, y belongs to A. We have

$$\begin{aligned} d_{A_{n_i}}(y_{n_i}) + \epsilon &\geq d_{A_{n_i}}(x_{n_i}) - |x_{n_i} - y_{n_i}| + \epsilon \quad \text{(since } d_{A_{n_i}} \text{ is non-expanding)} \\ &> d_{A_{n_i}}(x_{n_i}) - d_A(x_{n_i}) \quad \text{(by (4.3.8.5))} \\ &> 3\epsilon \text{ (by (4.3.8.3))}. \end{aligned}$$

Now since $y_{n_i} \to y$ as $i \to \infty$, we have, for all n_i large enough,

$$d_{A_{n_i}}(y) > d_{A_{n_i}}(x_{n_i}) - \epsilon > 3\epsilon - 2\epsilon = \epsilon.$$

This contradicts the fact that y is in $\liminf A_n$. Thus, Inequality (4.3.8.3) cannot occur. Suppose now that (4.3.8.4) is satisfied. For each i, we choose a point u_{n_i} in A_{n_i} such that

$$|u_{n_i} - x_{n_i}| < d_{A_{n_i}}(x_{n_i}) + \epsilon. \tag{4.3.8.6}$$

The sequence (u_{n_i}) is bounded, since (x_{n_i}) is bounded and we have

$$\begin{aligned} d_A(u_{n_i}) + \epsilon &\geq d_A(x_{n_i}) - |x_{n_i} - u_{n_i}| + \epsilon \quad \text{(since } d_A \text{ is non-expanding)} \\ &> d_A(x_{n_i}) - d_{A_{n_i}}(x_{n_i}) \quad \text{(by (4.3.8.6))} \\ &> 3\epsilon \text{ (by (4.3.8.4))}. \end{aligned}$$

Thus, we obtain

$$d_A(u_{n_i}) > 2\epsilon. \tag{4.3.8.7}$$

The space X being proper, since the sequence (u_{n_i}) is bounded, it has an accumulation point u, which is in A since $A = \limsup A_n$. This contradicts (4.3.8.7). Thus, we have $d_p(A_n, A) \to 0$ as $n \to \infty$. This completes the proof of Proposition 4.3.8.□

4.4 Metrics on the isometry group

We first recall the following

Definition 4.4.1 (Isometry and isometric spaces). Let X and Y be two metric spaces. A map $f: X \to Y$ is called an *isometry* if f is distance-preserving and onto. The spaces X and Y are said to be *isometric* if there exists an isometry $f: X \to Y$.

It follows easily from this definition that an isometry is continuous and invertible, that its inverse is also an isometry and that the relation of being isometric is an equivalence relation between metric spaces.

In this section, we are interested in the case where $Y = X$. In this case, the composition of two isometries is always well-defined and the set of isometries of the metric space X forms a group that we denote by $\mathrm{Isom}(X)$.

We recall that a *topological group* is a group G equipped with a topology satisfying the following two properties:

- the map $G \to G$ defined by $g \mapsto g^{-1}$ is continuous;
- the map $G \times G \to G$ defined by $(f, g) \mapsto fg$ is continuous. (The space $G \times G$ is equipped with the natural product topology.)

Proposition 4.4.2. *Let X be a compact metric space. Then the map*

$$d: \mathrm{Isom}(X) \times \mathrm{Isom}(X) \to \mathbb{R}$$

defined by

$$d(f, g) = |f - g| = \sup_{x \in X} |f(x) - g(x)|$$

is a metric on $\mathrm{Isom}(X)$. *This metric is invariant under the left and the right actions of the group* $\mathrm{Isom}(X)$ *on itself, and this group, equipped with this metric, is a compact topological group.*

Proof. First we note that since X is compact, then for every f and g in $\mathrm{Isom}(X)$, the value of $|f(x) - g(x)|$ is uniformly bounded (by the diameter of X). Thus $d(f, g)$ is finite. It is easy to check that d is a metric. To prove left-invariance, we let f, g and h be arbitrary elements of $\mathrm{Isom}(X)$. Since h is an isometry, we have, for each x in X,

$$|h \circ f(x) - h \circ g(x)| = |f(x) - g(x)|.$$

Taking the supremum over all x in X, we obtain the equality $|h \circ f - h \circ g| = |f - g|$, that is, the left-invariance of the action. The right invariance can be proved in the same way.

Now let us prove that $\mathrm{Isom}(X)$, equipped with this metric, is compact. Let $(f_n)_{n\geq 0}$ be a sequence of elements in $\mathrm{Isom}(X)$. Since X is compact, it is separable and by Ascoli's Theorem, (f_n) has a convergent subsequence $(f_{n_i})_{i\geq 0}$. Let f be the limit of that subsequence. We must show that f is an isometry. For each x and y in X and for each integer $i \geq 0$, we have $|f_{n_i}(x) - g_{n_i}(y)| = |x - y|$. Letting i tend to infinity, we obtain $|f(x) - g(x)| = |x - y|$. Thus, f is distance-preserving. To prove that f is surjective, we can use the fact that any distance-preserving map from a compact metric space to itself is surjective (Corollary 3.3.5), or we can use the sequence of inverses $(f_{n_i}^{-1})_{i\geq 0}$ to produce an inverse for f. Thus, f is an isometry. This shows that $\mathrm{Isom}(X)$ is compact.

Finally, let us show that $\mathrm{Isom}(X)$, equipped with this metric, is a topological group. For any f and g in $\mathrm{Isom}(X)$, we have

$$\begin{aligned}
|f - g| &= \sup_{x\in X} |f(x) - g(x)| \\
&= \sup_{x\in X} |g^{-1} \circ f(x) - x| \text{ (since } g^{-1} \text{ is an isometry)} \\
&= \sup_{x\in X} \left|g^{-1}\big(f(x)\big) - f^{-1}\big(f(x)\big)\right| \\
&= \sup_{y\in X} |g^{-1}(y) - f^{-1}(y)| \text{ (by setting } y = f(x)) \\
&= |f^{-1} - g^{-1}|.
\end{aligned}$$

This proves that the map $f \mapsto f^{-1}$ is distance-preserving and therefore continuous. Let us prove now that the map $(f, g) \mapsto fg$ is continuous. We recall that the topology on $G \times G$ is induced by the metric $\big((f_1, f_2), (g_1, g_2)\big) \mapsto |f_1 - f_2| + |g_1 - g_2|$. We have

$$\begin{aligned}
|f_2 \circ g_2 - f_1 \circ g_1| &= \sup_{x\in X} |f_2 \circ g_2(x) - f_1 \circ g_1(x)| \\
&\leq \sup_{x\in X} |f_2(g_2(x)) - f_1(g_2(x))| + |f_1(g_2(x)) - f_1(g_1(x))| \\
&= \sup_{x\in X} |f_2(g_2(x)) - f_1(g_2(x))| + |g_2(x) - g_1(x)| =
\end{aligned}$$

$$= \sup_{y \in X} |f_2(y) - f_1(y)| + |g_2(x) - g_1(x)|$$
$$= |f_2 - f_1| + |g_2 - g_1|.$$

This proves the continuity of the map $(f, g) \mapsto fg$, which completes the proof of Proposition 4.4.2. □

In the case where X is not compact, the map d of Proposition 4.4.2 does not define, in general, a distance on $\mathrm{Isom}(X)$. For instance, with such a definition, if we take X to be the Euclidean plane, then the distance between the identity and any isometry that is not a translation would be infinite. Busemann defines in [28] a distance function on the isometry group of an arbitrary metric space by modifying the formula for d. In fact, he defines a family of metrics on $\mathrm{Isom}(X)$, in which the parameter space is the space X itself, and we now recall his definition.

Let us choose a point p in X. We define a map d_p on $\mathrm{Isom}(X) \times \mathrm{Isom}(X)$ by setting, for every f and g in $\mathrm{Isom}(X)$,

$$d_p(f, g) = |f - g|_p = \sup_{x \in X} |f(x) - g(x)| e^{-|p-x|}.$$

Proposition 4.4.3. *For every metric space X and for every p in X, the map d_p defines a metric on* $\mathrm{Isom}(X)$.

Proof. We first show that for every f and g in $\mathrm{Isom}(X)$, we have $|f - g|_p < \infty$. For x in X, we have

$$\begin{aligned} |f(x) - g(x)| &\le |f(x) - p| + |p - g(x)| \\ &= |x - f^{-1}(p)| + |g^{-1}(p) - x| \\ &\le |x - p| + |p - f^{-1}(p)| + |x - p| + |p - g^{-1}(p)| \\ &= 2|x - p| + |p - f^{-1}(p)| + |p - g^{-1}(p)|. \end{aligned}$$

Therefore, we have

$$\begin{aligned} |f(x) - g(x)| e^{-|p-x|} &\le (2|x - p| + |p - f^{-1}(p)| + |p - g^{-1}(p)|) e^{-|p-x|} \\ &= 2|x - p| e^{-|p-x|} + (|p - f^{-1}(p)| + |p - g^{-1}(p)|) e^{-|p-x|}. \end{aligned}$$

Using the fact that $e^{-t} \le 1$ and $te^{-t} < 1$ for all $t \ge 0$, we obtain

$$|f(x) - g(x)| e^{-|p-x|} \le 2 + |p - f^{-1}(p)| + |p - g^{-1}(p)|.$$

Thus, $|f(x) - g(x)| e^{-|p-x|}$ is bounded by a constant that is independent of x, which implies that $|f - g|_p$ is finite.

Now we must prove that d_p satisfies the properties of a distance function. All the properties except the triangle inequality are trivially satisfied. Let us prove the triangle inequality. For all $\epsilon > 0$, for every x in X and for every f, g and h in $\mathrm{Isom}(X)$, we have

$$\begin{aligned} |f-h|_p - \epsilon &\leq |f(x)-h(x)|e^{-|p-x|} \\ &\leq (|f(x)-g(x)| + |g(x)-h(x)|)e^{-|p-x|} \\ &\leq |f-g|_p + |g-h|_p. \end{aligned}$$

By making $\epsilon \to 0$, we obtain $|f-g|_p \leq |f-g|_p + |g-h|_p$. This completes the proof of Proposition 4.4.3. □

Proposition 4.4.4. *Let X be a metric space. Then,*

(i) *the metric d_p on* $\mathrm{Isom}(X)$ *is invariant by the left-action of* $\mathrm{Isom}(X)$ *on itself;*
(ii) *for every f, g and h in* $\mathrm{Isom}(X)$*, we have $|f \circ h - g \circ h|_p = |f-g|_{h(p)}$.*

Proof. Left-invariance can be proved in the same way as left-invariance of the metric d in Proposition 4.4.2. We prove (ii). We have

$$\begin{aligned} |f \circ h - g \circ h|_p &= \sup_{x \in X} |f \circ h(x) - g \circ h(x)|e^{-|p-x|} \\ &= \sup_{y \in Y} |f(y) - g(y)|e^{-|p-h^{-1}(y)|} \\ &= \sup_{y \in Y} |f(y) - g(y)|e^{-|h(p)-y|} \\ &= |f-g|_{h(p)}. \end{aligned}$$

□

Proposition 4.4.5. *Let X be a metric space. Then,*

(i) *for any p and q in X, the metrics d_p and d_q on* $\mathrm{Isom}(X)$ *are commensurable;*
(ii) $\mathrm{Isom}(X)$*, equipped with any metric d_p is a topological group.*

Proof. For every p, q and x in X, we have $|q-x| \geq |p-x| - |p-q|$. Therefore, $e^{-|q-x|} \leq e^{-|p-x|}e^{|p-q|}$. Thus, we obtain

$$\begin{aligned} |f-g|_q &= \sup_{x \in X} |f(x)-g(x)|e^{-|q-x|} \\ &\leq \sup_{x \in X} |f(x)-g(x)|e^{-|p-x|}e^{-|p-q|} \\ &= |f-g|_p e^{-|p-q|}. \end{aligned}$$

In the same way (or by symmetry), we have $|f-g|_p \leq |f-g|_q e^{-|p-q|}$. This implies that the two metrics are commensurable. This proves (i).

The fact that $\mathrm{Isom}(X)$ is a topological group can be proved in the same way as in the proof Proposition 4.4.2, using left-invariance of the metric d_p and using Property (ii) of Proposition 4.4.4 instead of right-invariance. □

The next proposition compares pointwise convergence of isometries with convergence with respect to the metric d_p. It shows in particular that the two notions coincide in the case where the space X is proper.

Proposition 4.4.6. *Let X be a metric space and let p be in X. Then,*

(i) *for any sequence $(f_n)_{n\geq 0}$ in* $\mathrm{Isom}(X)$*, for any element f in* $\mathrm{Isom}(X)$ *such that $|f_n - f|_p \to 0$ and for every x in X, we have $|f_n(x) - f(x)| \to 0$;*

(ii) *if X is proper, if $(f_n)_{n\geq 0}$ is a sequence in* $\mathrm{Isom}(X)$ *and if f is an element in* $\mathrm{Isom}(X)$ *satisfying $|f_n(x) - f(x)| \to 0$ for every x in X, we have $|f_n - f|_p \to 0$.*

Proof. Property (i) follows from the fact that for every x in X, we have $|f_n(x) - f(x)|_X \leq |f_n - f|_p\, e^{|p-x|}$.

We prove (ii) by contradiction. Suppose that $|f_n - f|_p$ does not converge to 0. For every $n \geq 0$, we set $\alpha_n = \sup_{x\in X} |f_n(x) - f(x)|e^{-|p-x|}$. We can find a subsequence $(f_{n_i})_{i\geq 0}$ of $(f_n)_{n\geq 0}$ and a real number $\epsilon > 0$ satisfying $\alpha_{n_i} \geq \epsilon$ for every $i \geq 0$. Thus, taking $\alpha = \epsilon/2$, we can find, for every integer $i \geq 0$, a point x_{n_i} in X such that $|f_{n_i}(x_{n_i}) - f(x_{n_i})|e^{-|p-x_{n_i}|} \geq \alpha > 0$.

Now since X is proper, we can suppose, up to replacing the sequence (f_{n_i}) by a subsequence, that either $|p - x_{n_i}| \to \infty$ as $i \to \infty$, or that the sequence (x_{n_i}) converges to a point x in X. We show that neither of these two cases can occur; this will imply that $|f_n - f|_p \to 0$ as $n \to \infty$.

If the first case occurs, then we have

$$\begin{aligned}
\alpha &\leq |f_{n_i}(x_{n_i}) - f(x_{n_i})|e^{|p-x_{n_i}|}\\
&\leq (|f_{n_i}(x_{n_i}) - f_{n_i}(p)| + |f_{n_i}(p) - f(p)| + |f(p) - f(x_{n_i})|)e^{-|p-x_{n_i}|}\\
&= (|x_{n_i}p| + |f_{n_i}(p) - f(p)| + |p - x_{n_i}|)e^{-|p-x_{n_i}|}\\
&\leq 2|p - x_{n_i}|e^{-|p-x_{n_i}|} + |f_{n_i}(p) - f(p)|e^{-|p-x_{n_i}|}.
\end{aligned}$$

Since $|p - x_{n_i}| \to \infty$ as $i \to \infty$, the expression in the last line tends to 0 as $i \to \infty$, and this contradicts the fact that it is bounded below by α.

In the second case, we have

$$\begin{aligned}
\alpha &\leq |f_{n_i}(x_{n_i}) - f(x_{n_i})|e^{-|p-x_{n_i}|}\\
&\leq |f_{n_i}(x_{n_i}) - f_{n_i}(x)| + |f_{n_i}(x) - f(x)| + |f(x) - f(x_{n_i})|\\
&= 2|x - x_{n_i}| + |f_{n_i}(x) - f(x)|.
\end{aligned}$$

Since $|x - x_{n_i}| \to 0$ as $i \to \infty$, the expression in the last line tends to 0, and this gives again a contradiction. This completes the proof of Proposition 4.4.6. □

The following result is also due to Busemann (cf. [28] p. 17).

Theorem 4.4.7. *Suppose that X is a proper metric space. Then for every p in X, the metric space* $(\mathrm{Isom}(X), d_p)$ *is proper.*

Proof. To prove that $(\mathrm{Isom}(X), d_p)$ is proper, consider an arbitrary bounded sequence $(f_n)_{n\geq 0}$ in $\mathrm{Isom}(X)$, and let us show that we can extract from it a convergent subsequence. Since (f_n) is bounded, there exists a real number $\alpha > 0$ satisfying $|f_n - f_m|_p < \alpha$ for each m and $n \geq 0$. This implies that for each x in X, we have $|f_m(x) - f_n(x)| < \alpha e^{-|p-x|}$. Thus, for each x in X, the sequence $\big(f_n(x)\big)_{n\geq 0}$ is bounded. Furthermore, since every map f_n is distance-preserving, the sequence of maps $(f_n)_{n\geq 0}$ is equicontinuous. By Ascoli's theorem (Theorem 1.4.9), there exists a subsequence $(f_{n_i})_{i\geq 0}$ of $(f_n)_{n\geq 0}$ and a map $f : X \to X$ such that $|f_{n_i}(x) - f(x)| \to 0$ for each x in X. By taking limits, we obtain $|f(x) - f(y)| = |x - y|$ for every x and y in X. Thus, f is distance-preserving.

Now let us prove that f is surjective. Let y be an arbitrary element of X. For each $i \geq 0$, we have $|f_{n_i}^{-1}(y) - p| = |y - f_{n_i}(p)|$, which converges to $|y - f(p)|$ as $i \to \infty$. Thus, the sequence $(f_{n_i}^{-1}(y))_{i\geq 0}$ is bounded. Therefore, up to taking a subsequence, we can assume that $f_{n_i}^{-1}(y)$ converges to a point x in X. We then have

$$|y - f(x)| = \lim_{i\to\infty} |y - f_{n_i}(x)| = \lim_{i\to\infty} |f_{n_i}^{-1}(y) - x| = 0.$$

Thus, we have $y = f(x)$, which shows that f is surjective.

Since $f_{n_i}(x)$ converges to $f(x)$ for every x in X, we have, by Proposition 4.4.6, $|f_{n_i} - f|_p \to 0$. Thus, the sequence (f_{n_i}) converges to f, and this completes the proof of Theorem 4.4.7. □

Notes on Chapter 4

Distances between subsets. In [66], Hausdorff defined (what is now called) the Hausdorff distance $d_{\mathcal{H}}(A, B)$, which he denoted by $\overline{AB}$, between subsets A and B of a metric space, and he established its main properties. The modified version $d_p(A, B)$ which we study in Section 1 was defined by Busemann in [28], §3. In [66], Hausdorff considered several other "distances" between subsets of a metric space. For instance, if A and B are subsets of a metric space X, he defined their "lower distance" by

$$\delta(A, B) = \inf_{x\in A, y\in B} |x - y|$$

and their "upper distance" by

$$d(A, B) = \sup_{x\in A, y\in B} |x - y|.$$

In particular, $d(A, A)$ is the diameter of A and for x in X, $\delta(\{x\}, A)$ is equal to the distance from x to A, a quantity that we denoted by $d_A(x)$.

The value of $d(A, B)$ is finite if and only if both A and B are bounded.

The "distance functions" δ and d do not necessarily separate points in X (that is, the distance between two sets can be zero without the sets being equal). For instance, take A to be a non-closed set and $B = \bar{A}$.

Hausdorff also defined the quantity $\rho(A, B)$ by

$$\rho(A, B) = \inf\{\epsilon \geq 0 \text{ such that } B \subset N(A, \epsilon)\}.$$

It is easy to see that in general this function ρ is not symmetric. As Hausdorff notes, one always has the inequalities

$$\delta(A, B) \leq \rho(A, B) \leq d(A, B).$$

The Hausdorff distance $d_{\mathcal{H}}(A, b)$ is a symmetrization of $\rho(A, B)$:

$$d_{\mathcal{H}}(A, B) = \max\big(\rho(A, B), \rho(B, A)\big).$$

Floyd's distance and Floyd's boundary of a finitely generated group. The idea of changing a distance function by scaling it by a factor that depends on the distance to a basepoint, as in the definitions of the Busemann–Hausdorff metric on subsets of a metric space X or in that of the metric d_p on the isometry group of X, is recurrent in geometry. We mention as another beautiful example the distance defined in the paper [48] by W. Floyd on the Cayley graph of a finitely generated group. This distance is used to define what is now called "Floyd's boundary" of the group. We recall the definition. Let Γ be a finitely generated group equipped with a finite generating set S and let $C(\Gamma, S)$ be the associated Cayley graph equipped with the length metric d in which the length of each edge is equal to 1 (see Example 2.1.3 (v)). Floyd's metric on $C(\Gamma, S)$ is then obtained by scaling the metric d in the following way: we choose a basepoint, say a vertex v, in $C(\Gamma, S)$ and we multiply the length of each edge e by a factor $f(M)$ where $M = M(e, v)$ is the least number of edges separating v from e and where $f\colon \mathbb{N} \to \mathbb{R}_+^*$ is a function that satisfies the following two conditions:

- $\sum_{M=0}^{\infty} f(M) < \infty$;
- for any k in $\mathbb{N}$, there exists two positive constants c_1 and c_2 such that $c_1 f(M) \leq f(kM) \leq c_2 f(M)$ for any M in $\mathbb{N}$.

For instance, one can take $f(M) = M^{-2}$. Then, the Floyd metric d' is the length metric on $C(\Gamma, S)$ associated to the new set of lengths of edges. This metric is not complete (except if the group is finite) and its completion $\overline{C(\Gamma, S)}$ is called a *Floyd completion* of the group Γ. The set $\overline{C(\Gamma, S)} \setminus C(\Gamma, S)$ is *Floyd's boundary* of Γ. This construction is reminiscent of the definition of the boundary S^{n-1} of n-dimensional hyperbolic space, as it appears in Example 2.4.3. Indeed, the Euclidean metric of the unit ball B^n can be obtained by scaling the hyperbolic metric (we recall that at each point of B^n, the hyperbolic length element is a multiple of the Euclidean length element by a factor that depends only on the distance from the point to the origin of B^n, and this factor satisfies the properties required for the above function f). The

Euclidean metric of the ball is not complete, and its completion is the closed ball $\overline{B^n}$. The boundary S^{n-1} of hyperbolic space is then the set $\overline{B^n} \setminus B^n$. For the definitions and for applications of the Floyd boundary, we refer the reader to Floyd's paper [48].

Chapter 5

Convexity in vector spaces

Introduction

Our purpose in this chapter is to describe several facets of the classical theory of affine convex subsets of vector spaces, with a view on applications of convexity in general metric spaces.

We recall that all the vector spaces that we consider in these notes are real vector spaces.

An (affinely) convex set in a vector space E is a subset X of E such that for any pair of points in X, the affine segment joining them is contained in X.

The outline of this chapter is as follows.

In Section 1, we give definitions, examples and some basic properties of affinely convex sets. Convexity is stable under taking closure, interior, intersection, increasing union, sums and products.

In Section 2, we study the notions of convex hull, closed convex hull and convex kernel.

In Section 3, we consider normed vector spaces. A norm on a vector space provides some natural classes of convex subsets of this space. Closed balls and open balls are examples of such subsets, but of course there are many others. Normed vector spaces are important examples of geodesic metric spaces, but, as we shall see, these spaces are not necessarily uniquely geodesic. We shall also see that any metric space can be embedded by a distance-preserving map in a complete geodesic metric space.

In Section 4, we present a construction due to Minkowski that associates to each convex subset B in a finite-dimensional vector space E, such that B is closed, bounded, symmetric about the origin and contains the origin in its interior, a norm on E whose closed unit ball is B.

In Section 5, we describe the Hilbert metric associated to a nonempty bounded open convex subset of $\mathbb{R}^n$. Besides the intrinsic importance of this metric, it will provide us with beautiful examples of geodesic metric spaces. A special case of the Hilbert metric is a model for hyperbolic space $\mathbb{H}^n$, namely, the Klein model.

5.1 Affinely convex subsets

Definition 5.1.1 (Affine segment). Let E be a vector space and let x and y be two points in E. The subset $\{(1-t)x + ty \ : t \in [0,1]\}$ of E is called the *affine segment joining x and y* and is denoted by $[x, y]$.

We note right away that if the space E is equipped with a metric induced from a norm (and this will be the case for any metric that we shall consider on E in this book), then any affine segment in E is a geodesic segment, so that the notation $[x, y]$ that we use here is consistent with the one we introduced before, where $[x, y]$ denoted a geodesic segment joining x and y. However, there is a large class of normed vector spaces in which there exist geodesic segments that are different from the affine segment joining them.

Definition 5.1.2 (Affinely convex subset). Let E be a vector space. A subset $X \subset E$ is said to be *affinely convex* if for all x and y in X, the affine segment $[x, y]$ is contained in X (Figure 5.1).

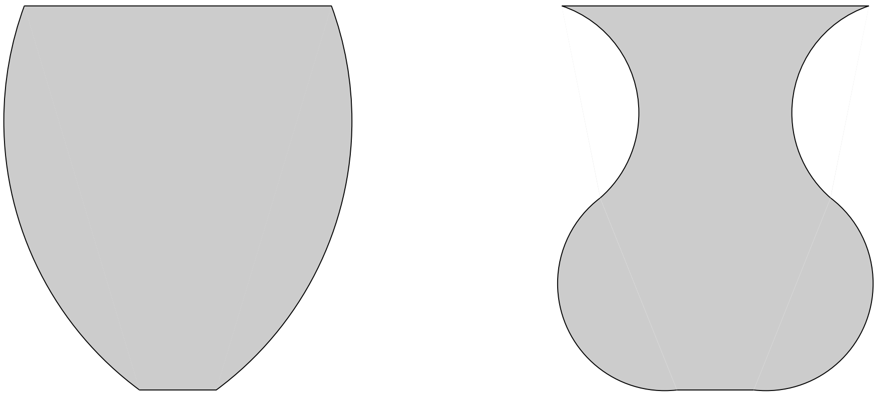


Figure 5.1. The left vase is convex, and the one on the right is not convex.

We shall also say a *convex subset* instead of an *affinely convex subset*, provided there is no ambiguity. (We recall that we already encountered other notions of convexity: geodesic convexity in uniquely geodesic metric spaces and Menger convexity in arbitrary metric spaces, and we shall see other notions of convexity in the sequel.)

We start with the following two convexity criteria:

Proposition 5.1.3. *Let E be a vector space and let X be a subset of E. Then X is convex if and only if for every x in X and for every t in* $[0, 1]$*, the homothety of center x and factor t sends the set X into itself.*

Proof. For every x and y in E and for every t in $[0, 1]$, we can write $(1 - t)x + ty = x + t(y - x)$, which shows that the point $(1-t)x + y$ is the image of y by the homothety of center x and factor t. From this, the result follows easily. □

Proposition 5.1.4. *Let E be a vector space and let X be a subset of E. The following two properties are equivalent:*

(i) *X is convex;*

(ii) *for every integer $n \geq 0$, for every $t_0, \ldots, t_n$ in $[0, 1]$ satisfying $t_0 + \cdots + t_n = 1$ and for every $x_0, \ldots, x_n$ in X, we have $t_0x_0 + \cdots + t_nx_n \in X$.*

Proof. Property (ii) with $n = 1$ implies (i). Let us prove that (i) $\Rightarrow$ (ii). We use induction. Property (ii) is trivially satisfied for $n = 0$. For $n = 1$, this property follows from (i). Thus, let us suppose that (ii) is satisfied for some integer $n \geq 1$ and let us show that for every $t_0, \ldots, t_{n+1}$ in $[0, 1]$ satisfying $t_0 + \cdots + t_{n+1} = 1$ and for every $x_0, \ldots, x_{n+1}$ in X, we have $t_0x_0 + \cdots + t_{n+1}x_{n+1} \in X$. We can assume that $t_0 \neq 1$, for otherwise every t_i for $i \geq 1$ would be 0 and the desired result would be trivially satisfied. Let $t = t_1 + \cdots + t_{n+1}$. Then $t \neq 0$. For every $i = 1, \ldots, n+1$, let us set $t'_i = t_i/t$. Then $t'_1 + \cdots + t'_{n+1} = 1$. The induction hypothesis implies that $t'_1x_1 + \cdots + t'_{n+1}x_{n+1} \in X$. Since X is convex, we also have $(1-t)x_0 + t(t'_1x_1 + \cdots + t'_{n+1}x_{n+1}) \in X$, that is, $t_0x_0 + \cdots + t_{n+1}x_{n+1} \in X$, which completes the proof of Proposition 5.1.4. □

The following is a useful notion that generalizes the notion of a point on the affine segment joining two points.

Definition 5.1.5 (Affine convex combination). Let E be a vector space and let $\{x_0, \ldots, x_n\}$ be a finite set of points in E. An *affine convex combination* of $\{x_0, \ldots, x_n\}$ is a point of the form $t_0x_0 + \cdots + t_nx_n$ where for each $i \geq 0$, t_i is a nonnegative real number and $t_0 + \cdots + t_n = 1$.

Thus, Proposition 5.1.4 says that a subset X of a vector space E is convex if and only if for every finite subset F of E, any affine convex combination of F is in X.

Now, let us give a few examples of convex sets:

Examples 5.1.6 (Convex sets).

(i) *Affine subspaces.* For any vector space E, any vector subspace of E and, more generally, any affine subspace of E, is a convex subset of E. (We recall that a subset A of E is said to be an affine subspace if there exists a vector subspace V of E and an element a of E such that $A = V + a$.)

(ii) *Convex subsets of* $\mathbb{R}$. If $E = \mathbb{R}$, then the convex subsets of E are the intervals.

(iii) *Function spaces.* Let $E = \mathcal{C}([a, b])$ be the vector space of real-valued functions defined on an interval $[a, b]$. It is easy to see that the subsets

$$\{f \in E \text{ such that } |f(x)| \leq 1 \quad \text{for all } x \in E\}$$

and

$$\{f \in E \text{ such that } |f(x)| < 1 \quad \text{for all } x \in E\}$$

of E are convex. This is a special case of a general result (Proposition 5.3.13 below) which says that in a normed vector space, closed balls and open balls are convex.

(iv) *Images and inverse images by affine maps.* Let E and F be two vector spaces and let $f: E \to F$ be a linear map or, more generally, an affine map (that is to say, a linear map up to an additive constant). Then the image by f of any convex subset of E is a convex subset of F and the inverse image by f of any convex subset of F is a convex subset of E. A non-zero linear form or, more generally, a non-constant affine form $f: E \to \mathbb{R}$ defines a *hyperplane* in E; this is the set of points x in E satisfying $f(x) = 0$. Hyperplanes are convex. The linear form f defines two other special closed (respectively open) convex subsets of E that are bordered by the hyperplane $f^{-1}(0)$; these are the sets $f^{-1}(]-\infty, 0])$ and $f^{-1}([0, \infty[)$ (respectively $f^{-1}(]-\infty, 0[)$ and $f^{-1}(]0, \infty[)$). These sets are called the *closed* (*respectively open*) *half-spaces* bordered by the hyperplane $f^{-1}(0)$. All these sets are convex, as inverse images of intervals of $\mathbb{R}$ by affine maps.

Related to the last example, we mention the following notion, which is of central importance in convexity theory:

Definition 5.1.7 (Convex polyhedron). A convex polyhedron in a vector space is the intersection of a finite number of closed half-spaces.

A convex polyhedron is an affinely convex subset, since it is the intersection of affinely convex subsets.

It is clear from Definition 5.1.7 that a finite intersection of convex polyhedra is a convex polyhedron.

Example 5.1.8 (Convex polyhedra). Closed half-spaces, cubes and simplices are examples of convex polyhedra in $\mathbb{R}^n$. The *standard cube* of dimension n of $\mathbb{R}^n$, denoted by Cube_n, is the convex polyhedron defined as

$$\mathrm{Cube}_n = \{(x_1, \dots x_n)\ ,\ |x_i| \leq 1 \text{ for all } i = 1, \dots, n\}.$$

The *standard co-cube* of dimension n in $\mathbb{R}^n$, denoted by Co-cube_n, is the convex polyhedron defined as

$$\text{Co-cube}_n = \Big\{(x_1, \dots x_n),\ \sum_{i=1}^{n} |x_i| \leq 1\Big\}.$$

The standard cube and co-cube are the closed unit balls of the ℓ^∞ and the ℓ^1 norms respectively on $\mathbb{R}^n$, and we shall deal with them below. Let us note right away that the standard cube and co-cube of dimensions 1 or 2 are isometric, but that in dimension $n \geq 3$, this is not the case. Indeed, for all $n \geq 1$, the faces of a standard cube of dimension $n + 1$ are cubes, whereas those of a standard co-cube of dimension $n + 1$ are (regular) n-simplices and n-simplices and cubes of dimension n are combinatorially different if $n \geq 2$.

Let us review a few operations that preserve convexity of sets.

Intersection. It follows directly from Definition 5.1.2 that the intersection of a family of convex subsets of E is convex.

Increasing union. Likewise, an increasing union of convex subsets of E is convex.

Homothety. If X is a convex subset of E, then for every real number λ, the set λX defined as

$$\lambda X = \{\lambda x \text{ such that } x \in X\}$$

is convex.

Vector sum of subsets. If X and Y are two subsets of E, we recall that $X + Y$ is the subset of E defined as

$$X + Y = \{x + y \text{ such that } x \in X \text{ and } y \in Y\}.$$

If X and Y are convex, then $X + Y$ is also convex. To see this, let us take two arbitrary elements $x + y$ and $x' + y'$ in $X + Y$, with x and x' in X, and y and y' in Y. If X and Y are convex, then for all t in [0, 1] we have $(1 - t)x + ty \in X$ and $(1 - t)x' + ty' \in Y$, which implies that $(1 - t)(x + x') + t(y + y')$ is in $X + Y$. This shows that $X + Y$ is convex.

The notion of vector sum of subsets is due to Minkowski, and the vector sum of two convex bodies is sometimes called their *Minkowski sum*. Equipped with this operation, the set of nonempty convex subsets of a vector space is an abelian semigroup, whose identity element is the origin of E. This semigroup, equipped with the extra operation of homothety, has the structure of a convex cone.[1]

Product. Let E and F be two vector spaces and let X and Y be convex subsets of E and F respectively. Then the product $X \times Y$ is a convex subset of the product space $E \times F$.

Now, we study some properties of convexity that involve topology.

Any finite-dimensional vector space has a natural topology, which is the topology associated to an arbitrary norm. The fact that these norms induce the same topology follows from the fact that all norms on a finite-dimensional vector space are commensurable. Thus, the topological results that we collect below are valid in any finite-dimensional vector space, which we shall always assume to be equipped with its natural topology, without making reference to any norm.

We also note that even though we restrict ourselves here to finite-dimensional vector spaces, these results are in general valid in any topological vector space E.

---

[1] In relation to this abelian semigroup structure, we mention the following nice convexity criterion due to J. M. Borwein and R. C. O'Brien (see [19]) which uses cancellation in this semi-group: a compact subset X of a finite-dimensional vector space E is convex if and only if for every compact convex subsets Y and Z of E, we have

$$X + Y = X + Z \Rightarrow Y = Z.$$

We shall denote the closure and the interior of a subset X of E by $\overline{X}$ and $\overset{\circ}{X}$ respectively. We start with the following:

Proposition 5.1.9 (The closure of a convex set is convex). *Let E be a topological vector space and let X be an affinely convex subset of E. Then the closure of X is also affinely convex.*

Proof. By Proposition 5.1.3, X is convex if and only if for every x in X and for every t in $[0, 1]$, the homothety of factor t and center x sends X into itself. In a topological vector space, a homothety is continuous, and if a continuous map sends a subset X into itself, then it sends its closure $\overline{X}$ into itself. Thus, any homothety centered at a point in X and of factor $t \in [0, 1]$ sends the space $\overline{X}$ into itself. By continuity, any homothety of E centered at a point in $\overline{X}$ and of factor $t \in [0, 1]$ sends $\overline{X}$ into itself. This shows that $\overline{X}$ is convex. This proves Proposition 5.1.9 □

There is a similar result concerning the interior of a convex subset. Before stating it, we prove a lemma that will be useful in the proof of that result as well as in other circumstances.

We first introduce a notation. If x_0 and x_1 are two points in a vector space E, then the *open affine segment with endpoints x_0 and x_1* is the set of points in E that are of the form $(1 - t)x_0 + tx_1$ with $0 < t < 1$. We denote this set by $]x_0, x_1[$.

Lemma 5.1.10. *Let E be a topological vector space and let X be an affinely convex subset of E. For every x_0 in X and for every x_1 in $\overset{\circ}{X}$, the open affine segment $]x_0, x_1[$ is contained in $\overset{\circ}{X}$.*

Proof. Let $B = B(x_1, r)$ be the open ball in E of radius $r > 0$ and center x_1, where r is small enough so that this ball is contained in X. Since X is convex, then, by Proposition 5.1.3, for all t in $]0, 1]$, the homothety centered at x_0 and of factor t sends B homeomorphically onto an open ball B_t, that is contained in X. The open ball B_t has positive radius rt and it is centered at $x_t = (1 - t)x_0 + tx_1$. Therefore, the point x_t is contained in $\overset{\circ}{X}$. This proves the lemma. □

We deduce the following

Proposition 5.1.11 (The interior of a convex set is convex). *Let E be a topological vector space and let X be an affinely convex subset of E. Then the interior of X is also affinely convex.*

Proof. By Lemma 5.1.10, if x_0 and x_1 are in $\overset{\circ}{X}$, then the whole segment $[x_0, x_1]$ is in $\overset{\circ}{X}$. □

5.2 Convex hull

Definition 5.2.1 (Affine convex hull). Let E be a vector space and let X be a subset of E. The (*affine*) *convex hull* of X, denoted by $C(X)$, is the intersection of all the affinely convex subsets of E that contain X.

The convex hull $C(X)$, being an intersection of convex subsets of E, is convex. It is the smallest (with respect to inclusion) convex subset of E containing X. The following properties are easy to check:

If X and Y are two subsets of X, then $C(X + Y) = C(X) + C(Y)$. If E' is another vector space and if X and Y are respectively subsets of E and E', then $C(X \times Y) = C(X) \times C(Y)$. If $f : E \to E'$ is a linear (or, more generally, an affine) map, then $C(f(X)) = f(C(X))$.

A *polytope* in a vector space E is, by definition, the affine convex hull of a finite subset of E. Examples of polytopes are convex polygons in a plane. Other examples are the n-simplices, which are affine convex hulls of affinely independent $n+1$ points in E (that is, $n + 1$ points that span n-dimensional affine subspaces.)

Proposition 5.2.2. *Let E be a vector space and let X be a subset of E. The convex hull $C(X)$ of X is the set of affine convex combinations of finite subsets of X. In other words, $C(X)$ is equal to the subset $L(X)$ of E defined by*

$$L(X) = \Big\{ \sum_{i=0}^{n} t_i x_i \text{ with } n \geq 0,\, x_i \in X \text{ such that } t_i \geq 0 \text{ for all } i = 1, \ldots, n \text{ and } t_0 + \cdots + t_n = 1 \Big\}.$$

(We shall prove in the next proposition that if E has finite dimension d, then we can take $n = d$.)

Proof. Let us first prove that $L(X)$ is contained in $C(X)$. It suffices to prove that if X' is an arbitrary affinely convex subset of E that contains X, then X' contains $L(X)$. Let x be an element in $L(X)$. Then there exists a sequence $x_0, \ldots, x_n$ in X and a sequence of nonnegative real numbers $t_0, \ldots, t_n$ satisfying $t_0 + \cdots + t_n = 1$ such that $x = t_0 x_0 + \cdots + t_n x_n$. By Proposition 5.1.4, x is in X'. This shows that $L(X) \subset C(X)$.

Now, we prove that $C(X)$ is contained in $L(X)$. It suffices to prove that $L(X)$ is affinely convex. Let x and y be two elements in $L(X)$. Then there exist nonnegative real numbers $t_0, \ldots, t_n$ and $t'_1, \ldots, t'_m$ satisfying $t_0 + \cdots + t_n = 1$ and $t'_1 + \cdots + t'_m = 1$ such that $x = \sum_{i=1}^{n} t_i x_i$ for some $x_0, \ldots, x_n$ in X, and $y = \sum_{i=1}^{m} t'_i y_i$ for some $y_0, \ldots, y_m$ in X. We must show that the affine segment $[x, y]$ is contained in $L(X)$. Let t be a real number in $[0, 1]$. Then

$$(1 - t)x + ty = (1 - t)t_0 x_0 + \cdots + (1 - t)t_n x_n + tt'_1 y_0 + \cdots + tt'_m y_m$$

and

$$(1-t)t_0 + \cdots + (1-t)t_n + tt'_1 + \ldots tt'_m = (1-t) + t = 1.$$

Therefore, the point $(1-t)x + ty$ is in $L(X)$, which shows that $[x, y] \subset L(X)$. This completes the proof of Proposition 5.2.2. □

For example, if x and y are two points in E, then $C(\{x, y\})$ is the affine segment $[x, y]$ joining these two points. More generally, one can show that if X and Y are two convex subsets of E, then $C(X \cup Y)$ is the union of all the affine segments of the form $[x, y]$ with x in X and y in Y.

The following result is usually referred to as Carathéodory's theorem.

Proposition 5.2.3 (Carathéodory). *If E is a vector space of dimension d, then, for every subset X of E, every element in the convex hull $C(X)$ is an affine convex combination of $d+1$ elements in X.*

Proof. Let x be in $C(X)$. By Proposition 5.2.2, we can find a nonnegative integer n such that there exist $n+1$ elements $x_0 \ldots, x_n$ in X and $n+1$ nonnegative real numbers $t_0, \ldots, t_n$ satisfying $t_0 + \cdots + t_n = 1$ and $x = t_0x_0 + \cdots + t_nx_n$. Let n be the smallest such integer. We claim that $n \leq d$, and this will prove the proposition.

We prove this claim by contradiction. Suppose that $n > d$. Then, the $n+1$ vectors $x_0, \ldots, x_n$ are linearly dependent and we can find $n+1$ real numbers $\lambda_0, \ldots, \lambda_n$ that are not all equal to 0 and such that $\lambda_0x_0 + \cdots + \lambda_nx_n = 0$. By elementary linear algebra, we can assume without loss of generality that $\lambda_0 + \cdots + \lambda_n = 0$. We then consider the following subset of $\mathbb{R}$:

$$K = \{t \in \mathbb{R} \text{ such that } t\lambda_i + t_i \geq 0 \text{ for all } i = 0, \ldots, n\}.$$

The set K, being the intersection of finitely many closed sets, is closed. It contains 0, and it is not equal to $\mathbb{R}$, since otherwise we would have $\lambda_i = 0$ for all $i = 0, \ldots, n$. Let L be the least upper bound of K. We have $L > 0$, again because otherwise we would have $\lambda_i = 0$ for all $i = 0, \ldots, n$. The real number L belongs to the frontier of K, which means that for each $i = 0, \ldots, n$, we have $L\lambda_i + t_i \geq 0$, and there exists $j \in \{0, \ldots, n\}$ such that $L\lambda_j + t_j = 0$. Now we can write

$$x = (L\lambda_0 + t_0)x_0 + \cdots + (L\lambda_n + t_n)x_n,$$

with the coefficient $L\lambda_j + t_j$ equal to 0, which contradicts the minimality in the definition of the integer n. We conclude that $n \leq d$, which completes the proof of Proposition 5.2.3. □

Proposition 5.2.4. *Let E be a topological vector space and let X be an arbitrary open subset of E. Then the convex hull $C(X)$ of X is open in E.*

Proof. Since X is contained in $C(X)$ and since X is open, X is in the interior $\overset{\circ}{C}(X)$ of $C(X)$. Since $C(X)$ is convex, its interior $\overset{\circ}{C}(X)$ is also convex (Proposition 5.1.11). Since $C(X)$ is contained in any convex set containing X, then $C(X) \subset \overset{\circ}{C}(X)$. Thus, $C(X) = \overset{\circ}{C}(X)$, which shows that $C(X)$ is open. □

Proposition 5.2.5. *Let E be a finite-dimensional topological vector space and let X be a compact subset of E. Then the convex hull $C(X)$ of X is compact.*

Proof. Let $(x_n)_{n\geq 0}$ be a sequence in $C(X)$ and let us prove that this sequence has a convergent subsequence. Let $d = \dim(E)$. By Proposition 5.2.3 (Carathéodory), for every $n = 0, 1, \ldots,$ we can write

$$x_n = \sum_{k=0}^{d} t_{n,k} p_{n,k}$$

with $p_{n,k} \in X$, $t_{n,k} \geq 0$ and $t_{n,1} + \ldots t_{n,d} = 1$ for every $k = 0, \ldots, d$.

The sequence $(p_{n,0})_{n\geq 0}$, being in the compact set X, has a convergent subsequence. Let $(p_{n_1,0})_{n\geq 0}$ be such a subsequence. Next, consider the sequence $(p_{n_1,1})_{n\geq 0}$ in X. It also has a convergent subsequence $(p_{n_2,1})_{n\geq 0}$. Then, consider the sequence $(p_{n_2,2})_{n\geq 0}$ and so on. Applying this reasoning $d+1$ times, we end up with a sequence $(n_d)_{n\geq 0}$ of nonnegative integers such that the $d+1$ sequences $(p_{n_d,0})_{n\geq 0}$, $(p_{n_d,1})_{n\geq 0}, \ldots,$ $(p_{n_d,d})_{n\geq 0}$ are convergent.

We consider now the sequence of real numbers $(t_{n_d,0})_{n\geq 0}$. Being in the compact interval $[0, 1]$, it has a convergent subsequence. Repeating the argument above $d+1$ more times, we end up with a sequence $(n_{2d+1})_{n\geq 0}$ of nonnegative integers such that the $d+1$ sequences $(p_{n_{2d+1},0})_{n\geq 0}$, $(p_{n_{2d+1},1})_{n\geq 0}$, $\ldots$, $(p_{n_{2d+1},d})_{n\geq 0}$ in X are convergent and the $d+1$ sequences $(t_{n_{2d+1},0})_{n\geq 0}$, $(t_{n_{2d+1},1})_{n\geq 0}, \ldots, (t_{n_{2d+1},d})_{n\geq 0}$ in $[0, 1]$ are convergent. Then (by taking a linear combination) it is easy to see that the subsequence $(x_{n_{2d+1}})_{n\geq 0}$ of $(x_n)_{n\geq 0}$ is convergent. □

After Propositions 5.2.4 and 5.2.5, it is good to note that the convex hull of a closed set is not always closed, as the following example shows:

Example 5.2.6. Let D be a straight line in $\mathbb{R}^2$, let P be a point in $\mathbb{R}^2$ that is not on D and let X be the subset $P \cup \{D\}$ of $\mathbb{R}^2$. Then $C(X) = R \cup D \cup \{P\}$ where R is the open region contained between the line D and the parallel line to D passing by P (Figure 5.2). Thus, X is a closed subset of $\mathbb{R}^2$ but its convex hull $C(X)$ is not closed.

Nevertheless, there is another notion of "convex hull" that can be useful and which, applied to a closed set, gives a closed set. It is defined as follows.

Definition 5.2.7 (Closed convex hull). Let E be a vector space and let X be an arbitrary subset of E. The *closed (affine) convex hull* of X, denoted by $\overline{C}(X)$, is the intersection of all the closed convex subsets of E that contain X.

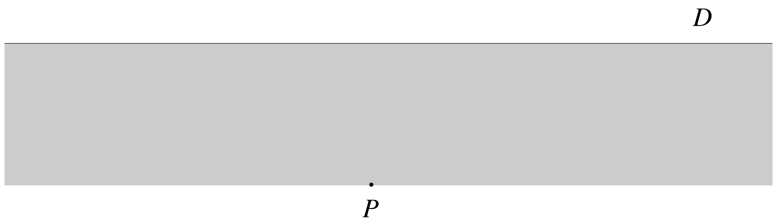


Figure 5.2. The convex hull of the line D and the point P.

As an intersection of convex sets, $\overline{C}(X)$ is convex, and as an intersection of closed sets, $\overline{C}(X)$ is closed. Furthermore, $\overline{C}(X)$ contains $C(X)$ since $C(X)$ is contained in any convex set that contains X, and, in fact, we have the following:

Proposition 5.2.8. *The closed convex hull $\overline{C}(X)$ is equal to the closure $\overline{C(X)}$ of the convex hull $C(X)$ of X.*

Proof. Since the set $\overline{C}(X)$ is a closed set that contains $C(X)$, it contains the closure $\overline{C(X)}$. Conversely, the closed set $\overline{C(X)}$ is convex (Proposition 5.1.9) and it contains $C(X)$, therefore it contains $\overline{C}(X)$. Thus, we have $\overline{C}(X) = \overline{C(X)}$. □

Definition 5.2.9 (Star-shaped subset). Let E a vector space, let X be a subset of E and let x be a point in X. We say that X is *star-shaped with respect to* x if for any y in X, the affine segment $[x, y]$ is contained in X.

It is plain from the definitions that a subset X of E is convex if and only if it is star-shaped with respect to any of its points.

Another useful notion is that of convex kernel, which in a certain sense is dual to that of convex hull. It is defined as follows:

Definition 5.2.10 (Convex kernel). Let E be a vector space and let X be a subset of E. The *convex kernel* of X is the set of points in X with respect to which X is star-shaped. We shall denote the convex kernel of X by $\mathrm{Ker}(X)$.

The notion of convex kernel is due to Hermann Brunn [23], and the following proposition is sometimes referred to as "Brunn's Theorem".

Proposition 5.2.11. *Let E a vector space. For every subset X of E, its convex kernel* $\mathrm{Ker}(X)$ *is convex.*

Proof. Let x and y be two distinct points in $\mathrm{Ker}(X)$. We must prove that $[x, y] \subset \mathrm{Ker}(X)$. Since X is star-shaped with respect to x, then for every point p in X we have $[x, p] \subset X$. Now, since X is star-shaped with respect to y, then X contains any segment joining y to a point on $[x, p]$. Thus, the triangle xyp is contained in X and for every u on $[x, y]$, the segment $[u, p]$ is contained in X. Since p is an arbitrary point in X, u is in $\mathrm{Ker}(X)$. □

The following step-by-step construction of the convex hull is analogous to the construction given in Proposition 2.5.5 in the context of uniquely geodesic metric spaces; it is usually attributed to Hermann Brunn.

Proposition 5.2.12. *Let E be a vector space and let X be a subset of E. We set $C_0(X) = X$ and for every $n \geq 0$, we let $C_{n+1}(X)$ be the union of all geodesic segments joining pairs of points in $C_n(X)$. Then, the convex hull $C(X)$ of X is given by*

$$C(X) = \bigcup_{n \geq 0} C_n(X).$$

Proof. The proof is the same as that of Proposition 2.5.5. (We note however that this proposition is not a special case of Proposition 2.5.5 since a vector space in general is not a uniquely geodesic space.) □

Related to this construction, one makes the following

Definition 5.2.13 (Brunn's number). The Brunn number of a subset X in a vector space E is the least element k in $\mathbb{N} \cup \infty$ such that

$$C(X) = \bigcup_{0 \leq n \leq k} C_n(X).$$

In the case where E is finite-dimensional, k is an element of $\mathbb{N}$ and in [22], Brunn gives a lower and an upper bound for k in terms of the dimension of E.

5.3 Convexity in normed vector spaces

We recall the definition of a norm. Of course, we assume that most of our readers are familiar with this notion, but we need to record its various components in order to refer to them later on.

Definition 5.3.1 (Norm). A *normed vector space* E is a vector space equipped with a map $x \mapsto \|x\|$ from E to $[0, \infty[$ (called the *norm*) that satisfies the following three properties:

- (homogeneity) $\|\lambda x\| = |\lambda|.\|x\|$ for all x in E and for λ in $\mathbb{R}$;
- (triangle inequality) $\|x + y\| \leq \|x\| + \|y\|$ for all x and y in E;
- (positive definiteness) $\|x\| = 0 \Rightarrow x = 0$.

A finite-dimensional normed vector space is called a *Minkowski space*.

The class of Minkowski spaces gives various interesting examples of metric spaces. In fact, Minkowski spaces are certainly the first non-Riemannian metric spaces whose geometry has been thoroughly studied. We should mention that these spaces are also important because as tangent spaces, they play in some sort the role of local spaces in Riemannian geometry and more generally in Finsler geometry.

A normed vector space is equipped with a canonical metric, defined by setting the distance between two points x and y to be equal to $\|x - y\|$. Equipped with this metric, a normed vector space is a geodesic metric space. More generally, we have the following

Proposition 5.3.2. *Let X be an affinely convex subset of a normed vector space E. Then X, equipped with the induced metric, is a geodesic metric space.*

Proof. For every x and y in X with $x \neq y$, let $\gamma\colon [0, \|x - y\|] \to E$ be the map defined by

$$t \mapsto \gamma(t) = \left(1 - \frac{t}{\|x - y\|}\right)x + \frac{t}{\|x - y\|}y.$$

Then for every t_1 and t_2 in $[0, \|x - y\|]$, we have

$$\begin{aligned}
|\gamma(t_1) - \gamma(t_2)|_E &= \Bigg\| \left(\left(1 - \frac{t_1}{\|x - y\|}\right)x + \frac{t_1}{\|x - y\|}y\right) \\
&\qquad - \left(\left(1 - \frac{t_2}{\|x - y\|}\right)x + \frac{t_2}{\|x - y\|}\right)\Bigg\| \\
&= \frac{1}{\|x - y\|}\|(t_2 - t_1)x + (t_1 - t_2)y\| \\
&= \frac{1}{\|x - y\|}|t_1 - t_2|.\|x - y\| \\
&= |t_1 - t_2|.
\end{aligned}$$

This shows that the path γ is geodesic. For every t in $[0, \|x - y\|]$, let us set $t' = t/\|x-y\|$. Since X is affinely convex and since t' is in $[0, 1]$, $\gamma(t) = (1-t')x+t'y$ is in X. We conclude that the image of γ is a geodesic segment in X joining x to y. This completes the proof of Proposition 5.3.2. □

Corollary 5.3.3. *A normed vector space is a geodesic metric space.*

Proof. This follows from Proposition 5.3.2 by taking $X = E$. □

We shall prove a result (Corollary 5.3.6 below) which implies that in principle, the study of metric spaces can be reduced to the study of complete normed vector spaces and their subsets. (Of course, this is only theoretical.)

Let X be an arbitrary metric space and let $\ell^\infty(X)$ be the vector space of bounded real-valued functions on X equipped with the ℓ^∞ norm, that is, the norm defined by $\|f\|_\infty = \sup_{x\in X} |f(x)|$ for any bounded function $f: X \to \mathbb{R}$.

Proposition 5.3.4. *The normed vector space $\ell^\infty(X)$ is complete.*

Proof. Let $(f_n)_{n\geq 0}$ be a Cauchy sequence in $\ell^\infty(X)$. We fix a positive real number ϵ. There exists an integer $N = N(\epsilon) \geq 0$ such that if m and n are integers that are $\geq N$, then $\|f_m - f_n\|_\infty \leq \epsilon$, that is, $sup_{x\in X}|f_m(x) - f_n(x)| \leq \epsilon$. Thus, for any fixed x in X, the sequence of real numbers $(f_n(x))_{n\geq 0}$ is a Cauchy sequence, and therefore it is convergent. Let $f(x)$ be its limit. For a fixed $n \geq N$, the sequence $(|f_m(x) - f_n(x)|)_{m\geq 0}$ converges to $|f(x) - f_n(x)|$ as $m \to \infty$. Thus, we have $|f(x) - f_n(x)| \leq \epsilon$ for all $n \geq N$, that is, $\sup_{x\in X} |f(x) - f_n(x)| \leq \epsilon$ for all $n \geq N$. This shows that the map $f_n - f$ is in $\ell^\infty(X)$, for all $n \geq N$. Since f_n is in $\ell^\infty(X)$, we have $f \in \ell^\infty(X)$. The last inequality also shows that $\|f - f_n\|_\infty \to 0$ as $n \to \infty$. Therefore, f_n converges to f as $n \to \infty$. This proves that $\ell^\infty(X)$ is complete. □

Let E be a normed vector space and let us choose a basepoint x_0 in X. We define an embedding $i: X \to E$ by taking as the image of each point x in X the map

$$i_x: X \to \mathbb{R}$$

defined by

$$i_x(y) = |x_0 - x|_X - |x - y|_X$$

for each y in X. Using the triangle inequality, we have $|i_x(y)| \leq |x_0 - x|_X$. Therefore, the map $i_x: X \to \mathbb{R}$ is in $\ell^\infty(X)$ for all x.

Proposition 5.3.5. *The map $i: X \to \ell^\infty(X)$ is distance-preserving.*

Proof. If x and x' are two arbitrary points in X, we have

$$\begin{aligned} |i_x - i_{x'}|_E &= \sup_{y\in X} \big||x_0 - y|_X - |x - y|_X - (|x_0 - y|_X - |x' - y|_X)\big| \\ &= \sup_{y\in X} \big||x - y|_X - |x' - y|_X\big| \\ &= |x - x'|_X. \end{aligned}$$

□

Corollary 5.3.6. *Any metric space admits a distance-preserving embedding in a complete vector space.*

Proof. This is a consequence of Propositions 5.3.4 and 5.3.5. □

Now we turn back to the study of affine segments in an arbitrary normed vector space.

Proposition 5.3.7. *Let E be a normed vector space and let x and y be two arbitrary points in E. Then the affine segment $[x, y]$ is a geodesic segment; in fact, it is the image of the geodesic path $\gamma : [0, \|x-y\|] \to X$ that joins x and y and that is defined by the formula*

$$\gamma(t) = \left(1 - \frac{t}{\|x-y\|}\right)x + \frac{t}{\|x-y\|}y.$$

Proof. The proof is contained in that of Proposition 5.3.2. □

Definition 5.3.8 (Affine geodesic). We shall call the geodesic γ of Proposition 5.3.7 the *affine geodesic* joining x and y

In an arbitrary normed vector space, there might exist other geodesic paths than the affine one, as we shall now see.

Even though in a finite-dimensional vector space all the norms define the same topology, the metrics associated to these norms can be quite different. For instance, the metric on $\mathbb{R}^n$ that is associated to the Euclidean norm, which is defined by $\|x\|_2 = (\sum_{i=}^n |x_i|^2)^{1/2}$, is uniquely geodesic (the affine segment joining two arbitrary points is the unique geodesic joining them), but neither the metric associated to the ℓ^1 norm, defined by $\|x\|_1 = \sum_{i=}^n |x_i|$, nor the one associated to the ℓ^∞ norm, defined by $\|x\|_\infty = \sup_{i=1,\dots,n} |x_i|$, is uniquely geodesic. Let us see some examples of distinct geodesic segments joining two points in these spaces. We can limit ourselves to the case of dimension 2.

Example 5.3.9 (Non-uniquely geodesic Minkowski spaces). Consider the metric on $\mathbb{R}^2$ induced by the ℓ^∞ norm. It is easy to see that any segment that has the form of an escalator (Figure 2.6 in Chapter 2), or, more generally (by passing to the limit), any segment that is the image of a monotonous function is a geodesic segment for that metric. Thus, if two points x and y in $\mathbb{R}^2$ have distinct first coordinates and distinct second coordinates, there are infinitely many geodesic segments joining them. Likewise, consider the metric on $\mathbb{R}^2$ induced from the ℓ^1 norm. We have represented in Figure 5.3 two distinct geodesic segments between the origin of $\mathbb{R}^2$ and the point x whose coordinates are $(1, 1)$: the affine segment $[0, x]$ and the segment obtained by concatenating the affine segments $[0, y]$ and $[y, x]$. In this figure, the square $ABCD$ is the unit ball of the norm ℓ^1 and the point y can be taken to be any point on the face AB of that ball.

It should be noted that since there is an automorphism of $\mathbb{R}^2$ sending the unit ball of the ℓ^1 norm onto the unit ball of the ℓ^∞ norm, the two spaces in Example 5.3.9 are isometric, and therefore showing that one of them is non-uniquely geodesic implies the same property for the other one. However, $\mathbb{R}^n$, for $n \geq 2$, equipped with the ℓ^1 norm is not isometric to $\mathbb{R}^n$ equipped with the ℓ^∞ norm. (We already mentioned that the standard cube and co-cube in dimension $n \geq 3$ are not isomorphic, see Example 5.1.8 above).

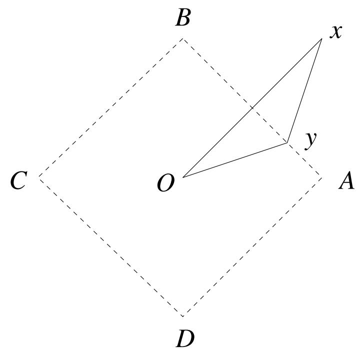


Figure 5.3. $\mathbb{R}^2$ is equipped with the ℓ^1 norm. Two distinct geodesic segments joining O to x. The square in dotted lines is the unit ball of the ℓ^1 norm.

The following proposition establishes a relation between affine convexity and Menger convexity (see Section 5 of Chapter 2).

Proposition 5.3.10. *Let E be a Minkowski space and let X be a closed subset of E equipped with the induced metric. Then X is Menger convex if and only if X is affinely convex.*

Proof. The vector space E, being finite-dimensional, is proper, and X being a closed subset of a proper space is itself proper for the induced metric. Then, Theorem 2.5.2 says that X is Menger convex if and only if it is a geodesic space, that is, if and only if two arbitrary points in X can be joined by a geodesic in X. A priori, this geodesic is not necessarily the affine geodesic in E that joins these two points, but the proof of Theorem 2.5.2 shows that we can take it to be the affine geodesic. Thus, the space X is Menger convex if and only if it is an affinely convex subset of E. □

In a normed vector space, there are convex subsets that are particularly interesting. The open and closed balls in such a space are examples of such sets, but there are several others.

We recall that for any subset A of a metric space X and for any nonnegative real number ϵ, the closed ϵ-neighborhood of A is the set

$$N(A, \epsilon) = \{x \in X \text{ such that } d_A(x) \leq \epsilon\}.$$

We now turn back to the special case where X is a normed vector space. It is easy to see that in this case, if B is an open ball in E of center x and radius $r > 0$, then, for any $\epsilon \geq 0$, $N(B, \epsilon)$ is the closed ball in E of center x and radius $r + \epsilon$.

The following description of the closed ϵ-neighborhood of a set is useful:

Proposition 5.3.11. *Let E be a normed vector space and let U denote the closed unit ball in X. Then for every subset A of X and for every $\epsilon \geq 0$, we have*

$$N(A, \epsilon) = A + \epsilon U.$$

Proof. The proof follows easily from the definitions. □

Proposition 5.3.12. *Let E be a normed vector space. Then, for every convex subset A of X, the closed ϵ-neighborhood $N(A, \epsilon)$ of A is convex.*

Proof. We first consider the case where A is a single point x, and therefore $N(A, \epsilon)$ is the open ball B of center x and radius ϵ. For every y and z in B and for every t in $[0, 1]$, we have

$$\begin{aligned}\|x - \big((1-t)y + tz\big)\| &= \|(1-t)(x-y) + t(x-z)\| \\ &\leq \|(1-t)(x-y)\| + \|t(x-z)\| \\ &= (1-t)\|x-y\| + t\|x-z\| \\ &\leq (1-t+t)\epsilon = \epsilon.\end{aligned}$$

Thus, $(1-t)y + tz$ is in B, which shows that B is convex. In the case where A is an arbitrary convex subset of E, we have, by Proposition 5.3.11, $N(A, \epsilon) = A + \epsilon U$ where U is the unit closed ball in E which, by the special case we considered, is convex. Using the fact that convexity is preserved by homothety and by vector sum of subsets, we deduce that $N(A, \epsilon)$ is convex. □

Let us note two consequences.

We already mentioned the notion of capsule in $\mathbb{R}^n$ (see Figure 4.1 of Chapter 4). A *capsule* in a vector space is an ϵ-neighborhood of some geodesic segment, for some $\epsilon \geq 0$. We shall again encounter this notion in Chapter 9.

We deduce the following from Proposition 5.3.12:

Corollary 5.3.13. *Let E be a normed vector space. Then, any capsule in E is convex.*

Proposition 5.3.14. *Let E be a normed vector space. Then, any closed or open ball in E is convex.*

Proof. The case of a closed ball is contained in Proposition 5.3.12. For an open ball, the proof is the same as for a closed ball, up to replacing some of the large inequalities by strict inequalities. □

5.4 Limits of convex sets

Let E be a normed vector space. We recall that if X and Y are subsets of E, then their Hausdorff distance is equal to

$$d_{\mathcal{H}}(X, Y) = \sup_{x \in E} |d_X(x) - d_Y(x)|,$$

where $d_X(x)$ denotes the distance from the point x to the set X. In Chapter 4, we studied general properties of Hausdorff distances between subsets of arbitrary metric spaces. For subsets of vector spaces, the Hausdorff distance satisfies further properties, for instance, the following one, which says that the Hausdorff distance is translation-invariant:

Proposition 5.4.1. *Let E be a normed vector space and let X and Y be two subsets of E. Then, for any vector v in E, we have*

$$d_{\mathcal{H}}(X, Y) = d_{\mathcal{H}}(X + v, Y + v).$$

Proof. Since the distance $|x - y|$ between any two points x and y in E is translation-invariant, we have

$$\begin{aligned} d_{\mathcal{H}}(X, Y) &= \sup_{x \in E} |d_X(x) - d_Y(x)| \\ &= \sup_{x \in E} |d_X(x + v) - d_Y(x + v)| \\ &= d_{\mathcal{H}}(X + v, Y + v) \end{aligned}$$

□

Proposition 5.4.2. *Let E be a finite-dimensional normed vector space and let X and Y be two nonempty subsets of E. Then, we have*

$$d_{\mathcal{H}}(C(X), C(Y)) \leq d_{\mathcal{H}}(X, Y).$$

Proof. Let $\delta = d_{\mathcal{H}}(X, Y)$. Since E is finite-dimensional, it is proper, and by Proposition 4.1.6 , we have

$$Y \subset N(X, \delta) \subset N(C(X), \delta).$$

By Proposition 5.3.12, $N(C(X), \delta)$ is convex. Therefore, we have $C(Y) \subset N(C(X), \delta)$. By symmetry, we have also $C(X) \subset N(C(Y), \delta)$. This implies $d_{\mathcal{H}}(C(X), C(Y)) \leq \delta$.

□

We obtain immediately the following

Corollary 5.4.3. *Let E be a finite-dimensional normed vector space, let $(X_n)_{n \geq 0}$ be a sequence of nonempty subsets of E and A a nonempty subset of E such that $d_{\mathcal{H}}(X_n, A) \to 0$ as $n \to \infty$. Then, $d_{\mathcal{H}}(C(X_n), C(A)) \to 0$.* □

We recall that the Hausdorff distance restricted to the set of nonempty closed bounded subsets of E (which is also the set of compact subsets of E, since E is finite-dimensional) is a genuine metric (Proposition 4.1.1). Since in a metric space limits of subsequences are unique, we have the following consequence of Corollary 5.4.3:

Corollary 5.4.4. *Let E be a finite-dimensional normed vector space and let $(X_n)_{n\geq 0}$ be a sequence of nonempty compact convex subsets of E that is convergent with respect to the Hausdorff metric. Then, the limit of this sequence is convex.*

□

5.5 Minkowski's construction

We already observed that in any normed vector space, the open and closed unit balls are affinely convex (Proposition 5.3.13). Conversely, there is a beautiful construction due to Minkowski (cf. [111]) that associates a norm to any (say closed) convex subset B of a vector space containing the origin and satisfying certain natural conditions that we now state. B will be the unit ball for that norm. In this section, we describe this construction.

First, we recall a few classical definitions.

If B is a subset of a metric space, then the *frontier* of B is the set $\overline{B} \setminus \overset{\circ}{B}$.

Let E be a vector space and let x be a point in E that is distinct from the origin. Then the (*linear*) *ray* in E passing through x is the subset r_x of E defined as

$$r_x = \{\lambda x, \text{ with } \lambda \geq 0\}.$$

The next proposition follows easily from the definitions:

Proposition 5.5.1. *Let E be a vector space with origin O and let B be a subset of E that is star-shaped with respect to O and that contains O in its interior. Then, for every x in E that is distinct from O, the intersection of B with the ray r_x is an affine segment of the form $[O, e(x)]$ where $e(x)$ is a point in the frontier of B and where the open affine segment $]O, e(x)[$ is contained in $\overset{\circ}{B}$.*

Proof. Since B is star-shaped with respect to O, the intersection of B with the ray r_x is convex. Since B is bounded, this intersection is bounded and therefore it is an affine segment. Thus, it is of the form $[O, e(x)]$. By Lemma 5.1.10, any point on the open segment $]O, e(x)[$ is in $\overset{\circ}{B}$. Therefore, the point $e(x)$ is the unique point in $B \cap r_x$ that is in the frontier of B. □

The notion that we now introduce is due to Minkowski. It is important in convexity theory, and we shall use it below.

Definition 5.5.2 (Convex body). Let E be a topological vector space. A *convex body* is a convex subset of E that has non-empty interior.

We start with the following result.

Proposition 5.5.3. *Let B be a convex body in E. Then $\overset{\circ}{B}$ is dense in B.*

Proof. Let x_0 be a point in B and let x be an interior point of B. Without loss of generality, we can assume that x is distinct from x_0. By convexity, the segment $[x, x_0]$ is contained in B and, by Lemma 5.1.10, any point on the open segment $]x, x_0[$ is in $\overset{\circ}{B}$. Thus, any point in B is a limit of a sequence of points in $\overset{\circ}{B}$. This completes the proof of Proposition 5.5.3. □

In the rest of this section, E is a finite-dimensional vector space, and its origin is denoted by O.

Given a convex body B in E, we shall consider the following three properties:

- B is closed and bounded. (We note that boundedness of B means that this set has finite diameter for some norm on E or, equivalently, for any norm on E.)
- B contains the origin O in its interior.
- B is symmetric with respect to O. In other words, for all x in E, we have the following equivalence: $x \in B \iff -x \in B$.

Definition 5.5.4 (The Minkowski function). Let B be a convex body in E satisfying the three properties above. The *Minkowski function*

$$\mu_B \colon E \to [0, \infty[$$

associated to B is defined by

$$\mu_B(x) = \begin{cases} 0 & \text{if } x = O \\ \lambda(x) & \text{if } x \neq O, \end{cases}$$

where $\lambda(x)$ is the unique real number satisfying $x = \lambda(x)e(x)$, using the notations of Proposition 5.5.1.

One can easily see that for every x in E that is distinct from the origin, we have

$$\mu_B(x) = \inf \left\{ \lambda > 0 \text{ such that } \frac{x}{\lambda} \in B \right\}. \tag{5.5.4.1}$$

From this, we deduce the following useful property:

For every $\lambda > 0$ and for every x in X, we have

$$\frac{x}{\lambda} \in B \iff \mu_B(x) \leq \lambda. \tag{5.5.4.2}$$

Theorem 5.5.5. *Let B be a convex body in E satisfying the three properties that are stated after Definition 5.5.3. Then the associated Minkowski function μ_B is a norm on E. Furthermore, B is the closed unit ball of this norm, the interior of B is its open unit ball and the frontier of B is its unit sphere.*

Proof. We must show that the three properties of Definition 5.3.1 (Norm) are satisfied.

Let us start by showing that the map μ_B is homogeneous. First of all, it is clear from the definition of μ_B that this map is positively homogeneous, that is, that for all $\lambda > 0$ and for all x in B, we have $\mu_B(\lambda x) = \lambda\mu_B(x)$. As the convex body B is symmetric with respect to the origin, we have $\mu_B(-x) = \mu_B(x)$. From this, we deduce that $\mu_B(\lambda x) = \lambda\mu_B(x)$ for all x in B and for all $\lambda \in \mathbb{R}$, which proves the homogeneity.

Next, let us prove that μ_B satisfies the triangle inequality. Let us fix a positive real number ϵ, let x and y be two points in E and let us choose two real numbers λ and μ satisfying

$$\mu_B(x) < \lambda < \mu_B(x) + \epsilon$$

and

$$\mu_B(y) < \mu < \mu_B(y) + \epsilon.$$

By (5.5.4.2), the points x/λ and y/μ are in B. Let

$$z = \frac{x+y}{\lambda+\mu} = \left(\frac{\lambda}{\lambda+\mu}\right)\frac{x}{\lambda} + \left(\frac{\mu}{\lambda+\mu}\right)\frac{y}{\mu}.$$

The point z belongs to the affine segment $[x/\lambda, y/\mu]$. Since B is convex, z is in B. Therefore, using again (5.5.4.2), we have

$$\mu_B(x+y) \leq \lambda + \mu < \mu_B(x) + \mu_B(y) + 2\epsilon.$$

Since ϵ is an arbitrary positive number, we conclude that $\mu_B(x+y) \leq \mu_B(x)+\mu_B(y)$, which proves the desired property.

Finally, let us show that the map μ_B is positive definite. Let x be a point in E that is distinct from the origin. Since B is compact, we can find a positive real number λ_0 such that for all λ in $[0, \lambda_0]$, we have $x/\lambda \notin B$, which implies (using 5.5.4.2) that $\mu_B(x) \geq \lambda_0$. Therefore, we have $\mu_B(x) > 0$. Thus, the map μ_B is a norm on E.

By (5.5.4.2), we have the equivalence $x \in B \iff \mu_B(x) \leq 1$, which shows that B is the closed unit ball of the norm.

Let S be the frontier of B and let us show that S is the unit sphere of μ_B, *i.e.*, that $S = \{x \in E : \mu_B(x) = 1\}$. This will also imply that the interior of B is the open unit ball of μ_B.

Let x be a point in E such that $\mu_B(x) = 1$. By (5.5.4.1), we can find a sequence of real positive numbers $(\lambda_i)_{i\geq 0}$ satisfying $x/\lambda_i \in B$ for all $i \geq 0$, with $\lambda_i \to 1$ as $i \to \infty$. Since $\lambda_i \to 1$, the sequence (x/λ_i) converges to x, and therefore x is in $\overline{B}$. On the other hand, x is not in $\overset{\circ}{B}$, since the sequence $(\lambda_i x)$, which is in the complement of B, also converges to x. Therefore, x is in S.

Now suppose that for some x in E, we have $\mu_B(x) < 1$. Then we have $x \in [O, e(x)[$, where $e(x)$ belongs to B. By Lemma 5.1.10, we obtain $x \in \overset{\circ}{B}$. This implies that B is the open unit ball of the norm μ_B. This completes the proof of Theorem 5.5.5. □

Thus, all the beautiful convex bodies in $\mathbb{R}^3$ are the unit balls of certain norms on this space. For instance, the regular octahedron centered at the origin and of radius 1 is the unit ball for the ℓ^1 norm, that is, the norm defined by $\|(x_1, x_2, x_3)\| = |x_1|+|x_2|+|x_3|$. In fact, this octahedron is the standard co-cube of $\mathbb{R}^3$.

There are several interesting metric characterizations of Minkowski spaces that are contained in the two books by Busemann [24] and [28]; see also [27]. We note also that the Minkowski construction gives a Euclidean norm if and only if the convex body B is an ellipsoid (see [24] p. 58).

As an application of Minkowski's construction, we give the following proposition which besides being important in itself, gives a preliminary taste of a general result that we shall prove later on, that says that any convex metric space convex is contractible (Proposition 8.1.9). We note by the way that this proposition remains valid without the hypothesis that A is bounded.

Proposition 5.5.6. *Let A be an open nonempty bounded convex subset of E. Then A is homeomorphic to E.*

Proof. By performing a translation, we can assume, without loss of generality, that the set A contains the origin of E. Let B be the closure of A. Then B is a closed convex body satisfying the hypotheses of Definition 5.5.4. Let $f: E \to E$ be the map defined by setting, for all x in E,

$$f(x) = \frac{x}{\mu_B(x) + 1},$$

where $\mu_B: E \to [0, \infty[$ is Minkowski's function that is associated to B. Since μ_B is homogeneous, we have, for all x in E,

$$\mu_B(f(x)) = \frac{\mu_B(x)}{\mu_B(x) + 1} < 1.$$

By Theorem 5.5.5, we obtain $f(x) \in \overset{\circ}{B} = A$. Now let us show that any point in A is in the image of f. Given y in A, we set

$$x = \frac{y}{1 - \mu_B(y)}.$$

Then we have

$$f(x) = \frac{x}{\mu_B(x) + 1} = \frac{y/(1 - \mu_B(y))}{(\mu_B(y)/1 - \mu_B(y)) + 1} = y,$$

which shows that y is in the image of f. This formula also shows that as a map from E to A, f is continuous and possesses an inverse map, defined by

$$x \mapsto \frac{x}{1-\mu_B(x)},$$

which is clearly continuous. Thus, f defines a homeomorphism between E and A. This completes the proof of Proposition 5.5.6 □

5.6 The Hilbert geometry

In all this section, by a *line* we mean a Euclidean line.

We start by recalling the definition of the cross ratio.

Definition 5.6.1 (Cross ratio). Consider four ordered points a_1, a_2, b_1, b_2 in $\mathbb{R}^n$ satisfying $a_1 \neq b_1$ and $a_2 \neq b_2$. Then the *cross ratio* $[a_1, a_2, b_2, b_1]$ is defined by the formula

$$[a_1, a_2, b_2, b_1] = \frac{d(a_2, b_1)d(a_1, b_2)}{d(a_1, b_1)d(a_2, b_2)},$$

where d denotes the Euclidean metric of $\mathbb{R}^n$.

In all that follows, we shall only use the cross ratio for points that are aligned. We start with a elementary property, which is usually referred to as the "cocycle property".

To avoid lengthy considerations of special cases, in the rest of this chapter, each time we prove a property involving the cross ratio of four points, we shall tacitly assume that these points are pairwise distinct. One can easily deal with the cases where some of the points coincide, when such considerations are needed.

Proposition 5.6.2 (Cocycle property). *Let b_1, a_1, a_2, a_3, b_2 be five points in $\mathbb{R}^n$. Then we have*

$$[a_1, a_2, b_2, b_1] \times [a_2, a_3, b_2, b_1] = [a_1, a_3, b_2, b_1].$$

Proof. From the definition of the cross ratio, we have

$$[a_1, a_2, b_2, b_1] = \frac{d(a_2, b_1)d(a_1, b_2)}{d(a_1, b_1)d(a_2, b_2)}$$

and

$$[a_2, a_3, b_2, b_1] = \frac{d(a_3, b_1)d(a_2, b_2)}{d(a_2, b_1)d(a_3, b_2)}.$$

We then obtain

$$[a_1, a_2, b_2, b_1] \times [a_2, a_3, b_2, b_1] = \frac{d(a_3, b_1)d(a_1, b_2)}{d(a_1, b_1)d(a_3, b_2)} = [a_1, a_3, b_2, b_1],$$

which proves Proposition 5.6.2. □

We shall use an invariance property of the cross ratio, and before stating it we need to recall the following classical notion:

Definition 5.6.3 (Perspectivity). Let b_1, a_1, a_2, b_2 be four points in $\mathbb{R}^n$ situated in that order on a line D, let b'_1, a'_1, a'_2, b'_2 be four points in the same space that are also contained in a line and let O be an arbitrary point in the extended space $\mathbb{R}^n \cup \{\infty\}$. We say that the ordered quadruple b_1, a_1, a_2, b_2 is *obtained from the ordered quadruple b'_1, a'_1, a'_2, b'_2 by a perspectivity of center O* if b_1, a_1, a_2, b_2 is the image of b'_1, a'_1, a'_2, b'_2 by the projection of center O on the line D. The restriction of this projection to the set $\{b'_1, a'_1, a'_2, b'_2\}$ is called a *perspectivity of center O*.

In Figures 5.4 and 5.5 respectively, we have represented the effect of a perspectivity in the case where O is a point in $\mathbb{R}^n$ and in the case where O is the point ∞.

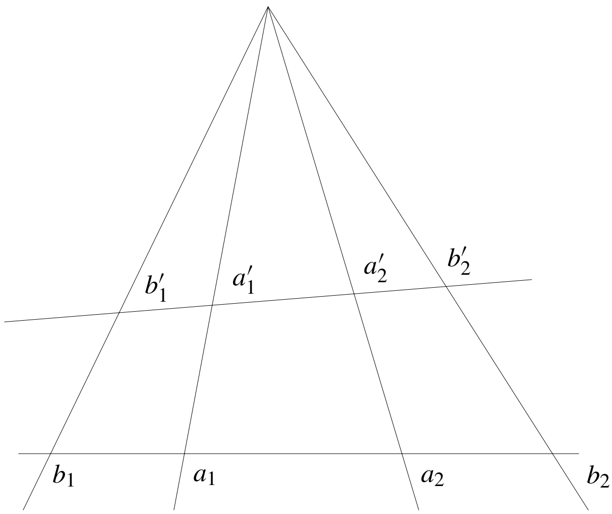


Figure 5.4. A perspectivity.

Proposition 5.6.4 (Cross ratio is invariant under perspectivities). *Let b'_1, a'_1, a'_2, b'_2 be an ordered quadruple of aligned pairwise distinct points in $\mathbb{R}^n$ and let b_1, a_1, a_2, b_2 be an ordered quadruple that is obtained from b'_1, a'_1, a'_2, b'_2 by a perspectivity of center $O \in \mathbb{R}^n \cup \{\infty\}$. Then we have $[a_1, a_2, b_2, b_1] = [a'_1, a'_2, b'_2, b'_1]$.*

Proof. We first consider the case where $O \neq \infty$. The proof is based on the computation of the areas of the triangles Oa_2b_1, Oa_1b_2, Oa_1b_1 and Oa_2b_2 in Figure 5.4. Let L be the distance from the point O to the line D. This distance is a height for each of

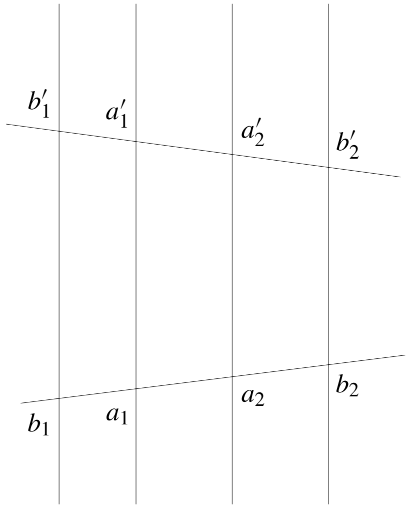


Figure 5.5. A perspectivity whose center is the point at infinity.

these triangles and therefore we have:

$$\text{Area}(Oa_2b_1) = \frac{1}{2}Ld(a_2, b_1) = \frac{1}{2}d(O, a_2)d(O, b_1)\sin\angle a_2Ob_1,$$

$$\text{Area}(Oa_1b_2) = \frac{1}{2}Ld(a_1, b_2) = \frac{1}{2}d(O, a_1)d(O, b_2)\sin\angle a_1Ob_2,$$

$$\text{Area}(Oa_1b_1) = \frac{1}{2}Ld(a_1, b_1) = \frac{1}{2}d(O, a_1)d(O, b_1)\sin\angle a_1Ob_1$$

and

$$\text{Area}(Oa_2b_2) = \frac{1}{2}Ld(a_2, b_2) = \frac{1}{2}d(O, a_2)d(O, b_2)\sin\angle a_2Ob_2.$$

Thus, we obtain

$$\begin{aligned}
&[a_1, a_2, b_2, b_1] \\
&\quad = \frac{d(a_2, b_1)d(a_1, b_2)}{d(a_1, b_1)d(a_2, b_2)} \\
&\quad = \frac{\big(d(O, a_2)d(O, b_1)\sin\angle a_2Ob_1/L\big)\big(d(O, a_1)d(O, b_2)\sin\angle a_1Ob_2/L\big)}{\big(d(O, a_1)d(O, b_1)\sin\angle a_1Ob_1/L\big)\big(d(O, a_2)d(O, b_2)\sin\angle a_2Ob_2/L\big)} \\
&\quad = \frac{\sin\angle a_2Ob_1 \sin\angle a_1Ob_2}{\sin\angle a_1Ob_1 \sin\angle a_2Ob_2}.
\end{aligned}$$

This shows that the cross ratio $[a_1, a_2, b_2, b_1]$ depends only on the angles that the four Euclidean lines Oa_1, Oa_2, Ob_1 and Ob_2 make at the point O, and therefore it is equal to $[a'_1, a'_2, b'_2, b'_1]$.

In the case where $O = \infty$, the perspectivity of center O is simply a parallel projection (Figure 5.5), and instead of the above computation, an application of the theorem of Thales gives the result. □

Let A be a nonempty bounded open convex subset of $\mathbb{R}^n$. For any pair of distinct points a_1, a_2 in A, consider the Euclidean line D containing them. This line intersects the boundary of A in exactly two points (as follows for instance from Lemma 5.1.10). Let us denote by b_1 and b_2 these two intersection points, the names being chosen in such a way that the four points b_1, a_1, a_2, b_2 are aligned in that order on D (Figure 5.6). We define the map $h_A : A \times A \to \mathbb{R}$ by the formula

$$h_A(a_1, a_2) = \begin{cases} \ln[a_1, a_2, b_2, b_1] & \text{if } a_1 \neq a_2, \\ 0 & \text{if } a_1 = a_2. \end{cases}$$

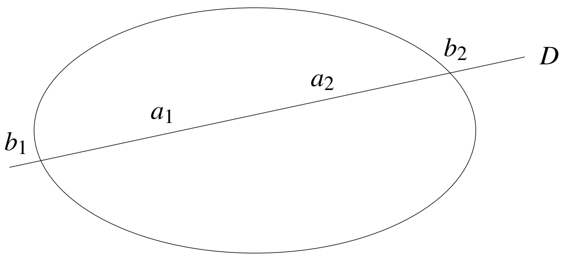


Figure 5.6

It follows from the fact that the four points b_1, a_1, a_2, b_2 are aligned in that order that we have $d(a_2, b_1)/d(a_1, b_1) > 1$ and $d(a_1, b_2)/d(a_2, b_2) > 1$, which implies $[a_1, a_2, b_2, b_1] > 1$. Therefore $h_A(a_1, a_2) > 0$ if $a_1 \neq a_2$. We shall prove that h_A defines a metric on A. The definition of this metric is due to David Hilbert who described it in [69], and h_A is referred to as the *Hilbert metric* of A.

We first consider the case $n = 1$. Here, A is an open finite-length interval of $\mathbb{R}$. Although the next proposition is a special case of Theorem 5.6.6 below, we start by proving it because we shall use it in the proof of that theorem.

Proposition 5.6.5 (The Hilbert metric of a bounded open interval). *Let I be a finite-length open interval of $\mathbb{R}$. Then the map $h_I : I \times I \to [0, \infty[$ is a distance function, and the metric space (I, h_I) is a proper geodesic metric space that is isometric to the real line $\mathbb{R}$ equipped with its usual metric.*

Proof. We already saw that $h_I(a_1, a_2) \geq 0$ for every a_1 and a_2 in A and that $h_I(a_1, a_2) > 0$ if and only if $a_1 \neq a_2$. To prove symmetry, we call b_1 and b_2, with

$b_1 < b_2$, the endpoints of the interval I. We can suppose, without loss of generality, that $a_1 < a_2$. Then we have

$$h_I(a_1, a_2) = \ln[a_1, a_2, b_2, b_1] = \ln\left(\frac{d(a_2, b_1)d(a_1, b_2)}{d(a_1, b_1)d(a_2, b_2)}\right)$$

and

$$h_I(a_2, a_1) = \ln[a_2, a_1, b_1, b_2] = \ln\left(\frac{d(a_1, b_2)d(a_2, b_1)}{d(a_2, b_2)d(a_1, b_1)}\right).$$

This shows that $h_I(a_1, a_2) = h_I(a_2, a_1)$. Finally, for the proof of the triangle inequality, we consider three points a_1, a_2 and a_3 in A and we suppose without loss of generality that $a_1 < a_2 < a_3$. By Proposition 5.6.2 we have

$$[a_1, a_2, b_2, b_1] \times [a_2, a_3, b_2, b_1] = [a_1, a_3, b_2, b_1]$$

which implies

$$\ln[a_1, a_2, b_2, b_1] + \ln[a_2, a_3, b_2, b_1] = \ln[a_1, a_3, b_2, b_1].$$

Thus, we have the (degenerate) triangle inequality

$$h_I(a_1, a_3) = h_I(a_1, a_2) + h_I(a_2, a_3).$$

Thus proves that h_I is a metric on I. It is clear from the definition of this metric that the closed balls are compact and therefore the space (I, h_I) is proper. From the equality $h_I(a_1, a_3) = h_I(a_1, a_2) + h_I(a_2, a_3)$ for any a_1 and a_2 and a_3 satisfying $a_1 < a_2 < a_3$, it follows that the metric space (I, h_I) is a geodesic metric space. (We can apply the criterion that uses Menger convexity, Theorem 2.5.2.) Finally, since $\lim h_A(a_1, a_2) = \infty$ as $a_1 \to b_1$ and $\lim h_I(a_1, a_2) = \infty$ as $a_2 \to b_2$, we can construct, for any point in I, a geodesic line in I that starts at this point and which establishes an isometry between $\mathbb{R}$ and the space (I, h_I). This completes the proof of Proposition 5.6.5. □

Next, we consider the case where the domain A is the interior of a triangle, and we prove the following

Proposition 5.6.6 (The case of a triangle).[2] *Let w, u and t be three non-collinear points in $\mathbb{R}^2$ and let A be the open region bounded by the triangle wut. Let v be a point on the segment $[w, t]$, let z be a point on the segment $[u, w]$ and let c be the intersection of the segments $[u, v]$ and $[z, t]$. Finally, let a and b be two points on the segments $[u, c]$ and $[t, c]$ respectively. Then we have*

$$h_A(a, b) = h_A(a, c) + h_A(c, b).$$

[2] The Hilbert metric of a triangle has beautiful properties. For instance, B. B. Phadke proved in [118] that circles in such a space are hexagons. P. de la Harpe showed in [64] that the isometry group of that space is isometric to $\mathbb{R}^2$ equipped with a norm whose unit ball is a regular hexagon. The paper [64] contains other interesting results on the isometry group of a simplex, some of them valid in all dimensions, as well as several open problems.

Proof. Let a' and b' be the intersection of $[a, b]$ with $[u, w]$ and $[t, w]$ respectively and let d be the intersection point of $[w, c]$ with $[a, b]$ (Figure 5.7). By the invariance of the cross ratio by perspectivities (Proposition 5.6.4), we have $[a, c, v, u] = [a, d, b', a']$ and $[c, b, t, z] = [d, b, b', a']$. By the cocycle property (Proposition 5.6.2), we have

$$[a, b, b', a'] = [a, d, b', a'] \times [d, b, b', a'].$$

Therefore we obtain

$$[a, b, b', a'] = [a, c, v, u] \times [c, b, t, z],$$

which implies

$$\ln[a, b, b', a'] = \ln[a, c, v, u] + \ln[c, b, t, z],$$

that is, $h_A(a, b) = h_A(a, c) + h_A(c, b)$. □

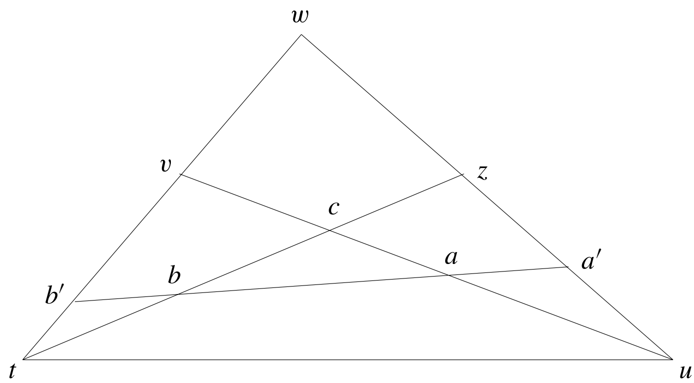


Figure 5.7

Theorem 5.6.7 (The Hilbert metric of a bounded open convex body). *For each non-empty bounded open convex subset A of $\mathbb{R}^n$, the associated map $h_A \colon A \times A \to [0, \infty[$ is a metric, the metric space (A, h_A) is a proper geodesic space and each affine segment in A is a geodesic segment for the metric h_A. Furthermore, the metric space (A, h_A) is uniquely geodesic if and only if the (Euclidean) boundary of the convex set A does not contain any pair of affine segments that span a two-dimensional affine plane.*

Proof. All the properties that make h_A a metric are contained in Proposition 5.6.5 (The Hilbert metric of a bounded open interval), except the triangle inequality. For the proof of this property, we follow Hilbert's argument contained in [69]. Consider three distinct points a, b and c in A. By Proposition 5.6.5, if these points are collinear, then the three distances between them satisfy (degenerate) triangle inequalities. Thus we now suppose that these three points are not collinear. We prove that in this case, the values of the map h_A on the three pairs of points (a, b), (a, c) and (b, c) satisfy the triangle inequalities and that these inequalities are strict if the boundary of A does not contain two affine segments whose span is a two-dimensional affine plane.

By taking the intersection of the convex body A with the affine plane containing the three points a, b and c, we are reduced to the case where A is 2-dimensional. We use the following notations (see Figure 5.8), which are those used by Hilbert in [69]:

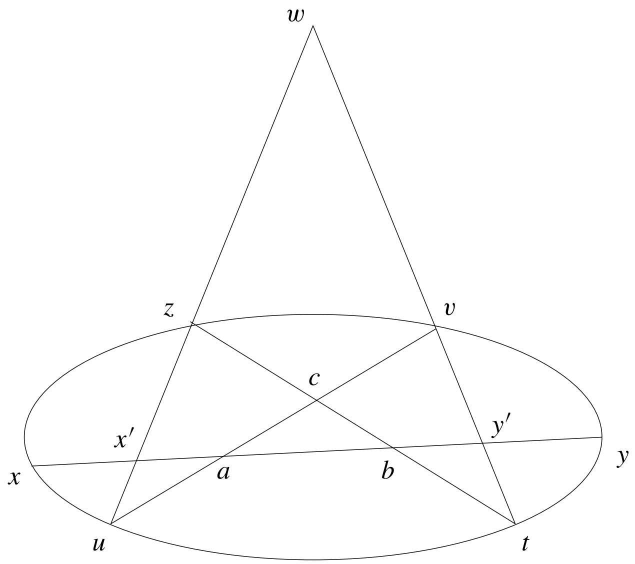


Figure 5.8

The points x and y are the intersection points of the segment $[a, b]$ with the boundary of A and the points x, a, b, y are aligned in that order.

The points u and t are the intersection points of the segment $[a, c]$ with the boundary of A and the points u, a, c, v are aligned in that order.

The points z and t are the intersection points of the segment $[c, b]$ with the boundary of A and the points z, c, b, t are aligned in that order.

The point x' is the intersection point of the segments $[u, z]$ and $[x, y]$.

The point y' is the intersection point of the segments $[t, v]$ and $[x, y]$.

The point w is the intersection point of the lines uz and tv. (It is possible that this point w is the point ∞.)

We have

$$h_A(a, b) = \ln[a, b, y, x],$$

$$h_A(a, c) = \ln[a, c, v, u],$$

and

$$h_A(c, b) = \ln[c, b, t, z].$$

Let B denote the interior of the triangle wut. Then B is a bounded open convex subset of the plane A, and it is equipped with its own map $h_B : B \times B \to [0, \infty[$. We can see directly from the definitions of the maps h_A and h_B that we have $h_A(a, c) = h_B(a, c)$ and $h_B(c, b) = h_B(c, b)$. Now let us compare the values $h_A(a, b)$ and $h_B(a, b)$. We have

$$[a, b, y', x'] = \frac{d(b, x')}{d(a, x')} \frac{d(a, y')}{d(b, y')} \leq \frac{d(b, x)}{d(a, x)} \frac{d(a, y)}{d(b, y)} = [a, b, y, x].$$

Thus, we obtain

$$h_A(a, b) \leq h_B(a, b) \tag{5.6.7.1}$$

By Proposition 5.6.6, we have $h_B(a, b) = h_B(a, c) + h_B(c, b)$. Thus, we obtain $h_A(a, b) \leq h_A(a, c) + h_A(c, b)$, which completes the proof of the fact that h_A is a metric on A. The closed balls for this metric are clearly compact and therefore this metric is proper.

Proposition 5.6.5 implies that any affine segment in A is a geodesic segment for the metric h_A. Thus, (A, h_A) is a geodesic metric space. Furthermore, the proof of Inequality (5.6.7.1) shows that this inequality is strict unless we have $x' = x$ and $y' = y$, that is, unless the affine segments $[u, z]$ and $[t, v]$ are contained in the boundary of A. Thus, if there are no affine segments in the boundary of A that span a 2-dimensional plane, the triangle inequality $h_A(a, b) \leq h_A(a, c) + h_A(c, b)$ is strict provided the three points a, b and c are not collinear. We conclude that in this case, the affine segments in A are the unique geodesic segments for the metric h_A.

Conversely, suppose that there exist two affine segments in the boundary of A whose span is an affine plane. Consider the plane spanned by these segments. In that plane, we can find, using the preceding argument, three non-collinear points a, b and c such that Inequality (5.6.7.1) is an equality, and therefore the distances between these points satisfy $h_A(a, b) = h_A(a, c) + h_A(c, b)$. This implies that the union of the two affine segments $[a, c]$ and $[c, b]$ is a geodesic segment joining a and c and therefore that the affine segment $[a, b]$ is not the unique geodesic segment joining these points. Thus, in the case considered, the metric space (A, h_A) is not uniquely geodesic. This completes the proof of Theorem 5.6.7. □

Notes on Chapter 5

Convexity. Convexity is one of the most beautiful subjects in mathematics. As a field in itself, convexity theory started at the end of the 19th century, but its roots are in Greek antiquity. For instance, the Pythagoreans of the fifth century B.C. already knew about the classification of the regular solid convex polyhedra in 3-dimensional space, and Archimedes, in the third century B.C., did substantial work on the properties of convex subsets of the plane. In fact, we owe to him the first definition of a convex curve in the plane and of a convex surface in 3-space. In his work *On the sphere and the cylinder* (see [67]), Archimedes studies lines in the plane which he describes as "concave in the same direction". He characterizes such a line by the property that "if any two points on it are taken, then either all the straight lines connecting the points fall on the same side of the line, or some fall on one and the same side while others fall on the line itself, but none on the other side". (This is Heath's translation, cf. [67] p. 2.) Of course, the work of Archimedes on centers of gravity, in his book "On the equilibrium of planes, or the centers of gravity of planes" (see [67] p. 189), is also intimately related to convexity theory. The work on convex curves that was started by Archimedes was pursued in the 17th century by several eminent mathematicians including Descartes, Fermat and Huygens, who contributed to convexity theory while working in various different areas. One also has to mention Johannes Kepler (1571–1630) who worked extensively on the relations between astronomy, convex polyhedra and music (relations that also date back to the Pythagoreans). The major writing by Kepler on this subject is his *Harmonices Mundi* (see [83]). After the 17th century, research on convexity went on uninterrupted, and we recall the works of Euler, Cauchy and Legendre on the rigidity of convex polyhedra in 3-space that we already mentioned in the introduction of this book.

The first modern definition of a convex set is due to H. Minkowski. Minkowski's interest in convex sets was motivated by number theory, and, more precisely, by a subject that is usually called the "geometry of numbers" (see [110]). To give an idea of the relation between convexity and number theory, we mention Minkowski's famous theorem on convex bodies (1896), which asserts that for any closed convex body A in $\mathbb{R}^n$ of volume V that is symmetric with respect to the origin and for any point lattice in $\mathbb{R}^n$ whose determinant is bounded from above by $2^{-n}V$, there exists a point in the lattice that is distinct from the origin and that belongs to A. Minkowski later on made a systematic study of convex sets (see [111])

There is an extensive literature on convexity in vector spaces, and a good introduction to this subject is the book [139] by F. Valentine.

The space of convex subsets. Let E be a finite-dimensional normed vector space and let $\mathcal{CB}(E)$ be the set of compact convex nonempty subsets of E. We know from Proposition 4.1.11 that $\mathcal{CB}(E)$, equipped with the Hausdorff distance, is a metric space. This is an interesting space to study. There is a natural geodesic segment joining any two points A and B in that space, namely, the segment

$$[A, B] = \{tA + (1-t)B, t \in [0, 1]\}.$$

But there may be more than one geodesic segment joining A and B, and in his paper [78], F. Jongmans analyzed geodesic segments in $\mathcal{CB}(E)$. He proved that given any two points A and B in $\mathcal{CB}(E)$, either there are infinitely many geodesic segments joining them, or there is a unique one (which is the natural geodesic described above). We know that the first case occurs for instance if E is a non-uniquely geodesic normed vector space, with $A = \{x\}$ and $B = \{y\}$, where x and y are any two distinct points in E (we shall see several examples of such spaces in Chapter 7 below). In the same paper, Jongmans asked for a characterization of pairs of points in a finite-dimensional normed vector space that are joined by a unique geodesic segment. R. Schneider gave an answer in [127]. Schneider's result is formulated in terms of the uniqueness of the middle point of the two points A and B, that is, of a point C satisfying $|B - C| = |A - C| = (1/2)|A - B|$ (see Proposition 2.6.2). Schneider proved that A and B have a unique middle point if and only if one of the following holds:

• there exists some $r > 0$ such that $A = B + B(O, r)$ or $B = A + B(O, r)$, where $B(O, r)$ is the closed unit ball in E centered at the origin and of radius r,

or

• A and B are contained in parallel hyperplanes and $A = B + t$, where t is some vector in E that is orthogonal to these hyperplanes.

Finally, let us mention a result on the isometries of $\mathcal{CB}(E)$. In [57], P. M. Gruber and G. Lettl show that a map $f : \mathcal{CB}(E) \to \mathcal{CB}(E)$ is an isometry if and only if there exists a rigid motion i of E and an element D of $\mathcal{CB}(E)$ such that f is the map $C \mapsto i(C) + D$.

Minkowski's construction. Minkowski's construction, which is the subject of Section 5, can be made in the more general setting of a locally convex topological vector space E containing a non-empty bounded open set, that is, an open set V such that for every open neighborhood N of the origin of E, there exists a positive real number α satisfying $S \subset \alpha N$ (cf. [139], Part III). Of course, a finite-dimensional vector space equipped with its natural topology satisfies this property. We have described Minkowski's construction only in the case of finite-dimensional vector spaces to avoid introducing more terminology. This setting of finite-dimensional vector spaces (that is, of "Minkowski spaces") is the setting that Minkowski worked in.

The cross ratio and the Hilbert metric. The Hilbert metric is studied by Busemann in [28] §18 and by Busemann and Kelly in [31]. The most important property of the cross ratio is certainly the fact that it is invariant under projective transformations of $\mathbb{R}^n \cup \{\infty\}$ (considered as the real projective space $\mathbb{P}^n$). In fact, this is essentially the content of Proposition 5.6.4 which says that cross ratio is invariant by perspectivities. (For the general case, one has to deal with the case where one of the points a_1, a_2, b_1, b_2

is the point ∞, and then study compositions of projectivities.) The definition of the Hilbert metric of a bounded open convex body in $\mathbb{R}^n$ can be made without reference to the Euclidean metric of that space, but using only its affine structure. In the case where A is the interior of an ellipsoid, then A, equipped with the Hilbert metric, is a model for hyperbolic space $\mathbb{H}^n$. In other words, A, equipped with its Hilbert metric, is isometric to the models that we described in Example 2.1.3 (v). In the special case where A is the interior of the unit disk, this model of hyperbolic geometry is usually called the *Klein model*. Thus, Hilbert metrics on convex bodies constitute in some sense a non-Riemannian generalization of hyperbolic geometry. Hilbert gave the definition of this metric in a letter he wrote to F. Klein on August 14, 1894, and he published the part of the letter containing this definition as a memoir (cf. [69]). Later on, he included this memoir in the second edition of his major work, the Foundations of Geometry (*Grundlagen der Geometrie*, [70]). We also recall that Number IV of the famous Hilbert problems (see [71]) asks for the characterization of the metrics on subsets of finite-dimensional projective space $\mathbb{P}^n$ for which the geodesic lines are straight lines in this space. Busemann calls such a space a "Desarguesian" space. Busemann made a thorough study of these spaces in [28], Chapter II and in [30], Chapter II. From the definition, it follows that a Desarguesian space is uniquely geodesic. The Hilbert metric is an example of such a metric space.[3]

[3] Of course, the formulation of Hilbert's Problem IV for abstract metric spaces is posterior to Hilbert, but it is clear that Hilbert asked the question for non-Riemannian metrics, since for Riemannian metrics the problem was already solved by Beltrami in 1865. Indeed, Eugenio Beltrami proved in [12] that any Riemannian metric on a connected open subset of projective space whose geodesic lines are straight lines is a metric of constant curvature. In fact, Beltrami only considered the case of the projective plane $\mathbb{P}^2$, but the result for projective space $\mathbb{P}^n$ for any $n \geq 2$ follows easily from that special case; see also [28] §15.

Chapter 6
Convex functions

Introduction

In this chapter, we collect some general properties of real-valued convex functions that will be needed in the rest of this book.

A real-valued function defined on a convex subset C of a real vector space is said to be convex if for every x and y in C and for every t in [0, 1], we have

$$f\big((1-t)x + ty\big) \leq (1-t)f(x) + tf(y).$$

Convexity of functions can also be expressed in terms of affine convexity of subsets: f is convex if and only if its epigraph (that is, the set of points in $C \times \mathbb{R}$ that are situated above the graph of f) is convex.

The definition of a convex function leads very rapidly to non-trivial results. We shall prove the following properties of convex functions of one real variable: existence of left and right derivatives at each point (Proposition 6.2.8), continuity on the interior of the domain (Corollary 6.2.8), existence of a derivative on the complement of a countable set (Proposition 6.2.7) and the remarkable property that says that if a function is locally convex, then it is convex (Theorem 6.2.16). The property of having a derivative is interesting in itself, but the main reason for which we include it here is that we use it in the proof of the local-implies-global convexity property (Theorem 6.2.16). This last property is at the basis of many applications of convex functions in geometry. In particular, it is one of the main ingredients in the proof of the important theorem that says that a complete geodesic locally compact locally convex metric space is globally convex, that is, a space that we call a "Busemann space" (Theorem 9.3.4 below).

The outline of this chapter is as follows:

Section 1 contains the definitions and the main basic properties of convex functions. We establish various conditions under which a function is convex. We study some examples of convex functions that arise in geometry, namely distance functions and projections onto closed convex subsets. These examples, that are defined on Euclidean vector spaces, give a preliminary taste for analogous examples in the more general context of Busemann metric spaces. In fact, most of the results of this section are valid in the context of Busemann spaces, as we shall see in Chapter 8.

Section 2 contains additional basic results for the case of convex functions of one variable that will be useful in the following chapters.

We recall that all the vector spaces considered in these notes are real vector spaces. We recall also that for all $n \geq 1$, $\mathbb{E}^n$ denotes the space $\mathbb{R}^n$ equipped with the Euclidean norm.

6.1 Convex functions

We start with the following classical definition:

Definition 6.1.1 (Convex and strictly convex function). Let E be a vector space and let $C \subset E$ be an affinely convex subset of E. Then a function $f : C \to \mathbb{R}$ is said to be *convex* if for every x and y in C and for every t in $[0, 1]$, we have

$$f\big((1-t)x + ty\big) \leq (1-t)f(x) + tf(y).$$

The function f is said to be *strictly convex* if for every distinct points x and y in C and for every t in $]0, 1[$, we have

$$f\big((1-t)x + ty\big) < (1-t)f(x) + tf(y).$$

Proposition 6.1.2 (Restriction). *Let E be a vector space and let C be an affinely convex subset of E. If $f : C \to \mathbb{R}$ is convex* (*respectively strictly convex*) *and if C' is an affinely convex subset of C, then the restriction of f to C' is convex* (*respectively strictly convex*).

Proof. The proof follows clearly from the definitions. □

Proposition 6.1.2 admits a sort of a converse that is non-trivial, which says that a locally convex function is convex. We shall prove this result in the special case where E is one-dimensional (Theorem 6.2.16 below). We note that there is an analogous result that is valid in any dimension.

Now let us give a few examples of convex functions:

Examples 6.1.3 (Convex functions).

(i) *Linear, affine and sublinear maps.* Let E be a vector space. Any linear map $f : E \to \mathbb{R}$ is convex. More generally, any affine map is convex. We recall that a map $f : E \to \mathbb{R}$ is affine if and only if there exists x_0 in $\mathbb{R}$ and a linear map $g : E \to \mathbb{R}$ such that $f(x) = g(x) + x_0$ for all x in E. We also note that an affine map is never strictly convex. Another generalization of the class of linear maps is that of sublinear maps. We recall that a map $f : \mathbb{R} \to \mathbb{R}$ is said to be *sublinear* if it satisfies $f(x + y) \leq f(x) + f(y)$ and $f(\lambda x) = \lambda f(x)$ for every x and y in E and for every λ in $\mathbb{R}$. It follows easily from the definitions that sublinear maps are convex.

(ii) *Norm.* Let E be a normed vector space. Then the map $x \mapsto \|x\|$ defined on E is convex. Indeed, for all x and y in E and for all t in $[0, 1]$, we have, by the triangle inequality,

$$\|(1-t)x + ty\| \leq \|(1-t)x\| + \|ty\| = (1-t)\|x\| + t\|y\|.$$

(iii) *Positive linear combinations of (strictly) convex functions.* Let C be an affinely convex subset of a vector space. For every $i = 1, \dots, n$, let $f_i : C \to \mathbb{R}$ be a convex (respectively strictly convex) function and let λ_i be a positive real number. Then the function $\sum_{i=1}^{n} \lambda_i f_i$ is convex (respectively strictly convex). We note that the result is false if we allow some of the λ_i's to be negative.

(iv) *Distance to a point.* If E is a normed vector space and if x_0 is a point in E, then the map $x \mapsto \|x - x_0\|$, defined on E, is convex. This result is more general than example (ii) above, and it is a particular case of Proposition 6.1.4 below.

(v) *Distance between two lines in $\mathbb{E}^n$.* Let $t \mapsto x(t)$ and $t \mapsto y(t)$ be two affinely parametrized lines in n-dimensional Euclidean space $\mathbb{E}^n$. Then the map $t \mapsto \|x(t) - y(t)\|$ is convex. Indeed, we can write $x(t) = ta + b$ and $y(t) = tc + d$ for some a, b, c and d in E, which gives $\|x(t) - y(t)\| = \|t(a - c) + b - d\|$. The result then follows from Proposition 6.1.18 below.

Example (v) is at the basis of the general theory of nonpositive curvature in the sense of Busemann. Indeed, Busemann spaces are defined as metric spaces in which the distance function between two arbitrary geodesic paths is convex. These spaces will be our main subject of study in later chapters of this book.

The next proposition provides us with a family of examples of convex functions defined on a normed vector space. It also establishes a relation between convex functions and convex subsets of that vector space.

We recall that if X is a metric space and if Y is a subset of X, we have a map $d_Y : X \to \mathbb{R}$, the "distance map from x to Y", defined by setting, for every x in X,

$$d_Y(x) = \inf_{y \in Y} \|x - y\|.$$

Proposition 6.1.4 (Distance to a convex set). *Let E be a normed vector space and let C be a nonempty closed convex subset of E. Then the map $d_C : E \to \mathbb{R}$ is convex.*

Proof. Let ϵ be a positive real number. Given x_0 and x_1 in E, let x_0' and x_1' be two points in C satisfying $\|x_0 - x_0'\| \leq d_C(x_0) + \epsilon$ and $\|x_1 - x_1'\| \leq d_C(x_1) + \epsilon$. For every t in [0, 1], let $x_t = (1 - t)x_0 + tx_1$ and let $x_t' = (1 - t)x_0' + tx_1'$. By convexity of C, the point x_t' is in C and we have

$$\begin{aligned} \|x_t - x_t'\| &= \|(1 - t)x_0 + tx_1 - (1 - t)x_0' - tx_1'\| \\ &= \|(1 - t)(x_0 - x_0') + t(x_1 - x_1')\| \\ &\leq (1 - t)\|x_0 - x_0'\| + t\|x_1 - x_1'\| \\ &\leq (1 - t)d_C(x_0) + td_C(x_1) + \epsilon. \end{aligned}$$

Letting ϵ tend to 0, we obtain $\|x_t - x_t'\| \leq (1 - t)d_C(x_0) + td_C(x_1)$. This implies that $d_C(x_t) \leq (1 - t)d_C(x_0) + td_C(x_1)$, which proves that the map d_C is convex. □

Definition 6.1.5 (Projection onto a subset). Let X be a metric space, let Y be a subspace of X and let x be a point in X. We say that a point x_0 of Y is a *projection* of x on Y if

$$\|x - x_0\| = \inf_{y \in Y} \|x - y\|.$$

It is easy to construct an example of a space X that contains a nonempty subspace Y and a point x that has no projection on Y. If Y is nonempty and closed, such a projection always exists but is not necessarily unique, even in the case where X is a normed vector space and where Y is a closed convex subset of X. In the case where the normed vector space is Euclidean space $\mathbb{E}^n$, the projection is unique. This will follow from Corollary 6.1.7 below. We shall see in Chapter 7 necessary and sufficient conditions on normed vector spaces so that the projection on any closed convex subset of such a space is unique. These conditions will depend on the geometry of the unit ball in that space.

Lemma 6.1.6 (Projection onto a convex subset of $\mathbb{E}^n$). *Let C be a nonempty convex subset of $\mathbb{E}^n$, let x be a point in $\mathbb{E}^n \setminus C$ and let x_0 be a projection of x on C. Then for every point x_1 in C that is distinct from x_0, the cosine of the angle made by the vectors $x - x_0$ and $x_1 - x_0$ is ≤ 0.*

Proof. By convexity of C, we have $(1-t)x_0 + tx_1 \in C$ for all t in $[0, 1]$. We also have

$$\|x - x_0\|^2 \leq \left\|x - \big((1-t)x_0 + tx_1\big)\right\|^2 = \|(x - x_0) - t(x_1 - x_0)\|^2.$$

Letting α be the angle made by the vectors $x - x_0$ and $x_1 - x_0$ in $\mathbb{E}^n$, we have, for all t in $[0, 1]$,

$$\|x - x_0\|^2 \leq \|x - x_0\|^2 - 2t\|x - x_0\|.\|x_1 - x_0\| \cos\alpha + t^2\|x_0 - x_1\|^2.$$

For all $t \in]0, 1]$, we therefore obtain

$$\cos\alpha \leq t\frac{\|x_1 - x_0\|}{\|x - x_0\|}.$$

Letting t tend to 0, we obtain $\cos\alpha \leq 0$, which proves Lemma 6.1.6. □

We note the following corollary, which also follows from more general results that are valid in strictly convex vector spaces (Proposition 7.1.4). We also note that there are results of this type that are valid in general Busemann metric spaces (cf. Proposition 8.4.7 below)

Corollary 6.1.7 (Uniqueness of the projection). *Let C be a nonempty closed convex subset of $\mathbb{E}^n$. For all x in $\mathbb{E}^n$, there exists a unique projection x_0 of x on C.*

Proof. If x belongs to C, then $x_0 = x$ is the unique projection of x on C. Therefore, we assume that $x \notin C$. Then the existence of x_0 follows from the fact that the set C is closed. Indeed, we start with any sequence $(y_n)_{n\geq 0}$ of points in C satisfying $\|x - y_n\| \to d_C(x)$ as $n \to \infty$. The sequence (y_n) is bounded, and up to replacing it by a subsequence, we can assume that it converges. Let x_0 be its limit. We have $\|x - y_n\| \to \|x - x_0\|$, which gives $d_C(x) = \|x - x_0\|$. To prove uniqueness, suppose that there exists a point x_1 in C that is distinct from x_0 and that satisfies $d_C(x) = \|x - x_1\|$. The triangle x, x_0, x_1 is non-degenerate, and in this triangle, we would have, by Lemma 6.1.6, two angles whose cosines are ≤ 0, which is a contradiction. This shows the uniqueness of the projection, and this proves Corollary 6.1.7. □

Proposition 6.1.8 (Projection is 1-Lipschitz). *Let C be a closed convex subset of $\mathbb{E}^n$. Then the map that assigns to each point in $\mathbb{E}^n$ its projection on C is 1-Lipschitz.*

Proof. Let x and y be two points in $\mathbb{E}^n$ and let x_0 and y_0 be their projections on C. We assume that the four points x, y, x_0 and y_0 are distinct (otherwise, the proof is simpler). Suppose that the quadrilateral xx_0y_0y is not contained in a 2-dimensional affine subspace of $\mathbb{E}^n$ and let A be the 3-dimensional affine subspace of $\mathbb{E}^n$ that contains this quadrilateral. We apply to A a rotation whose axis is x_0y_0, that transforms the point y into a point y_1 contained in the 2-dimensional affine subspace containing the points x, x_0 and y_0, and such that in that plane, y is on the same side than x with respect to the axis x_0y_0. We have $\|x - y\| \leq \|x - y_1\|$. (In fact, y_1 is the projection of x on the closed disk in A that contains y in its frontier, whose center is on the axis x_0y_0 and that is orthogonal to this axis.) Furthermore, the cosine of the angle made by the two vectors $y_1 - y_0$ and $x_0 - y_0$ is equal to the cosine of the angle made by the two vectors $y - y_0$ and $x_0 - y_0$. Now in the planar quadrilateral $xx_0y_0y_1$, the angles at the vertices x_0 and y_0 are obtuse, and therefore we have $\|x_0 - y_0\| \leq \|x - y_1\|$, which implies $\|x_0 - y_0\| \leq \|x - y\|$. This proves Proposition 6.1.8. □

Since we are talking about uniqueness of projections, we mention the following general result that concerns strictly convex functions:

Proposition 6.1.9 (Uniqueness of the minimum). *Let E be a normed vector space and let $C \subset E$ be an affinely convex subset. Let $f : C \to \mathbb{R}$ be a strictly convex function. Then there exists at most one point in C where f attains a minimum.*

Proof. Suppose that there exist two distinct points x and y in C such that $f(x) = f(y) = m = \inf_{x\in C} f(x)$. Then, by convexity of C, we have $(x + y)/2 \in C$, and, by strict convexity of f, we obtain

$$m \leq f\left(\frac{x+y}{2}\right) < \frac{1}{2}\big(f(x) + f(y)\big) = m,$$

which is a contradiction. This proves Proposition 6.1.9. □

Now, we give a few more properties of convex functions. Definition 6.1.10 below and the proposition that follows it say that theoretically, the study of convex functions is equivalent to the study of convex sets.

Definition 6.1.10 (Epigraph). Let E be a vector space, let C be an affinely convex subset of E and let $f : C \to \mathbb{R}$ be a function. The *epigraph* of f is the set

$$\mathrm{Ep}(f) = \{(x, t) \in C \times \mathbb{R},\ f(x) \le t\}$$

(see Figure 6.1).

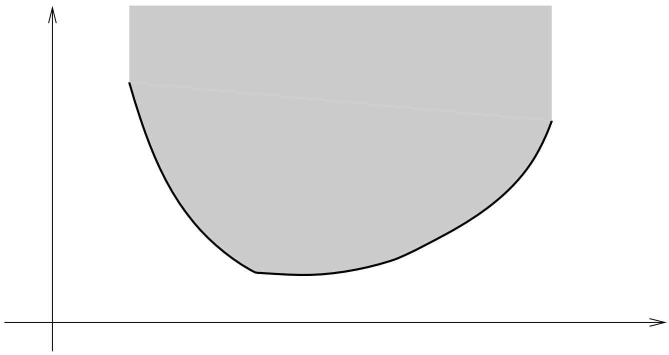


Figure 6.1. Epigraph.

We have the following relation between convexity of f and convexity of its epigraph.

Proposition 6.1.11 (Epigraph and convexity). *Let E be a vector space and let C be an affinely convex subset of E. A function $f : C \to \mathbb{R}$ is convex if and only if its epigraph is an affinely convex subset of $E \times \mathbb{R}$.*

Proof. Suppose that f is convex, let (x, u) and (y, v) be in $\mathrm{Ep}(f)$ and let t be in $[0, 1]$. By convexity of f, we have

$$f\big((1-t)x + ty\big) \le (1-t)f(x) + tf(y) \le (1-t)u + tv.$$

Hence,

$$\big((1-t)x + ty, (1-t)u + tv\big) \in \mathrm{Ep}(f),$$

that is,

$$(1-t)(x, u) + t(y, v) \in \mathrm{Ep}(f).$$

This shows that $\mathrm{Ep}(f)$ is convex.

Conversely, suppose that $\mathrm{Ep}(f)$ is convex, let x and y be two points in C and let t be in $[0, 1]$. Since $\big(x, f(x)\big)$ and $\big(y, f(y)\big)$ are in $\mathrm{Ep}(f)$, we have

$$\big((1-t)x + ty, (1-t)f(x) + tf(y)\big) \in \mathrm{Ep}(f),$$

which means

$$f\big((1-t)x + ty\big) \leq (1-t)f(x) + tf(y).$$

Thus, f is convex. □

Remark (Strict epigraph). Let us note that we can also define the *strict epigraph* of a function $f : C \to \mathbb{R}$ by replacing, in Definition 6.1.10, the inequality $f(x) \leq t$ by the strict inequality $f(x) < t$. Likewise, Proposition 6.1.11 remains valid if we replace in its statement the word "epigraph" by "strict epigraph". (The proof is the same, up to replacing some of the inequalities by strict inequalities.)

We can use the notion of epigraph to obtain examples of convex functions that are not necessarily continuous:

Example 6.1.12 (Non-continuity). Consider the non-continuous function $f : [0, 1] \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 0 & \text{if } 0 \leq x < 1, \\ 1 & \text{if } x = 1. \end{cases}$$

It is clear that the epigraph of f is a convex subset of $[0, 1] \times \mathbb{R}$. Thus, the function f is convex.

The proofs of the following two propositions give another example of how properties of convex subsets imply properties of convex functions.

Proposition 6.1.13. *Let E be a vector space, let C be a convex subset of E and let $f : C \to \mathbb{R}$ be a function. The following two properties are equivalent:*

(i) *f is convex;*

(ii) *for every integer $n \geq 1$, for every $t_1, \ldots, t_n$ in $[0, 1]$ satisfying $t_1 + \cdots + t_n = 1$ and for every $x_1, \ldots x_n$ in C, we have $f(t_1x_1 + \cdots + t_nx_n) \leq t_1 f(x_1) + \cdots + t_n f(x_n)$.*

Proof. We have (ii) ⇒ (i) by taking $n = 2$. Let us show that (i) ⇒ (ii). Let $E(f)$ be the epigraph of f, let $t_1, \ldots, t_n$ be n points in $[0, 1]$ satisfying $t_1 + \cdots + t_n = 1$ and let $x_1, \ldots x_n$ be n points in C. For every $i = 1, \ldots, n$, we have $(x_i, f(x_i)) \in \mathrm{Ep}(f)$. If f is convex, then $\mathrm{Ep}(f)$ is convex, and therefore, we have, by Proposition 5.1.4, $t_1(x_1, f(x_1)) + \cdots + t_n(x_n, f(x_n)) \in \mathrm{Ep}(f)$, which gives $f(t_1x_1 + \cdots + t_nx_n) \leq t_1 f(x_1) + \cdots + t_n f(x_n)$. □

Proposition 6.1.14 (Upper limit of convex functions). *Let E be a vector space, let C be an affinely convex subset of E and let $\{f_i\}_{i\in I} : C \to \mathbb{R}$ be a family of convex functions. Suppose that we have, for all x in C, $sup_{i\in I} f_i < \infty$. Then $f = \sup_{i\in I} f_i$ (which is called the upper limit of the family $\{f_i\}$) is a convex function on C.*

Proof. The epigraph of $f = \sup_{i\in I} f_i$ is the intersection of the epigraphs of the f_i's. Therefore, it is a convex subset of E. □

Proposition 6.1.11, together with its proof, give a "visual" characterization of convexity (see Figure 6.2), that we state as a proposition:

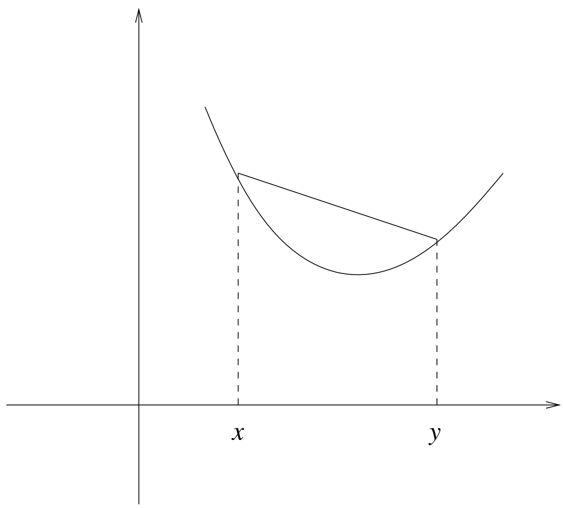


Figure 6.2. A convex function.

Proposition 6.1.15. *Let E be a vector space, let C be an affinely convex subset of E and let $f : C \to \mathbb{R}$ be a map. Then f is convex if and only if for all distinct points x and y in C, the affine segment joining $(x, f(x))$ and $(y, f(y))$ is above (in the large sense) the graph of the restriction of the map f to the segment $[x, y]$.* □

Proposition 6.1.16 (Pointwise limit of convex functions). *Let E be a vector space and let C be an affinely convex subset of E. If $f_i : C \to \mathbb{R}$ is a sequence of convex functions that converges pointwise to a function $f : I \to \mathbb{R}$, then f is convex.*

Proof. The proof follows easily from the definition of convexity. □

Proposition 6.1.17 (Precomposition of a convex function with an affine function). *Let E be a vector space and let C be an affinely convex subset of E. If $f: C \to \mathbb{R}$ is a convex function, if D is an affinely convex subset of a vector space E' and if $\psi: D \to C$ is affine, then $f \circ \psi: D \to \mathbb{R}$ is convex.*

Proof. Associated to the affine map ψ, there is a linear map $L: E' \to E$ and a vector v in E such that $\psi(x) = L(x) + v$ for every x in E'. Therefore, we can write, for every x and y in D and for every t in [0, 1],

$$\begin{aligned} f \circ \psi\big((1-t)x + ty)\big) &= f(L((1-t)x + ty) + v) \\ &= f\big((1-t)L(x) + tL(y) + tv\big) \\ &= f\big((1-t)(L(x) + v) + t(L(y) + v)) \\ &= f\big((1-t)(\psi(x)) + t\psi(y)\big) \\ &\leq (1-t)f \circ \psi(x) + tf \circ \psi(y). \end{aligned}$$

This shows that $f \circ \psi: D \to \mathbb{R}$ is convex. □

The following result, which is a consequence of Proposition 6.1.17, will be useful for us later on.

Proposition 6.1.18. *Let x and y be two points in a normed vector space E. Then the map $t \mapsto \|x + ty\|$, defined on $\mathbb{R}$, is convex.*

Proof. For any x and y in E, the map $t \mapsto x + ty$ defined on $\mathbb{R}$ is affine. We saw in Example 6.1.3 (ii) that the function $x \mapsto \|x\|$ is convex on E. Therefore, by Proposition 6.1.17, the function $t \mapsto \|x + ty\|$ is convex. □

It is not true that if $\psi: C \to \mathbb{R}$ is a convex function and if $f: \mathbb{R} \to \mathbb{R}$ is an arbitrary affine map, then the map $\psi \circ f$ is always convex. (For instance, take $f: \mathbb{R} \to \mathbb{R}$ be the map $x \mapsto x^2$ and $\psi: \mathbb{R} \to \mathbb{R}$ the map $x \mapsto -x$.) However, the map $\psi \circ f$ is convex in the case where the affine map ψ is increasing. This is a special case of Proposition 6.1.19 below. Likewise, he composition of two convex functions is not necessarily a convex function. For instance, consider the composition of the two functions $x \mapsto x^2 - 3$ and $x \mapsto x^2 - 4$ defined on $\mathbb{R}$. The result is a function that has more than one local extremum, and therefore it is not convex. Nevertheless, we have the following result that will be useful for us later on:

Proposition 6.1.19 (Composition of convex functions). *Let E be a vector space, let C be an affinely convex subset of E and let $f: C \to \mathbb{R}$ be a convex function. If $g: f(C) \to \mathbb{R}$ is an increasing convex (respectively strictly convex) function, then $g \circ f: C \to \mathbb{R}$ is convex (respectively strictly convex).*

Proof. First we consider the case where g is convex (in the large sense). For x and y in C and for t in $[0, 1]$, we have

$$f\big((1-t)x + ty\big) \leq (1-t)f(x) + tf(y),$$

which implies

$$g\big(f\big((1-t)x+ty\big)\big) \leq g\big((1-t)f(x)+tf(y)\big) \leq (1-t)g\circ f(x) + tg\circ f(y),$$

which shows that $g \circ f$ is convex. In the case where g is strictly convex, then, for all $t \notin \{0, 1\}$, the last inequality is strict, and this shows that $g \circ f$ is strictly convex. □

The following consequence of Proposition 6.1.19 is most useful, and we shall use it in Chapter 8.

Corollary 6.1.20. *If $f : I \to \mathbb{R}$ is a positive convex function, then, for any $\alpha > 1$, the function f^α is strictly convex.* □

Proposition 6.1.21. *Let E be a vector space, let C be an affinely convex subset of E and let $f : C \to \mathbb{R}$ be a continuous function satisfying*

$$f\left(\frac{x+y}{2}\right) \leq \frac{1}{2}\big(f(x)+f(y)\big) \qquad (6.1.21.1)$$

for all x and y in I. Then f is convex.

Proof. Let x and y be points in C and let us first consider a real number t in $[0, 1]$ that is of the form $t = p/2^q$, where p and q are natural numbers. Dividing the interval $[0, 1]$ into 2^q intervals of equal lengths and applying q times (6.1.21.1), we obtain $f\big((1-t)x + ty)\big) \leq (1-t)f(x) + tf(y)$. Now let t be an arbitrary number in $[0, 1]$. Then there exists a sequence $(t_i)_{i\geq 0}$ in $[0, 1]$ that converges to t, with $t_i = p_i/2^{q_i}$, with p_i and q_i natural numbers. Therefore, for all $i \geq 0$, we have $f\big((1-t_i)x + t_i y)\big) \leq (1-t_i)f(x) + t_i f(y)$. Since f is continuous, we obtain, by taking the limit as $i \to \infty$, $f\big((1-t)x + ty)\big) \leq (1-t)f(x) + tf(y)$. This proves Proposition 6.1.21. □

We shall see another relation between convex functions and convex sets. First we introduce the following definition:

Definition 6.1.22 (Sublevel set). Given a function $f : C \to \mathbb{R}$ and given a real number α, the *sublevel set of f of hight α* is defined as

$$f_\alpha = \{x \in C,\ f(x) \leq \alpha\}.$$

The *strict sublevel set of f of hight α* is defined as

$$f_\alpha^* = \{x \in C,\ f(x) < \alpha\}.$$

We have the following

Proposition 6.1.23. *Let E be a vector space, let C be an affinely convex subset of E and let $f : C \to \mathbb{R}$ be a convex function. Then, for every real number α, the sublevel set f_α and the strict sublevel set f_α^* are convex subsets of E.*

Proof. The proof is straightforward: for every α in $\mathbb{R}$ and for every x and y in f_α, we have $f(x) \leq \alpha$ and $f(y) \leq \alpha$. Therefore, for t in $[0, 1]$, we have, by the convexity of f,

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y) \leq (1-t)\alpha + t\alpha = \alpha.$$

This shows that f_α is convex. The proof for f_α^* is the same, up to replacing some large inequalities by strict inequalities. □

We note however that it is not always true that if all the sublevel sets f_α of a map f are convex then f is convex. For instance, the map $f : [0, \infty[\to \mathbb{R}$ defined by $f(x) = -x^2$ is not convex, although its sublevel sets are either the empty set or an interval of $\mathbb{R}$, and therefore they are convex. Of course, this example is somehow superficial since the map $x \mapsto -x^2$ is convex up to the minus sign. In fact, the property for a map of having all of its sublevel sets convex sets is interesting and the following definition can also be done in the general setting of metric spaces:

Definition 6.1.24 (Sublevel-convex function). Let E be a vector space and let C be an affinely convex subset of E. Then, a map $f : C \to \mathbb{R}$ is said to be *sublevel-convex* if for every real number α, the sublevel set f_α is convex.

We end this section by the following relation between general convex functions and convex functions of one real variable. This will make the link with Section 2 below.

Proposition 6.1.25. *Let E be a vector space and let C be an affinely convex subset of E. Then a function $f : C \to \mathbb{R}$ is convex (respectively strictly convex) if and only if for all x and y in C, the function $f_{x,y} : [0, 1] \to \mathbb{R}$ defined by $f_{x,y}(t) = f((1-t)x + ty)$ is convex (respectively strictly convex).*

Proof. The proof is clear from the definitions. □

Proposition 6.1.25 says in some sense that in the study of convex functions, it suffices to consider convex functions of one real variable. One can compare this proposition with the general definition of a convex function on a metric space X: a function $f : X \to \mathbb{R}$ is convex if its restriction to geodesic paths in X in convex (see Chapter 8). In any case, convex functions of one real variable play a prominent role in convexity theory, and this is one of the reasons for which the rest of this chapter is devoted to them.

6.2 Convex functions of one variable

In this section, we collect a few properties of convex functions of one real variable (that is, convex functions whose domain is an interval of $\mathbb{R}$) that will be useful for us in the sequel. As we already said, most of the important ideas in convexity theory can be expressed in this one-dimensional setting, and for our needs in the rest of this book, it suffices to consider the case of functions of one variable.

In the rest of this chapter, I denotes an arbitrary (finite or infinite) interval of $\mathbb{R}$.

Proposition 6.2.1. *Let $f\colon I \to \mathbb{R}$ be a map. The following four properties are equivalent:*

(i) *the map f is convex* (*respectively strictly convex*)*;*

(ii) *for all x, y and z in I satisfying $x < y < z$, we have*

$$\frac{f(y)-f(x)}{y-x} \le \frac{f(z)-f(x)}{z-x}$$

(*and the inequality is strict in the case where f is strictly convex*)*;*

(iii) *for all x, y and z in I satisfying $x < y < z$, we have*

$$\frac{f(y)-f(x)}{y-x} \le \frac{f(z)-f(y)}{z-y}$$

(*and the inequality is strict in the case where f is strictly convex*)*;*

(iv) *For all x, y and z in I satisfying $x < y < z$, we have*

$$\frac{f(z)-f(x)}{z-x} \le \frac{f(z)-f(y)}{z-y}$$

(*and the inequality is strict in the case where f is strictly convex*).

Proof. Let us show that (i) ⇒ (ii). For all $x < y < z$, there exists t in $]0, 1[$ such that $y = (1-t)x + tz$. (More precisely, we take $t = (y-x)/(z-x)$).

If f is convex, then $f(y) \le (1-t)f(x) + tf(z)$, or, equivalently,

$$f(y) - f(x) \le t\big(f(z) - f(x)\big),$$

that is,

$$f(y) - f(x) \le \frac{(y-x)}{(z-x)}\big(f(z) - f(x)\big),$$

which implies

$$\frac{f(y)-f(x)}{y-x} \le \frac{f(z)-f(x)}{z-x}.$$

In the case where f is strictly convex, we have strict inequalities. Thus, we have (i) ⇒ (ii). The proofs of (i) ⇒ (iii) and (i) ⇒ (iv) can be done in the same way.

Conversely, each of Properties (ii), (iii) or (iv) implies (i). For instance, to see that (ii) $\Rightarrow$ (i), we can follow backwards the arguments of the proof of (i) $\Rightarrow$ (ii) and we obtain $f\big((1-t)x + ty\big) \leq (1-t)f(x) + tf(y)$ for all $0 < t < 1$. For $t = 0$ or 1, the desired convexity inequality is obviously satisfied. This completes the proof of Proposition 6.2.1. □

Proposition 6.2.2. *Let $f : I \to \mathbb{R}$ be a convex function and let x, y and z be points in I satisfying $x < y < z$. Then the point $(z, f(z))$ in $\mathbb{R}^2$ is above (in the large sense) the line joining $(x, f(x))$ and $(y, f(y))$.*

Proof. This follows from Property (iv) of Proposition 6.2.1. □

Proposition 6.2.2 is useful for the proof of the following result that will serve us in Chapters 9 and 10.

Proposition 6.2.3. *Let $f : \mathbb{R} \to \mathbb{R}$ be a convex function. If f is bounded from above, then f is constant.*

Proof. Let $f : \mathbb{R} \to \mathbb{R}$ be an non-constant convex function. Then, we can find two points x and y in $\mathbb{R}$ satisfying $x < y$ and $f(x) \neq f(y)$. Let us first suppose that $f(x) < f(y)$. In this case, the slope of the straight line in the plane that passes through the points $(x, f(x))$ and $(y, f(y))$ is positive, and, by Proposition 6.2.2, the graph of the restriction of f to the interval $[y, \infty[$ is above the graph of this straight line. Thus, we have $f(x) \to \infty$ as $x \to \infty$. In the case where $f(x) > f(y)$, an analogous reasoning gives $f(x) \to \infty$ as $x \to -\infty$. Thus, if f is bounded above, it is necessarily constant. □

Proposition 6.2.4. *Let $f : I \to \mathbb{R}$ be a convex function, let x and y be two points in I satisfying $x < y$ and suppose that there exists t in $]0, 1[$ satisfying*

$$f\big((1-t)x + ty\big) = (1-t)f(x) + tf(y).$$

Then, this equality is satisfied for all t in $[0, 1]$. (In other words, the restriction of f to the interval $[x, y]$ is affine.)

Proof. Let $z = (1-t)x + ty$. Then we have $x < z < y$, and the hypothesis implies that the point $(z, f(z))$ is on the line D joining the points $(x, f(x))$ and $(y, f(y))$ in the plane. Consider a point M on the graph of f. By 6.2.2, the point M is above the line D. By convexity of the map f, the point M is below the line D. Therefore, the point M is on D. This completes the proof of the proposition. □

Corollary 6.2.5. *Let $f : I \to \mathbb{R}$ be a convex function. Then f is strictly convex if and only if there exist no points x and y in I satisfying $x < y$ such that the restriction of f to the interval $[x, y]$ is an affine map* □

Proposition 6.2.6. *Let $f : I \to \mathbb{R}$ be a convex (respectively strictly convex) function. Then, for all a in I, the function $f_a : I \setminus \{a\} \to \mathbb{R}$ defined by*

$$f_a(x) = \frac{f(x) - f(a)}{x - a}$$

is increasing (respectively strictly increasing).

Proof. The proof follows from Property (iv) of Proposition 6.2.1. □

Proposition 6.2.7 (Existence of left and right derivatives). *Let $f : I \to \mathbb{R}$ be a convex function. Then, at each point x in $\overset{\circ}{I}$ (the interior of I), f admits a left derivative $f_l'(x)$ and a right derivative $f_r'(x)$ that are both finite. Furthermore, for all x and y in $\overset{\circ}{I}$ satisfying $x < y$, we have*

$$f_l'(x) \leq f_r'(x) \leq f_l'(y) \leq f_r'(y).$$

Proof. Let us fix a point a in $\overset{\circ}{I}$. For all x in $\overset{\circ}{I}$ with $x \neq a$, the value $f_a(x) = (f(y) - f(a))/(y - a)$ is the slope of the affine line joining the points $(x, f(x))$ and $(a, f(a))$ on the graph of f. If f is convex, then, by Condition (iii) of Proposition 6.2.1, we have $f_a(x) < f_a(y)$ for all x and y in I satisfying $x < y < a$. Thus, $f_a(x)$ is an increasing function of x, for x in the interval $I \cap]-\infty, a[$. Furthermore, this function is bounded from above since, for all x and z in I satisfying $x < a < z$, we have, by Condition (iv) of Proposition 6.2.1, $f_a(x) \leq f_a(z)$. We conclude that $f_a(x)$ has a finite limit as x tends to a (with $x < a$). By definition, this limit is the left derivative $f_l'(a)$ of f at the point a. In the same manner, we can show that f has a right derivative $f_r'(a)$ at every point a in $\overset{\circ}{I}$. Furthermore, we have shown that for all x, y and a in $\overset{\circ}{I}$ satisfying $x < a < y$, we have

$$\frac{f(x) - f(a)}{x - a} \leq f_l'(a) \leq f_r'(a) \leq \frac{f(y) - f(a)}{y - a},$$

which implies

$$f_l'(x) \leq f_r'(x) \leq \frac{f(y) - f(x)}{y - x} \leq f_l'(y) \leq f_r'(y).$$

This completes the proof of Proposition 6.2.7. □

Corollary 6.2.8 (Continuity of convex functions). *Let $f : I \to \mathbb{R}$ be a convex function. Then f is continuous on $\overset{\circ}{I}$.*

Proof. From Proposition 6.2.7, at each point in $\overset{\circ}{I}$, the map f has a finite right and a finite left derivative. Therefore, this map has a right and a left limit at each point, and it is therefore continuous. □

Proposition 6.2.9 (Existence of a derivative in the complement of a countable set). *Let $f : I \to \mathbb{R}$ be a convex function. Then the set of points in I where f does not have a derivative is at most countable.*

Proof. Suppose that there exist two points x and y in I, with $x < y$, at which f does not have a derivative. By Proposition 6.2.7, the left and right derivatives at each of these points exist, but they are distinct. Thus, we have, again by Proposition 6.2.7, $f_l'(x) < f_r'(x) \leq f_l'(y) < f_r'(y)$. Therefore, the two open intervals $]f_l'(x), f_l'(x)[$ and $]f_l'(y), f_l'(y)[$ are disjoint. Thus, to each point x where f does not have a derivative, we can associate a rational number in the open nonempty interval $]f_l'(x) < f_r'(x)[$, and we obtain in this manner an injection from the set of points at which f does not have a derivative into the set of rational numbers. This shows that the set of points in I where f does not have a derivative is at most countable. □

Busemann introduced in [28] (p. 109) a new terminology to describe a certain property of a real-valued function of a real variable, which is more general than convexity and which he uses in his theory of geodesics in metric spaces. We recall it here:

Definition 6.2.10 (Peakless function). A continuous function $f : I \to \mathbb{R}$ is said to be *peakless* if there exists a sub-interval I_0 (which may be empty) of I on which f is constant and such that the complement of I_0 in I is the union of two sub-intervals I_l and I_r, with f strictly decreasing on I_l and strictly increasing on I_r. In the case where I_0 is nonempty, the interval I_l is either empty or situated to the left of I_0, and the interval I_r is either empty or situated to the right of I_0. The function f is said to be *strictly peakless* if I_0 is either a single point or empty.

Thus, the graph of a peakless function f has the aspect described in Figure 6.3.

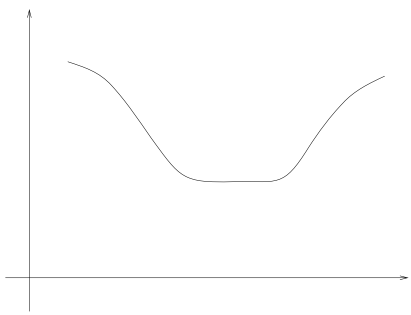


Figure 6.3. A peakless function.

It is easy to see that the three intervals I_0, I_l and I_r that appear in Definition 6.2.10 are uniquely determined by the property described there.

Peakless functions share some of the properties of convex functions. For instance, if a peakless function has a maximum, then it is constant. From Proposition 6.2.7, we deduce the following

Proposition 6.2.11 (Convex functions are peakless). *Let $f: I \to \mathbb{R}$ be a continuous convex function. Then f is peakless.*

Proof. This follows from the fact that the function $x \mapsto f_r'(x)$ is increasing on I (Proposition 6.2.7). The interval I_0 is the set of points where $f_r' = 0$, the interval I_l is the set of points where $f_r' < 0$ and the interval I_r is the set of points where $f_r' > 0$. □

In particular, a convex functions (as a peakless functions) has the property that if it is constant on some interval with nonempty interior, then every point of that interval is a global minimum point.

The following corollary will also be useful for us:

Corollary 6.2.12. *Let $[x, y]$ be a compact interval of $\mathbb{R}$ and let $f: [x, y] \to \mathbb{R}$ be a nonnegative convex function satisfying $f(x) = 0$. Then f is increasing.*

Proof. This follows directly from the fact that f is continuous on $\overset{\circ}{I}$ and from the description of the variation of f given in Proposition 6.2.11. □

Proposition 6.2.13. *Let $[x_0, x_1]$ be a compact interval of $\mathbb{R}$ and let $f: [x_0, x_1] \to \mathbb{R}$ be a convex function. Then, for every $x \in [x_0, x_1]$, we have*

$$f(x) \leq \max\{f(x_0), f(x_1)\}.$$

Furthermore, if for some x in $]x_0, x_1[$ we have

$$f(x) = \max\{f(x_0), f(x_1)\},$$

then f is constant on $[x_0, x_1]$.

Proof. Let $M = \max\{f(x_0), f(x_1)\}$. Every point in $[x_0, x_1]$ can be written as $x_t = (1 - t)x_0 + x_1$, with $t \in [0, 1]$. By convexity of f, we have

$$f(x_t) \leq (1 - t)f(x_0) + tf(x_1) \leq (1 - t)M + tM = M.$$

The second part of the statement follows from the fact that f is continuous on $]x_0, x_1[$ and from the fact that a continuous convex function is peakless. □

Proposition 6.2.14 (Characterization of convex functions). *Let $f: I \to \mathbb{R}$ be a map. Then the following conditions are equivalent:*

(i) *f is convex (respectively strictly convex);*

(ii) *f is continuous on $\overset{\circ}{I}$, there is a countable set in I on the complement of which f has a derivative, and the derivative of f, as a map defined on the set of points where this derivative exists, is increasing* (*respectively strictly increasing*).

Proof. We have (i) ⇒ (ii) by Propositions 6.2.8, 6.2.9 and 6.2.7. Let us prove (ii) ⇒ (i). Suppose that Condition (ii) holds and let $\mathcal{D}$ be the set of points where f has a derivative. We reason by contradiction. If f were not convex, then by Proposition 6.2.1, we can find points x, y and z in I such that $x < y < z$ and such that

$$\frac{f(z) - f(y)}{z - y} < \frac{f(y) - f(x)}{y - x}.$$

By the mean value theorem, we obtain

$$\frac{f(y) - f(x)}{y - x} \leq \sup_{u \in \mathcal{D} \cap]x,y[} f'(u)$$

and

$$\inf_{u \in \mathcal{D} \cap]y,z[} f'(u) \leq \frac{f(y) - f(x)}{y - x}.$$

Thus, we have

$$\inf_{u \in \mathcal{D} \cap]y,z[} f'(u) < \sup_{u \in \mathcal{D} \cap]x,y[} f'(u),$$

which contradicts the fact that f' is increasing on $\mathcal{D}$. □

Corollary 6.2.15. *Let $f : I \to \mathbb{R}$ be a $\mathcal{C}^2$-map. Then f is convex* (*respectively strictly convex*) *if and only if $f'' \geq 0$* (*respectively $f'' > 0$*). □

Finally, we have the following

Theorem 6.2.16 (Locally convex implies convex). *Let $f : I \to \mathbb{R}$ be a locally convex function.* (*In other words, suppose that for all x in I, there is an open interval $I_x \subset I$ containing x such that the restriction of f to I_x is a convex function.*) *Then f is convex.*

Proof. This follows directly from the characterization of convex functions that is given in Proposition 6.2.14. □

Notes on Chapter 6

Convex functions. Convex functions have remarkable properties, and they have applications in several branches of mathematics, including functional analysis, control theory, economics, optimisation and, of course, geometry.

The idea of studying a convex functions is already contained in [59], which is one of the early papers by Hadamard. In this paper, Hadamard, studies Riemann's zeta function and he introduces a class of real-valued functions of one real variable that have a derivative except at some finite set, and whose derivative is an increasing function, a property which of course is very close to the definition of a convex function.

Several of the results in this chapter will be proved in the more general context of convex functions defined on geodesically convex subspaces of Busemann spaces. For instance, Corollary 6.1.7 is a special case of Proposition 8.4.8 of Chapter 8.

Inequality (ii) of Proposition 6.1.13 is generally referred to as Jensen's inequality. J. L. Jensen is considered generally as being the first mathematician who studied convex functions in a systematic way. His foundational papers on the subject are [76] and [77].[1] In these papers, Jensen defines a convex function as a function f on a compact interval $[a, b] \subset \mathbb{R}$ satisfying inequality

$$(*) \qquad f\left(\frac{x+y}{2}\right) \leq \frac{1}{2}\big(f(x) + f(y)\big)$$

for all x and y in $[a, b]$ (without a continuity hypothesis). He proved that if such a function is bounded above, then it is continuous on the open interval $]a, b[$, and it possesses at every point of $]a, b[$ a left and a right derivative. It is now a classical result that if a function f defined on a convex subset C of a vector space satisfies $(*)$ and if f is bounded on some open subset of C, then f is convex.

Another result of Jensen says that inequality $(*)$ holds for all x and y in $[a, b]$ if and only if the following inequality

$$(**) \qquad f(t_1x_1 + \cdots + t_nx_n) \leq t_1 f(x_1) + \cdots + t_n f(x_n)$$

(that is, inequality (ii) of Proposition 6.1.13) holds for every integer $n \geq 1$, for all $t_1, \ldots, t_n \in [0, 1]$ satisfying $t_1 + \cdots + t_n = 1$ and for all $x_1, \ldots x_n \in I$. Inequality $(*)$ is inequality 6.1.21.1 of Proposition 6.1.21, in which we suppose that f is continuous. In [77], Jensen attributes Inequality $(**)$ to Otto Hölder (1859–1937) who, in his paper [72], establishes it for a $\mathcal{C}^2$-map f whose second derivative is nonnegative.

Otto Stolz (1842–1905) proved in [129] that if $f : [a, b] \to \mathbb{R}$ is continuous and if it satisfies $(*)$ for all x and y in $[a, b]$, then f has a right and a left derivative at each point of the open interval $[a, b]$.

Peakless functions. Peakless functions (Definition 6.2.10) were defined by Busemann in [28] p. 109. According to that definition, a continuous function $f : I \to X$ is peakless if for every x_1, x_2 and x_3 in I satisfying $x_1 < x_2 < x_3$, we have

$$f(x_2) \leq \max\{f(x_1), f(x_3)\},$$

[1] We mention that the Danish mathematician Johann Ludwig Jensen was as an engineer doing research in mathematics at his spare time. Furthermore, he was completely self-taught for what concerns higher mathematics.

and equality implies that $f(x_1) = f(x_3)$. (This is equivalent to Definition 6.2.10 above.) Thus, Proposition 6.2.13 is equivalent to the fact that the function f in that statement is peakless on the interval $]x_0, y_0[$.

Busemann introduced peakless functions to define a notion of a "metric space in which the distance function is peakless". Such spaces generalize his "negatively curved spaces" that is, the spaces we call Busemann spaces, which are characterized by the fact that the distance function between two geodesics is convex. We refer the reader to the notes on Chapter 9 below for a more precise statement.

Chapter 7

Strictly convex normed vector spaces

Introduction

In this chapter, we study strictly convex normed vector spaces as examples of geodesic spaces.

A normed vector space E is strictly convex if for all distinct x_0 and x_1 in E satisfying $\|x_0\| = \|x_1\| = 1$ and for all t in $]0, 1[$, we have

$$\|(1-t)x_0 + tx_1\| < 1.$$

Strictly convex normed vector spaces are uniquely geodesic. We note right away that strictly convex normed vector spaces are also examples of Busemann spaces, that we shall study in Chapter 8.

The outline of this chapter is as follows.

In Section 1, we give the definition and various characterizations of strictly convex normed vector spaces. Some of these characterizations can be formulated as follows (we shall give precise statements below):

- the spheres are strictly convex;
- the distance function to any point is strictly convex;
- any projection map onto a closed affinely convex subset is strictly convex.

In Section 2, we show that a normed vector space is strictly convex if and only if it is uniquely geodesic, and we give several other metric characterizations of strict convexity of normed vector spaces in terms of inequalities involving distances between a finite number of points in such a space.

In Section 3, we study the strict convexity of vector spaces equipped with norms arising from inner products and with the familiar ℓ^p norms.

7.1 Strictly convex normed vector spaces

We saw in Chapter 5 that normed vector spaces are examples of geodesic metric spaces. Not all these vector spaces are uniquely geodesic. The property of being uniquely geodesic is related to the geometry of the unit ball of the space. We already saw that the (closed or open) unit ball in a normed vector space is affinely convex (Proposition 5.3.14). We shall prove below that a normed vector space is uniquely geodesic if and only if it satisfies the following property, which in some sense expresses the fact that the closed unit ball is strictly convex:

Definition 7.1.1 (Strictly convex normed vector space). A normed vector space E is said to be *strictly convex* if for all distinct x_0 and x_1 in E satisfying $\|x_0\| = \|x_1\| = 1$ and for all t in $]0, 1[$, we have $\|(1-t)x_0 + tx_1\| < 1$.

The next proposition is merely a reformulation of this definition, but it expresses more clearly strict convexity of a normed vector space in terms of an affine property of its unit sphere (or of any sphere) in that space.

Proposition 7.1.2 (Strict convexity of spheres). *A normed vector space E is strictly convex if and only if its unit sphere* (*or equivalently, any sphere of positive radius*) *in E does not contain any affine segment that is not reduced to a point.*

Proof. The proof in the case of the unit sphere follows trivially from Definition 7.1.1. The case of an arbitrary sphere of positive radius follows by applying to that sphere a translation followed by a homothety of E, that send this sphere onto the unit sphere, and using the fact that these transformations preserve the affine properties. □

Thus, by looking at the pictures of the unit spheres (Figure 7.1), we can see that the vector space $\mathbb{R}^2$ equipped with the ℓ^2 norm is strictly convex, whereas the same vector space equipped with the ℓ^1 or with the ℓ^∞ norm is not strictly convex.

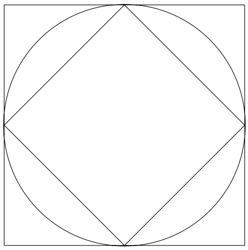


Figure 7.1. The outer square is the unit sphere for the ℓ^∞ norm, the inner square is the unit sphere for the ℓ^1 norm, and the circle in between is the unit sphere for the ℓ^2 norm.

Proposition 7.1.3 (Strict convexity of the distance function). *A normed vector space E is strictly convex if and only if for every point x in E and for every distinct points x_0 and x_1 in E satisfying $\|x - x_0\| = \|x - x_1\|$, the distance from x to the point $x_t = (1-t)x_0 + tx_1$ is a strictly convex function of t in* $[0, 1]$. *In other words, the map*

$$t \mapsto \|x - (1-t)x_0 - tx_1\|,$$

defined on $[0, 1]$, *is strictly convex.*

Proof. We reason by contradiction. Suppose that there exist three points x, x_0 and x_1 in E, with x_0 and x_1 distinct and satisfying $\|x - x_0\| = \|x - x_1\|$, and such that the distance function $t \mapsto \|x - x_t\|$ is not strictly convex. Performing, if necessary, a translation followed by a homothety of E, we can assume that x is the origin of E of and that x_0 and x_1 are at distance 1 from the origin. Thus, from Definition 7.1.1, E is not strictly convex.

Conversely, suppose that for any x in E and for any distinct points x_0 and x_1 in E satisfying $\|x - x_0\| = \|x - x_1\|$, the map $t \mapsto \|x - x_t\|$ is strictly convex. Taking x to be the origin of E and taking x_0 and x_1 to be any points at distance 1 from this origin, we see from Definition 7.1.1 that E is strictly convex. This completes the proof of Proposition 7.1.3. □

Proposition 7.1.4 (Projection onto a closed convex subset). *A normed vector space E is strictly convex if and only if for every x in E and for every affinely convex closed subset C of E, there is a unique projection of x on C.*

Proof. Suppose that E is strictly convex, let C be an affinely convex closed subset of E, let x be a point in E and suppose that x has two distinct projections x_0 and x_1 on C. Since C is convex, the affine segment $[x_0, x_1]$ is contained in C. For every t in $[0, 1]$, let $x_t = (1 - t)x_0 + tx_1$. We have $\|x - x_0\| = \|x - x_1\|$. Therefore, by Proposition 7.1.3, the map $t \mapsto \|x - x_t\|$ is strictly convex, which gives, for all t in $]0, 1[$,

$$\|x - x_t\| < \|x - x_0\| = \|x - x_1\|.$$

This contradicts the fact that x_0 and x_1 are projections of x on C. Thus, the projection of x on C is unique.

Conversely, suppose that for every x in E and for every affinely convex closed subset C of E, there is a unique projection of x on C. Suppose that the unit sphere of E contains an affine segment. This segment is a closed convex subset of E, and we call it C. The distance from the origin O to any point in C is equal to 1, and each point in C is therefore a projection of O on C. By the uniqueness of the projection, the segment C must be reduced to a point. Proposition 7.1.2 then implies that E is strictly convex. This completes the proof of Proposition 7.1.4. □

7.2 Uniquely geodesic spaces

The next proposition gives several characterizations of strict convexity for normed vector spaces.

Proposition 7.2.1. *Let E be a normed vector space and let O denote its origin. The following nine properties are equivalent:*

(i) *as a metric space, E is uniquely geodesic;*

(ii) *for every x in E, there exists a unique geodesic segment joining O to x;*

(iii) *if x and y are two points in E satisfying $\|x\| + \|y\| = \|x + y\|$, then either $y = O$ or there exists a nonnegative real number λ such that $x = \lambda y$;*

(iv) *if x, y and z are three points in E satisfying $\|x - y\| + \|y - z\| = \|x - z\|$, then there exists t in $[0, 1]$ such that $y = (1 - t)x + tz$;*

(v) *the normed vector space E is strictly convex;*

(vi) *for all distinct x and y in E satisfying $\|x\| = \|y\| = 1$, we have $\|x + y\| < 2$;*

(vii) *for every x and y in E that are not collinear, we have $\|x + y\| < \|x\| + \|y\|$;*

(viii) *for every x and y in E with $x \neq y$ and for every p in $]1, \infty[$, we have*

$$\left\| \frac{x + y}{2} \right\|^p < \frac{1}{2}(\|x\|^p + \|y\|^p);$$

(ix) *for every x and y in E with $x \neq y$ and for every p in $]1, \infty[$, the map on E defined by $x \mapsto \|x\|^p$ is strictly convex.*

Proof. Implication (i) ⇒ (ii) is trivial and (v) ⇒ (vi) follows from Definition 7.1.1 by taking $t = 1/2$.

We begin by proving (ii) ⇒ (iii). Let x and y be two points in E. If $x = O$, then we have $x = \lambda y$ with $\lambda = 0$ and the conclusion of (iii) is satisfied. Therefore we can assume that $x \neq O$. Consider the path $\gamma : [0, \|x\| + \|y\|] \to E$ defined by

$$\gamma(t) = \begin{cases} \frac{t}{\|x\|} x & \text{if } 0 \leq t \leq \|x\| , \\ \left(1 - \frac{t - \|x\|}{\|y\|}\right) x + \frac{t - \|x\|}{\|y\|}(x + y) & \text{if } \|x\| \leq t \leq \|x\| + \|y\|. \end{cases}$$

(Of course, the second case is useful in this definition if and only if $\|y\| \neq O$.)

The path γ is obtained by concatenating two affine geodesics, namely the affine geodesic joining O to x and the affine geodesic joining x to $x + y$. These two affine geodesics are parametrized by arclength. Therefore, by Proposition 1.2.7, γ is also parametrized by arclength. Thus, γ joins O to $x + y$, it is parametrized by arclength and it satisfies $L(\gamma) = \|x\| + \|y\|$. If $\|x\| + \|y\| = \|x + y\|$, then $L(\gamma) = \|x + y\|$ and Proposition 2.2.7 implies that γ is geodesic. Now if Condition (ii) is satisfied, then any point on the image of γ is on the image of the affine geodesic joining O to $x + y$. In particular, there exists a real number λ' in $[0, 1]$ such that $x = \lambda'(x + y)$. Hence, $(1 - \lambda')x = \lambda' y$, and if $y \neq O$ we have $\lambda' \neq 1$, which implies $x = \lambda y$ with $\lambda = \lambda'/(1 - \lambda')$. This proves (ii) ⇒ (iii).

Now let us prove that (iii) ⇒ (iv). Let x and y be two points in E satisfying $\|x - y\| + \|y - z\| = \|x - z\|$. If (iii) is satisfied, then either $y = z$ or there exists $\lambda \geq 0$ such that $x - y = \lambda(y - z)$, or, equivalently,

$$y = x/(1 + \lambda) + \lambda z/(1 + \lambda).$$

Therefore Property (iv) is satisfied with $t = \lambda/(1+\lambda)$.

Let us show that (iv) $\Rightarrow$ (i). Let x and z be two arbitrary points in X. The affine segment $\{(1-t)x + tz, t \in [0,1]\}$ is a geodesic segment joining x to z. Suppose that y is a point on a geodesic segment joining x and z. Then by Proposition 2.2.7, we have $\|x-y\| + \|y-z\| = \|x-z\|$. Therefore Property (iv) implies that y is on the affine segment $\{(1-t)x + tz,\ t \in [0,1]\}$. Thus, this segment is the unique geodesic segment joining x and z.

Let us show that (iii) $\Rightarrow$ (v). Let x and y be two distinct vectors of norm one and let t be in $]0,1[$. We claim that there does not exist any nonnegative real number λ satisfying $tx = (1-t)\lambda y$. Indeed, this equality would imply that x and y are situated on the same line through the origin, and since both vectors have norm 1, this would imply $x = \pm y$. Then $x = y$ using again the equality $tx = (1-t)\lambda y$. Thus, if Condition (iii) is satisfied, then

$$\|tx + (1-t)y\| < t\|x\| + (1-t)\|y\| = t + (1-t) = 1,$$

which implies that Condition (v) is satisfied.

Let us show that (v) $\Rightarrow$ (iii). Suppose that Condition (v) is satisfied and let x and y be two non-zero vectors in E. Let us set $u = x/\|x\|$, $v = y/\|y|$ and $t = \|x\|/(\|x\| + \|y\|)$. Then we have $\|u\| = \|v\| = 1$, $0 < t < 1$ and $tu + (1-t)v = (x+y)/(\|x\| + \|y\|)$. If no nonnegative real number λ exists such that $x = \lambda y$, then $u \neq v$, and therefore we obtain, from Condition (v), $\|tu + (1-t)v\| < 1$. Hence, $\|x+y\| < \|x\| + \|y\|$. Therefore Condition (iii) is satisfied.

We prove that (vi) $\Rightarrow$ (vii) by contradiction. Suppose that x and y are not collinear and that

$$\|x+y\| = \|x\| + \|y\|, \tag{7.2.1.1}$$

and let us show that (vi) cannot hold. Dividing the two members of (7.2.1.1) by $\max\{\|x\|, \|y\|\}$, we can suppose, up to interchanging the names of x and y, that $\|y\| = 1$ and $\|x\| \leq 1$. The map $t \mapsto \|tx + y\|$, defined on $[0, \infty[$, is convex (Proposition 6.1.18), and it is bounded above by the linear map $t \mapsto t\|x\| + \|y\|$, defined on the same interval. Furthermore, the two maps coincide for $t = 0$ and for $t = 1$. By Proposition 6.2.4, these two maps coincide on $[1, \infty[$. Thus, we have $\left\|tx + y\right\| = t\|x\| + \|y\|$ for all $t \geq 1$. Taking $t = 1/\|x\|$, we obtain

$$\left\| \frac{x}{\|x\|} + y \right\| = \frac{\|x\|}{\|x\|} + |y\| = 2,$$

which shows that (vi) is not satisfied. Thus, we have (vi) $\Rightarrow$ (vii).

Now we prove that (vii) $\Rightarrow$ (viii). If (vii) is satisfied, then if x and y are not collinear, we have, $\|x+y\| < \|x\| + \|y\|$ which implies, for all p in $]1, \infty[$,

$$\left\| \frac{x+y}{2} \right\|^p < \left(\frac{\|x\| + \|y\|}{2} \right)^p .$$

By convexity of the map $t \mapsto t^p$, we have

$$\left(\frac{\|x\| + \|y\|}{2}\right)^p < \frac{1}{2}(\|x\|^p + \|y\|^p).$$

Thus, we obtain

$$\left\|\frac{x+y}{2}\right\|^p < \frac{1}{2}(\|x\|^p + \|y\|^p),$$

which is the desired inequality.

Now suppose that x and y are collinear and let us write $y = tx$ for some t in $\mathbb{R}$. Then we have

$$\left\|\frac{x+y}{2}\right\|^p = \left\|\frac{x+tx}{2}\right\|^p = \left|\frac{1+t}{2}\right|^p \|x\|$$

and

$$\frac{1}{2}(\|x\|^p + \|y\|^p) = \frac{1}{2}(\|x\|^p + |t|^p\|x\|^p) = \frac{1}{2}(1 + |t|^p)\|x\|^p.$$

By elementary calculus, one can see that

$$\left(\frac{1+t}{2}\right)^p < \frac{1}{2}(1 + |t|^p)$$

for all $t \neq 1$ and for all $p > 1$.

Thus, we obtain again, in the case where x and y are collinear,

$$\left\|\frac{x+y}{2}\right\|^p < \frac{1}{2}(\|x\|^p + \|y\|^p).$$

This proves (vii) $\Rightarrow$ (viii).

Now we prove that (viii) $\Rightarrow$ (ix), that is, we prove that if (viii) is satisfied, then, for any two distinct vectors x and y in X, for any p in $]1, \infty[$ and for any t in $]0, 1[$, we have

$$\|(1-t)x + ty\|^p < (1-t)\|x\|^p + t\|y\|^p.$$

First, suppose that $0 \leq t \leq 1/2$. We claim that the map $x \mapsto \|x\|^p$ is convex for all p in $]1, \infty[$. Indeed, since the map $x \mapsto \|x\|$ is convex on E and the map $t \mapsto t^p$, is convex and increasing on $[0, \infty[$, for all p in $]1, \infty[$, then, by Proposition 6.1.19, $x \mapsto \|x\|^p$ is convex. Thus, we have

$$\begin{aligned}\|(1-t)x + ty\|^p &= \left\|\frac{2(1-2t)x + 2t(x+y)}{2}\right\|^p \\ &\leq (1-2t)\|x\|^p + 2t\left\|\frac{x+y}{2}\right\|^p.\end{aligned}$$

Now if Condition (viii) is satisfied, then

$$(1-2t)\|x\|^p + 2t\left\|\frac{x+y}{2}\right\|^p < (1-2t)\|x\|^p + t(\|x\|^p + \|y\|^p)$$
$$= (1-)\|x\|^p + t\|y\|^p.$$

Thus, we obtain

$$\|(1-t)x + ty\|^p < (1-2t)\|x\|^p + t\|y\|^p,$$

which is the desired inequality. The case where $1/2 \leq t \leq 1$ is handled in the same way. This shows that (viii) ⇒ (ix).

Finally, let us prove that (ix) ⇒ (v). If (ix) is satisfied, then for any two distinct points x and y in X, for any p in $]1, \infty[$ and for any t in $]0, 1[$, we have

$$\|(1-t)x + ty\|^p < (1-t)\|x\|^p + t\|y\|^p.$$

In particular, for $\|x\| = \|y\| = 1$, we obtain $\|(1-t)x + ty\|^p < 1$, that is, $\|(1-t)x + ty\| < 1$. Thus, (v) is satisfied. This completes the proof of Proposition 7.2.1. □

We end this section with another characterization of strict convexity of a normed vector space in terms of the geometry of its unit ball. Before stating it, we need to recall the following classical notion in convexity theory. It is important for us because it makes sense in an arbitrary geodesic metric space.

Definition 7.2.2 (Extreme point). Let E be a normed vector space and let X be an affinely convex subset of E. Then a point x in X is said to be an *extreme point* of X if the set $X \setminus \{x\}$ is convex.

Equivalently, a point x in X is an extreme point of X if for any x_0 and x_1 in X satisfying $x = (1-t)x_0 + tx_1$ with t in $]0, 1[$, we have $x_0 = x_1 = x$.

We already know that the closed unit ball in a normed vector space is convex. The following characterization of strictly convex normed vector spaces says that in such a space, the closed unit ball is strictly convex.

Proposition 7.2.3. *A normed vector space E is strictly convex if and only if every point on the unit sphere of E is an extreme point of the closed unit ball.*

Proof. Suppose that E is strictly convex and let B be the closed unit ball of E. Let x_t be a point on the unit sphere of E such that there exist two points x_0 and x_1 in B and a real number t in $]0, 1[$ satisfying $x_t = (1-t)x_0 + tx_1$. By the triangle inequality, we have $\|x_t\| \leq (1-t)\|x_0\| + t\|x_1\|$, and if any one of the two quantities $\|x_0\|$ or $\|x_1\|$ is < 1, then $\|x_t\| < 1$. This contradicts the fact that x_t is on the unit sphere. Thus, we have $\|x_0\| = \|x_1\| = 1$.

Now we claim that $\|x_0 + x_1\| = 2$. Indeed, if this were not the case, then we would have $\|x_0 + x_1\| < 2$, which implies

$$\begin{aligned}1 &= \|x_t\| = \|(1-t)x_t + tx_t\| \\ &= (1-t)\big((1-t)x_0 + tx_1\big) + t\big((1-t)x_0 + tx_1\big)\| \\ &= \|(1-t)^2x_0 + t(1-t)(x_0+x_1) + t^2x_1\| \\ &\leq (1-t)^2\|x_0\| + t(1-t)\|x_0+x_1\| + t^2\|x_1\|^2 \\ &= (1-t)^2 + t(1-t) + t^2 \\ &< (1-t)^2 + 2t(1-t) + t^2 = 1,\end{aligned}$$

which is a contradiction. Thus, we have $\|x_0 + x_1\| = 2$. Since E is strictly convex, this implies, by Property (vi) of Proposition 7.2.1, that $x_0 = x_1$ and therefore that $x_t = x_0 = x_1$. Thus, x_t is an extreme point of B.

To prove the converse, suppose that every point on the unit sphere of E is an extreme point of B and let us show that Property (vi) of Proposition 7.2.1 is satisfied. Let x and y be two distinct points in B satisfying $\|x\| = \|y\| = 1$. If $\|(x+y)/2\| = 1$, then $(x+y)/2$ would be a point on the unit sphere that is not an extreme point of B. Thus, we have $\|(x+y)/2\| < 1$, and Property (vi) is satisfied. □

7.3 Inner products and ℓ^p norms

In this section, we give a few examples of strictly convex normed vector spaces among the classical spaces. We start by spaces whose norms are defined by inner products.

We recall that an *inner product* on a real vector space E is a bilinear form $E \times E \to \mathbb{R}$, that we shall denote by $(x, y) \mapsto (x|y)$ and that satisfies the following two properties:

- the form is symmetric, that is, $(x|y) = (y|x)$ for all x and y in E;
- the form is positive definite, that is, $(x|x) > 0$ for all $x \neq 0$.

We recall also that any inner product satisfies the following inequality, which is usually referred to as *Minkowski's inequality*:

(7.3.1.1) $$\sqrt{(x+y|x+y)} \leq \sqrt{(x|x)} + \sqrt{(y|y)}$$

for all x and y in E, with equality if and only if either $y = 0$ or $y = \lambda x$ for some nonnegative real number λ.

The norm on E associated to this inner product is defined by setting $\|x\| = \sqrt{(x|x)}$ for every x in X.

Proposition 7.3.1. *Let E be a vector space equipped with a norm arising from an inner product. Then E is strictly convex.*

Proof. It suffices to see that Property (vii) of Proposition 7.2.1 is satisfied; this follows from Minkowski's inequality (7.3.1.1) and from its equality case, that we recalled above. □

For every integer $n \geq 1$ and for every p in $[1, \infty] = [1, \infty[\cup\{\infty\}$, we recall that the ℓ^p norm on $\mathbb{R}^n$, denoted by $\| \ \|_p$, is defined by

- $\|x\|_p = (\sum_{i=}^n |x_i|^p)^{1/p}$ for p in $[1, \infty[$;
- $\|x\|_\infty = \sup_{i=1,\dots,n}(|x_1|, \dots, |x_n|)$ for $p = \infty$.

Proposition 7.3.2. *Given any integer $n \geq 2$ and given p in $[1, \infty]$, consider the vector space $\mathbb{R}^n$ equipped with the ℓ^p norm. Then, for $p \in]1, \infty[$, this normed vector space is strictly convex. For $p = 1$ or $p = \infty$, this normed vector space is not strictly convex.*

Proof. For $p = 1$ and $p = \infty$, we saw that the unit sphere of ℓ^p contains a non-trivial affine segment (Figure 7.1). Therefore, by Proposition 7.1.2, this space is not strictly convex. (In fact, we already saw in Example 5.3.9 that such a space is not uniquely geodesic, and therefore the result also follows from Proposition 7.2.1) For p in $]1, \infty[$, we use the criterion for equality in Minkowski's inequality in Proposition 2.6.4, which says that for all x and y in $\mathbb{R}^n$, if x and y are not collinear, and if p is in $]1, \infty[$, then $\|x + y\|_p < \|x\|_p + \|y\|_p$. Proposition 7.2.1 implies that in that case $(\mathbb{R}^n, \ell^p)$ is not strictly convex. □

Notes on Chapter 7

There is an extensive literature on the subject of strict convexity in normed vector spaces. Most of the criteria of Proposition 7.2.1 are contained in the textbooks [16] and [75], and there are other criteria. For instance, a result of M. A. Khamsi says that a Banach space E is strictly convex if and only if for every nonexpansive map f defined on a convex subset C of E, the fixed point set of f is convex (Theorem 4 of [87]).

Chapter 8

Busemann spaces

Introduction

A Busemann space is a geodesic metric space X such that for any two geodesics $\gamma\colon [a,b] \to X$ and $\gamma'\colon [a',b'] \to X$, the map from $[a,b] \times [a',b']$ to $\mathbb{R}$ defined by

$$(t,t') \mapsto |\gamma(t) - \gamma'(t')|$$

is convex.

In this chapter, we study the basic properties of Busemann spaces. The outline is the following:

In Section 1, we give several characterizations of Busemann spaces. Then we give examples that include Euclidean and hyperbolic spaces, $\mathbb{R}$-trees and strictly convex normed vector spaces. We prove that Busemann spaces are uniquely geodesic and contractible.

In Section 2, we study local geodesics in Busemann spaces. We prove that any local geodesic in a Busemann space is a geodesic.

In Section 3, we collect some facts about geodesically convex subsets in a Busemann space and we prove another "local-implies-global" property, which concerns geodesically convex subsets in Busemann spaces.

In Section 4, we study convex functions defined on a geodesic metric space and then we specialize to the case of a Busemann space.

In the notes at the end of this chapter, we discuss convexity properties of Teichmüller space, equipped with its Teichmüller metric and with its Weil–Petersson metric.

Notation. In all this chapter, if a geodesic is denoted by $\gamma_{x,y}\colon [a,b] \to X$, then it is implicitly assumed that $\gamma(a) = x$ and $\gamma(b) = y$.

8.1 Busemann spaces

Definition 8.1.1 (Busemann spaces). A metric space X is said to be a *Busemann space* if X is a geodesic metric space and if for any two affinely reparametrized geodesics $\gamma\colon [a,b] \to X$ and $\gamma'\colon [a',b'] \to X$, the map $D_{\gamma,\gamma'}\colon [a,b] \times [a',b'] \to \mathbb{R}$ defined by

(8.1.1.1) $$D_{\gamma,\gamma'}(t,t') = |\gamma(t) - \gamma'(t')|$$

is convex.

Proposition 8.1.2 (Characterization of Busemann spaces). *Let X be a geodesic metric space. Then properties* (i) *to* (xi) *that follow are equivalent:*

(i) *X is a Busemann space.*

(ii) *Let $[x_0, x_1]$ and $[x_0', x_1']$ be two arbitrary geodesic segments in X. For every t in $[0, 1]$, let x_t be the point on $[x_0, x_1]$ satisfying $|x_0 - x_t| = t|x_0 - x_1|$ and let x_t' be the point on $[x_0', x_1']$ satisfying $|x_0' - x_t'| = t|x_0' - x_1'|$. Then we have*

$$|x_t - x_t'| \leq (1-t)|x_0 - x_0'| + t|x_1 - x_1'|.$$

(iii) *For all affinely reparametrized geodesics $\gamma\colon [a, b] \to X$ and $\gamma'\colon [a', b'] \to X$, the map $d_{\gamma,\gamma'}\colon [0, 1] \to X$ defined by*

$$d_{\gamma,\gamma'}(t) = |\gamma\big((1-t)a + tb\big) - \gamma\big((1-t)a' + tb'\big)| \tag{8.1.2.1}$$

is convex.

(iv) *Let $\gamma\colon [a, b] \to X$ and $\gamma'\colon [a', b'] \to X$ be two arbitrary affinely reparametrized geodesics in X and let $d_{\gamma,\gamma'}\colon [0, 1] \to X$ be the map defined by formula* (8.1.2.1) *Then we have, for all t and t' in* $[0, 1]$,

$$d_{\gamma,\gamma'}\left(\frac{t+t'}{2}\right) \leq \frac{1}{2}\big(d_{\gamma,\gamma'}(t) + d_{\gamma,\gamma'}(t')\big).$$

(v) *Let $[x_0, x_1]$ and $[x_0', x_1']$ be two arbitrary geodesic segments in X and let m and m' be their respective midpoints. Then we have*

$$|m - m'| \leq \frac{1}{2}\big(|x_0 - x_1| + |x_0' - x_1'|\big).$$

(vi) *Let $[x_0, x_1]$ and $[x_0, x_1']$ be two geodesic segments in X having a point x_0 as a common initial point. For all t in $[0, 1]$, let x_t and x_t' be the points situated respectively on $[x_0, x_1]$ and $[x_0, x_1']$ and satisfying $|x_0 - x_t| = t|x_0 - x_1|$ and $|x_0 - x_t'| = t|x_0 - x_1'|$. Then, for all t in $[0, 1]$, we have*

$$|x_t - x_t'| \leq t|x_1 - x_1'|.$$

(vii) *Let $[x_0, x_1]$ and $[x_0, x_1']$ be two geodesic segments in X having a common initial point x_0, and let m and m' be their respective midpoints. Then we have*

$$|m - m'| \leq \frac{1}{2}\,|x_1 - x_1'|.$$

(viii) *For all affinely reparametrized geodesics $\gamma\colon [0, 1] \to X$ and $\gamma'\colon [0, 1] \to X$, we have, for all t in* $[0, 1]$,

$$|\gamma(t) - \gamma'(t)| \leq (1-t)|\gamma(0) - \gamma'(0)| + t|\gamma(1) - \gamma'(1)|.$$

(ix) *For all affinely reparametrized geodesics* $\gamma\colon [0,1] \to X$ *and* $\gamma'\colon [0,1] \to X$ *satisfying* $\gamma(0) = \gamma'(0)$*, we have, for all* t *in* $[0,1]$*,*

$$|\gamma(t) - \gamma'(t)| \leq t|\gamma(1) - \gamma'(1)|.$$

(x) *For all affinely reparametrized geodesics* $\gamma\colon [0,1] \to X$ *and* $\gamma'\colon [0,1] \to X$ *satisfying* $\gamma(0) = \gamma'(0)$*, we have*

$$|\gamma(1/2) - \gamma'(1/2)| \leq (1/2)|\gamma(1) - \gamma'(1)|.$$

(xi) *For all affinely reparametrized geodesics* $\gamma\colon [0,1] \to X$ *and* $\gamma'\colon [0,1] \to X$*, we have*

$$|\gamma(1/2) - \gamma'(1/2)| \leq (1/2)|\gamma(0) - \gamma'(0)| + (1/2)|\gamma(1) - \gamma'(1)|.$$

Proof. We have (ii) $\iff$ (iii) $\iff$ (viii), since convexity of functions is preserved under precomposition by an affine map. The following equivalences also follow clearly from the definitions: (iv) $\iff$ (v) $\iff$ (xi), (vi) $\iff$ (ix) and (vii) $\iff$ (x). We have (iii) $\iff$ (iv) and (vi) $\iff$ (vii) by Proposition 6.1.21 (or by the same proof). Implication (v) $\Rightarrow$ (vii) is also clear. Therefore, it suffices to prove that (i) $\iff$ (ii) and that (vii) $\Rightarrow$ (v).

We start by proving that (i) $\Rightarrow$ (ii). Let $[x_0, x_1]$ and $[x'_0, x'_1]$ be two arbitrary geodesic segments in X and let $\gamma\colon [a,b] \to X$ and $\gamma'\colon [a',b'] \to X$ be two geodesics whose images are respectively $[x_0, x_1]$ and $[y_0, y_1]$. Let $D_{\gamma,\gamma'}\colon [a,b] \times [a',b'] \to \mathbb{R}$ be the map defined in (8.1.1.1). Since X is a Busemann space, we have, for all u and u' in $[a,b] \times [a',b']$ and for all t in $[0,1]$,

$$D_{\gamma,\gamma'}\big((1-t)u + tu'\big) \leq (1-t)D_{\gamma,\gamma'}(u) + tD_{\gamma,\gamma'}(u'). \tag{8.1.2.2}$$

Letting v and w denote respectively the pairs (a, a') and (b, b'), we have

$$\begin{aligned} D_{\gamma,\gamma'}(v) &= |\gamma(a) - \gamma'(a')| = |x - x'|, \\ D_{\gamma,\gamma'}(w) &= |\gamma(b) - \gamma'(b')| = |y - y'| \end{aligned}$$

and

$$\begin{aligned} D_{\gamma,\gamma'}\big((1-t)v + tw\big) &= D_{\gamma,\gamma'}\big((1-t)a + tb, (1-t)a' + tb'\big) \\ &= |\gamma\big((1-t)a + tb\big) - \gamma'\big((1-t)a' + tb'\big)| \\ &= |x_t - x'_t|. \end{aligned}$$

Thus, we obtain, from (8.1.2.2):

$$|x_t - x'_t| \leq (1-t)|x_0 - x'_0| + t|x_1, x'_1|,$$

which proves that (i) $\Rightarrow$ (ii).

Now we prove that (ii) ⇒ (i). Let us consider two geodesics $\gamma\colon [a,b] \to X$ and $\gamma'\colon [a',b'] \to X$ and let us take two arbitrary points in $[a,b] \times [a',b']$, whose coordinates are respectively (v, v') and (w, w'). We set $x = \gamma(v)$, $y = \gamma(w)$, $[x,y] = \gamma([v,w])$, $x' = \gamma'(v')$, $y' = \gamma'(w')$ and $[x',y'] = \gamma'([v',w'])$.

Then we have

$$D_{\gamma,\gamma'}(v,v') = |\gamma(v) - \gamma'(v')| = |x - x'| \tag{8.1.2.3}$$

and

$$D_{\gamma,\gamma'}(w,w') = |\gamma(w) - \gamma'(w')| = |y - y'|. \tag{8.1.2.4}$$

For all t in [0, 1], let x_t be the point on $[x,y]$ satisfying $|x_t - x| = t|x - y|$ and let x'_t be the point on $[x',y']$ satisfying $|x'_t - x'| = t|x' - y'|$. Then,

$$D_{\gamma,\gamma'}\big((1-t)(v,v') + t(w,w')\big) = |x_t - x'_t|. \tag{8.1.2.5}$$

By property (ii), we have

$$|x_t - x'_t| \le (1-t)|x - x'| + t|y - y'|. \tag{8.1.2.6}$$

Inserting (8.1.2.3), (8.1.2.4) and (8.1.2.5) in (8.1.2.6), we obtain

$$D_{\gamma,\gamma'}\big((1-t)(v,v') + t(w,w')\big) \le (1-t)D_{\gamma,\gamma'}(v,v') + tD_{\gamma,\gamma'}(w,w'),$$

which shows that (ii) ⇒ (i).

Finally, let us prove that (vii) ⇒ (v). Let $[x_0, x_1]$ and $[x'_0, x'_1]$ be two arbitrary geodesic segments in X. Since X is a geodesic space, there exists a geodesic segment $[x_0, x'_1]$ joining x_0 and x'_1. Let m, m' and m'' be the midpoints of $[x_0, x_1]$, $[x'_0, x'_1]$ and $[x_0, x'_1]$ respectively. Applying property (vii) to the segments $[x_0, x_1]$ and $[x_0, x'_1]$, we obtain

$$|m - m''| \le \frac{1}{2}\big(|x_1 - x'_1|\big).$$

Applying the same property to the segments $[x'_1, x_0]$ and $[x'_1, x'_0]$, we find

$$|m' - m''| \le \frac{1}{2}\big(|x_0 - x'_0|\big).$$

By the triangle inequality, we then obtain

$$|m - m'| \le |m - m''| + |m'' - m'| \le \frac{1}{2}\big(|x_1 - x'_1| + |x_0 - x'_0|\big),$$

which shows that (vii) ⇒ (v). This completes the proof of Proposition 8.1.2. □

Examples 8.1.3 (Busemann spaces).

(i) *Euclidean space* $\mathbb{E}^n$. If $[x, y]$ and $[x, z]$ are two geodesic segments in $\mathbb{E}^n$ and if m and m' are their respective midpoints, then, by the theorem of Thales, $|m - m'| = (1/2)|y - z|$. Property (vii) of Proposition 8.1.2 is therefore satisfied. (In fact, the equality instead of the large inequality for all geodesic segments $[x, y]$ and $[x, z]$ in $\mathbb{E}^n$ means that the space has zero curvature in the sense of Busemann.)

(ii) *Hyperbolic space* $\mathbb{H}^n$. Let us prove that for all $n \geq 2$, hyperbolic space $\mathbb{H}^n$ is a Busemann space. We prove that criterion (vii) of Proposition 8.1.2 is satisfied, using a well-known property of hyperbolic isometries of $\mathbb{H}^2$. Let $[x, y]$ and $[x, z]$ be two geodesic segments in $\mathbb{H}^n$ with a common initial point x. These two segments are contained in a plane (that is, a two-dimensional totally geodesic subspace) of $\mathbb{H}^n$. Such a plane is isometric to the space $\mathbb{H}^2$. Therefore, it suffices to prove the desired property in $\mathbb{H}^2$. We may assume that the points x, y and z are pairwise distinct (otherwise the result is trivial). Let m and m' be the midpoints of $[x, y]$ and $[x, z]$ respectively and let us call s_m and $s_{m'}$ the central symmetries of $\mathbb{H}^2$ with respect to the points m and m'. Then, the composed map $s_{m'} \circ s_m$ is an isometry of hyperbolic type whose axis is the hyperbolic line containing m and m', and whose minimal displacement is equal to $2|m - m'|$. On the other hand, we have $s_{m'} \circ s_m(y) = z$. Since y is not on the axis of the isometry $s_{m'} \circ s_m$, we have $|y - z| \geq 2|m - m'|$, which is the required inequality. (In fact, the last inequality is strict, and the strict inequality means, in the language of Busemann, that this space has negative curvature.)

(iii) *The Hilbert metric.* Let A be a bounded open convex subset of $\mathbb{R}^n$. In Section 6 of Chapter 5, we recalled the construction of the Hilbert metric of A. We saw that A, equipped with this metric, is a geodesic metric space for which the Euclidean segments are geodesic. We also saw that this metric space is uniquely geodesic if and only if the boundary of A does not contain two Euclidean segments that span a 2-dimensional affine subspace of $\mathbb{R}^n$. In his book [28], Busemann asked the following question: "Is a Hilbert geometry with negative curvature hyperbolic?" (this is Problem 34, p. 406 of [28]). P. J. Kelly and E. G. Straus answered this question in [81] and [82], where they proved that the convex set A equipped with its Hilbert metric is a Busemann space if and only if A is an ellipsoid. It was already known that an ellipsoid equipped with its Hilbert metric is (up to a constant multiplicative factor) a model of hyperbolic space $\mathbb{H}^n$. The open ball

$$\left\{(x_1, \dots, x_n) : \sum_{i=0}^{n} x_i^2 < 1\right\}$$

of $\mathbb{R}^n$, equipped with its Hilbert metric is called the Klein model of hyperbolic geometry.[1]

(iv) $\mathbb{R}$-*trees.* Let T be an $\mathbb{R}$-tree and let us prove that T is a Busemann space. Again, we shall use Criterion (vii) of Proposition 8.1.2.

[1]This result of Kelly and Straus should be put into parallel with the fact that the Minkowski metric associated to a convex body B is Euclidean if and only if B is an ellipsoid (cf. Section 6 of Chapter 5).

Let $[x, y]$ and $[x, z]$ be two geodesic segments in T. From the uniqueness of the topological segment joining two arbitrary points in an $\mathbb{R}$-tree, one can easily see that the intersection of the geodesic segments $[x, y]$ and $[x, z]$ is necessarily a segment (which could consist of a single point). One of the vertices of this segment is x. Let us call s the other vertex. The union $[x, y] \cup [x, z]$, equipped with the metric induced from that of T, is a space that is homeomorphic to a *tripod* equipped with a simplicial metric, that is, a graph formed by three edges having one vertex in common, the other vertices being monovalent (Figure 8.1). Each of these three edges is isometric to a compact interval of $\mathbb{R}$ and the induced metric on the tripod is a length metric. There are also degenerate cases, where one or two of the edges of the graph $[x, y] \cup [x, z]$ are reduced to a point; these cases can be treated as limiting cases of the non-degenerate case.

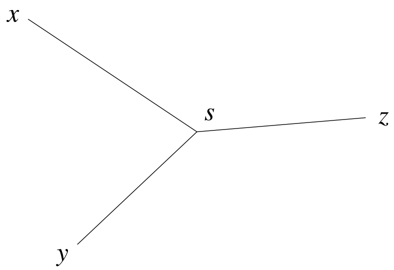


Figure 8.1. A tripod.

Thus, we shall reason in this metric tripod, whose monovalent vertices are naturally called x, y and z (we use for these points the same names as for their images in the space X). Likewise, the trivalent vertex of the tripod is called s. Let m and m' be respectively the midpoints of the segments $[x, y]$ and $[x, z]$ in this tripod. Then we have

$$|y - z| = |x - y| + |x - z| - 2|x - s|,$$
$$|x - y| = 2|x - m|$$

and

$$|x - z| = 2|x - m'|.$$

We break the argument into cases. Up to symmetries, there are only three cases to consider:

Case 1. The point m is on $[s, y]$ and the point m' is on $[s, z]$ (Figure 8.2 (a)). Then we have

$$|m - m'| = |x - m| + |x - m'| - 2|x - s|$$

and

$$|y - z| = |x - y| + |x - z| - 2|x - s|,$$

whence

$$|y - z| - 2|m - m'| = 2|x - s| \geq 0,$$

which implies $|m - m'| \leq (1/2)|y - z|$.

Case 2. The point m is on $[x, s]$ and the point m' is on $[s, z]$ (Figure 8.2 (b)). Then we have:

$$\begin{aligned}|m - m'| &= |m - s| + |s - m'| \\ &= |x - s| - |x - m| + |x - m'| - |x - s|,\end{aligned}$$

whence

$$\begin{aligned}|y - z| - 2|m - m'| &= (|x - y| - |x - s|) + (|x - m| - |x - s|) \\ &= 2|s - y| \geq 0,\end{aligned}$$

which implies $|m - m'| \leq (1/2)|y - z|$.

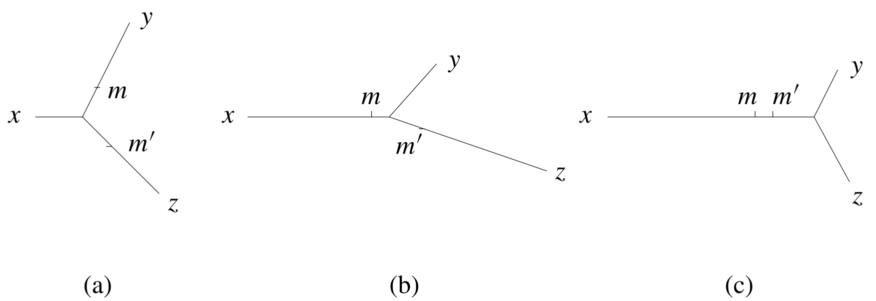


Figure 8.2. The three cases in Example 8.1.3 (iv).

Case 3. The points m and m' are on $[x, s]$, and the four points x, m, m', s are in this order on that segment (Figure 8.2 (c)). In this case, we have

$$|m - m'| = |x - m'| - |x - m|$$

and

$$\begin{aligned}|y - z| - 2|m - m'| &= |x - y| + |x - z| - 2|x - s| - 2|x - m'| + 2|x - m| \\ &= (|x - y| - |x - s|) + (2|x - m| - |x - s|) \\ &\quad + (|x - z| - 2|x - m'|) \\ &= (|x - y| - |x - s|) + 2|x - m| - |x - s| \\ &= (|x - y| - |x - s|) + (|x - y| - |x - s|) \\ &= 2|s - y| \geq 0,\end{aligned}$$

which also implies $|m - m'| \leq (1/2)|y - z|$.

The remaining cases can be deduced by symmetry from the cases considered.

Thus, the convexity inequality is valid in any case, and the $\mathbb{R}$-tree T is a Busemann space.

Proposition 8.1.4. *A Busemann space is uniquely geodesic.*

Proof. Let X be a Busemann space and let x and y be two arbitrary points in X. From the definition, a Busemann space is geodesic, therefore there exists a geodesic segment joining x and y. Suppose that there exist two geodesics joining these two points, $\gamma_{x,y} \colon [a,b] \to X$ and $\gamma'_{x,y} \colon [a',b'] \to X$. Then the map $d_{\gamma,\gamma'} \colon [0,1] \to X$ defined by

$$d_{\gamma,\gamma'}(t) = |\gamma\big((1-t)a + tb\big) - \gamma\big((1-t)a' + tb'\big)|$$

is a nonnegative convex map and it satisfies $d_{\gamma,\gamma'}(0) = d_{\gamma,\gamma'}(1) = 0$. Therefore, by Corollary 6.2.12, this map is the constant zero map, and this shows that the images of $\gamma_{x,y}$ and $\gamma'_{x,y}$ coincide. This proves Proposition 8.1.4. □

Proposition 8.1.5 (Subspaces). *Let X be a Busemann space and let X' be a subset of X which, equipped with the induced metric, is a geodesic space. Then X' is also a Busemann space.*

Proof. The proof is clear from the definitions. □

The next proposition and its corollary will provide us with other examples of Busemann spaces.

Proposition 8.1.6 (Normed vector spaces). *Let E be a normed vector space. The following three properties are equivalent:*

(i) *as a metric space, E is a Busemann space;*

(ii) *as a metric space, E is uniquely geodesic;*

(iii) *E is a strictly convex vector space.*

Proof. Implication (i) $\Rightarrow$ (ii) follows from Proposition 8.1.4. Let us prove (ii) $\Rightarrow$ (i). We use criterion (vii) of Proposition 8.1.2. Let $[x_0, x_1]$ and $[x_0, x_1']$ be two geodesic segments in E and let m and m' be their respective midpoints. If the space E is uniquely geodesic, then $[x_0, x_1]$ and $[x_0, x_1']$ are the images of affinely reparametrized geodesics joining respectively x_0 to x_1 and x_0 to x_1'. Therefore, we have $\|m - m'\| = 1/2(x_0 - x_1)$ and $m' = 1/2(x_0 - x_1')$. Thus, we obtain:

$$\|m - m'\| = \frac{1}{2}|x_1 - x_1'| = \left\| \frac{1}{2}(x_0 + x_1) - \frac{1}{2}(x_0 + x_1') \right\| = \frac{1}{2}\|x_1 - x_1'\|$$

that is,

$$|m - m'| = \frac{1}{2}|x_1 - x_1'|,$$

which proves (ii) $\Rightarrow$ (i). Finally, we have (ii) $\iff$ (iii) since a normed vector space is strictly convex if and only if it is uniquely geodesic (Corollary 7.2.1). $\square$

Example 8.1.7 (Norms on $\mathbb{R}^n$). We already saw that $\mathbb{R}^n$, equipped with the Euclidean norm, is a Busemann space. However, $\mathbb{R}^n$ equipped with the norm ℓ^1 or with the norm ℓ^∞ is not a Busemann space. Indeed, we already saw that this space is not strictly convex. Therefore, by Proposition 8.1.6, this space is not a Busemann space (Proposition 7.3.2).

Recall that a topological space E is said to be *contractible* if there exists a point x_0 in E and a continuous map $H\colon [0,1] \times E \to E$ satisfying $H(0,x) = x_0$ and $H(1,x) = x$ for all x in E. We have the following

Proposition 8.1.8. *A Busemann space is contractible.*

Proof. Let X be a Busemann space and let us fix a basepoint x_0 in X. For every x in X, there exists, by Proposition 8.1.4, a unique geodesic $\gamma_{x_0,x}\colon [0, |x - x_0|] \to X$ (uniqueness is up to a translation of the domain). Then we define the map $H\colon [0,1] \times X \to X$ by setting

$$H(t,x) = \gamma_{x_0,x}(t|x - x_0|)$$

for all x in X and for all t in $[0,1]$. In particular, we have $H(0,x) = x_0$ and $H(1,x) = x$ for all x in X. We prove that H is continuous. If x and x' are two arbitrary points in X, we have, for all t and t' in $[0,1]$,

$$\begin{aligned}|H(t,x) - H(t',x')| &= |\gamma_{x_0,x}(t|x - x_0|) - \gamma_{x_0,x'}(t'|x' - x_0|)| \\ &\leq |\gamma_{x_0,x}(t|x - x_0|) - \gamma_{x_0,x}(t'|x - x_0|)| \\ &\quad + |\gamma_{x_0,x}(t'|x - x_0|) - \gamma_{x_0,x'}(t'|x' - x_0|)|.\end{aligned}$$

Since the path γ_{x,x_0} is geodesic, we have

$$|\gamma_{x_0,x}(t|x - x_0|) - \gamma_{x_0,x'}(t'|x - x_0|)| = |t - t'|.|x - x_0|.$$

Since X is a Busemann space, the map

$$t \mapsto |\gamma_{x_0,x}(t|x - x_0|) - \gamma_{x_0,x'}(t|x' - x_0|)|,$$

defined on $[0,1]$, is convex. This maps takes the value 0 at $t = 0$ and the value $|x - x'|$ at $t = 1$. Therefore, we have, for all t' in $[0,1]$,

$$|\gamma_{x_0,x}(t'|x - x_0|) - \gamma_{x_0,x'}(t'|x' - x_0|)| \leq t'|x - x'|.$$

Finally, we obtain

$$|H(t,x) - H(t',x')| \leq |t - t'| \,.\, |x - x_0| + t'|x - x'|,$$

which implies the continuity of H. This completes the proof of Proposition 8.1.8. □

Thus, by Proposition 8.1.8, there is no Busemann space structure on any space whose fundamental group is non-trivial. By the same proposition, there is no Busemann metric space structure on a sphere S^2.

8.2 Local geodesics in Busemann spaces

Notation. In the same manner as for geodesics, we make the following convention: if we denote by $\gamma_{x,y}\colon [a,b] \to X$ an affinely reparametrized local geodesic, then we assume implicitly that we have $\gamma_{x,y}(a) = x$ and $\gamma_{x,y}(b) = y$.

Proposition 8.2.1. *Let X be a Busemann space and let $\gamma\colon [a,b] \to X$ and $\gamma'\colon [a',b'] \to X$ be two affinely reparametrized local geodesics in X. Then the map $t \mapsto |\gamma\big((1-t)a + tb\big) - \gamma'\big((1-t)a' + tb'\big)\big|$, defined on* $[0,1]$*, is convex.*

Proof. It is clear that this map is locally convex. Therefore, the proof of the proposition follows from Theorem 6.2.16. □

Let us note a few useful corollaries. The first corollary is a result that is stronger than Proposition 8.1.4 above.

Corollary 8.2.2. *Let X be a Busemann space and let x and y be two arbitrary points in X. Then there exists a unique affinely reparametrized local geodesic $\gamma_{x,y}\colon$* $[0,1] \to X$ *joining these points.*

Proof. The existence of a local geodesic follows from the existence of a geodesic, since X is a geodesic space. Now let $\gamma_{x,y}\colon [0,1] \to X$ and $\gamma'_{x,y}\colon [0,1] \to X$ be two local geodesics. The map $t \mapsto |\gamma_{x,y}(t) - \gamma'_{x,y}(t)|$, defined on $[0,1]$, is convex (Proposition 8.2.1), and it takes only nonnegative values. Furthermore, the value of this map is zero for $t = 0$ and $t = 1$. By Corollary 6.2.12, this map is the zero constant map, which implies that $\gamma_{x,y}(t) = \gamma'_{x,y}(t)$ for all t in $[0,1]$. □

We already saw that the sphere S^2 equipped with its canonical length metric is not a Busemann space. Another way to see this fact is to note that there exist at least two local geodesics joining any two distinct points in this space (there are at least two distinct arcs of a great circle joining these two points).

Corollary 8.2.3 (In a Busemann space, local geodesics are geodesics). *In a Busemann space X, every local geodesic is a geodesic.*

Proof. Let γ be a local geodesic and let x and y be its endpoints. Since the space X is geodesic, there exists a geodesic path γ' joining x and y. By Corollary 8.2.2, the images of γ and γ' coincide. Thus, the image of γ is a geodesic segment. It follows easily that γ is a geodesic path. □

The following proposition will also be useful:

Proposition 8.2.4 (Betweenness in Busemann spaces). *Let X be a Busemann space and let x, y, z and t be four points in X that are pairwise distinct. Then we have the following*

(8.2.4.1) *y lies between x and z and z lies between y and t*

$$\Downarrow$$

(8.2.4.2) *y and z both lie between x and t*

Proof. Suppose that (8.2.4.1) holds. If y lies between x and z, then y is on the unique geodesic segment joining x and z. Likewise, if z lies between y and t, then z is on the unique geodesic segment joining y and t. The union of these two geodesic segments is the image of a local geodesic path joining x and t and containing the points y and z. By Corollary 8.2.3, this path is geodesic. This implies that y and z both lie between x and t. □

We note that Proposition 8.2.4 is false without the hypothesis that X is a Busemann space, as one can see by considering the case where X is the two-dimensional sphere S^2 equipped with its canonical length metric, x and t the endpoints of an arc of a great circle in S^2 whose length is strictly larger than π, and y and z two distinct points on that arc satisfying $|x - z| < \pi$ and $|y - t| < \pi$.

Proposition 8.2.5. *Let X be a Busemann space, let $\gamma\colon [0, 1] \to X$ be an affinely reparametrized geodesic and let x be a point in X such that there is no geodesic segment in X containing the three points x, $\gamma(0)$ and $\gamma(1)$. Then, the map $f\colon [0, 1] \to X$ defined by $t \mapsto |x - \gamma(t)|$ is strictly convex.*

Proof. That the map f is convex is part of the definition of a Busemann space. Let us show that it is strictly convex. For all t in $[0, 1]$, let $x_t = (1 - t)\gamma(0) + t\gamma(1)$ and let $y_t = (1 - t)\gamma(0) + tx$ (we are using the natural parametrization of the geodesic segments $[\gamma(0), \gamma(1)]$ and $[\gamma(0), x]$). Suppose that the map f is not strictly convex. Then, there exists a point t in $]0, 1[$ satisfying

$$\begin{aligned}|x - x_t| &= (1 - t)|x - \gamma(0)| + t|x - \gamma(b)|\\ &= |x - y_t| + t|x - \gamma(b)|\\ &\geq |x - y_t| + |y_t - x_t|\\ &= |x - x_t|.\end{aligned}$$

Thus, the two large inequalities in the last sequence are equalities, which implies that the union $[x, y_t] \cup [y_t, x_t]$ is a geodesic segment. Since $[\gamma(0), \gamma(1)]$ is also a geodesic segment, the three points x, $\gamma(0)$ and $\gamma(1)$ are contained in a geodesic segment (use the fact that in a Busemann space, a local geodesic is a geodesic), which is a contradiction. This completes the proof of Proposition 8.2.5. □

Corollary 8.2.6. *If X, γ and x are as in Proposition* 8.2.5*, there is a unique projection of x on the segment $\gamma([a, b])$.*

Proof. A strictly convex function has a unique minimum. □

Corollary 8.2.6 will be generalized below (Proposition 8.4.10).

8.3 Geodesic convexity in Busemann spaces

In Chapter 2, we introduced the notion of geodesically convex set in an arbitrary uniquely geodesic metric space (Definition 2.5.1). Of course, the results that we proved there are valid for Busemann spaces.

It follows immediately from the definitions that if A is a geodesically convex subset of a Busemann space, then A, equipped with the induced metric, is itself a Busemann space. (In fact, this is Proposition 8.1.5 above). Balls are examples of such subsets:

Proposition 8.3.1. *Let X be a Busemann space. The open balls and the closed balls in X are geodesically convex.*

Proof. Let B be an open ball in X, of center x and radius r. For x_0 and x_1 in B, let $[x_0, x_1]$ be the unique geodesic segment joining these points. By property (ii) in Proposition 8.1.2, for every t in $[0, 1]$, the point x_t (using the natural parametrization of the segment $[x_0, x_1]$) satisfies

(8.3.1.1) $$|x - x_t| \leq (1-t)|x - x_0| + t|x - x_1| < (1-t)r + tr = r.$$

This shows that $[x_0, x_1] \subset B$. Thus, B is geodesically convex. In the case where B is a closed ball, the proof is the same up to replacing the strict inequality in (8.3.1.1) by a large one. □

Thus, we have the following

Corollary 8.3.2. *An open ball or a closed ball in a Busemann space, equipped with the induced metric, is itself a Busemann space.*

The following proposition gives an example of a "local-implies-global" property for geodesically convex subsets in a Busemann space.

Proposition 8.3.3 (Geodesic convexity implied by a local condition). *Let X be a proper Busemann space that is proper and let A be a closed subset of X which, equipped with the metric induced by that of X, is connected by rectifiable arcs. Suppose that every point x in A has an open neighborhood $V(x)$ in X such that for all y and z in $V(x)$, the geodesic segment (for the metric of X) that joins them is contained in A. Then A is convex.*

Proof. Let y and z be two points in A. By Proposition 1.4.12, there exists a path $\gamma\colon [a,b] \to A$ of minimal length (with respect to the induced metric on A) joining y to z. We show that γ is a geodesic path (with respect to the metric of X.) By Corollary 8.2.3, it suffices to show that γ is a local geodesic. Let t be in $]a,b[$ and let $x = \gamma(t)$. Let us choose an open neighborhood $V(x)$ of x with the property that for all y and z in $V(x)$, the geodesic segment (for the metric of X) that joins them is contained in A. Let $[c,d]$ be a closed sub-interval of $]a,b[$ that is contained in $\gamma^{-1}\big(V(x)\big)$ and that contains the point t in its interior. Since $\gamma(a)$ and $\gamma(b)$ are in $V(x)$, the unique geodesic segment $[\gamma(a),\gamma(b)]$ (for the metric of X) joining $\gamma(a)$ and $\gamma(b)$ is contained in A. Therefore, this geodesic segment is the image of the minimal length path in A joining $\gamma(a)$ and $\gamma(b)$. (In particular, such a minimal segment is unique.) Thus, the image of $\gamma_{|[a,b]}$ is a geodesic segment. This proves that γ is a local geodesic and therefore it is a geodesic. This completes the proof of Proposition 8.3.3. □

We note that the hypotheses of Proposition 8.3.3 do not imply that A is an open subset of X, as one can see by taking X to be a simplicial tree and A a geodesic segment having a vertex of valency ≥ 3 in its interior.

8.4 Convex functions on Busemann spaces

We start by introducing a notion that is useful in an arbitrary geodesic metric space.

Definition 8.4.1 (Convex function). Let X be a geodesic metric space and let $f\colon X \to \mathbb{R}$ be a map. We say that f is *convex* (respectively *strictly convex*) if for every geodesic path $\gamma\colon [a,b] \to X$, the map $f \circ \gamma\colon [a,b] \to \mathbb{R}$ is convex (respectively strictly convex).

For later use, we record a few properties of convex functions.

Proposition 8.4.2 (Limits convex functions). *Let X be a geodesic metric space and let $f_n\colon X \to \mathbb{R}$ $(n \geq 0)$ be a sequence of convex functions on X converging pointwise to a function f. Then f is convex.*

Proof. The proof is clear from the definitions. □

The proof of the following proposition is analogous to that of Proposition 6.1.19, and we omit it here.

Proposition 8.4.3 (Composition of convex functions). *X be a geodesic metric space, let $f: X \to \mathbb{R}$ be a convex function and let $g: f(X) \to \mathbb{R}$ be an increasing convex (respectively strictly convex) function. Then $g \circ f: X \to \mathbb{R}$ is convex (respectively strictly convex).* □

For any map $f: X \to \mathbb{R}$, we say that a point x_0 in X is a *minimum* for f if $f(x_0) = \min_{x \in X} f(x)$.

Proposition 8.4.4. *Let X be a geodesic metric space and let $f: X \to \mathbb{R}$ be a convex function. If x_0 and x_1 are two minima of f and if $\gamma: [a,b] \to \mathbb{R}$ is a geodesic path joining these points, then f is constant on the image of γ.*

Proof. Let $m = f(x_0) = f(x_1) = \min_{t \in [a,b]} f \circ \gamma(t)$. Since f is convex, the map $f \circ \gamma: [a,b] \to \mathbb{R}$ is convex on $[a,b]$, and it satisfies, for all t in $[a,b]$,

$$f(t) \leq (1-t) f \circ \gamma(a) + t f \circ \gamma(b) = (1-t)m + tm = m.$$

On the other hand, since x_0 and x_1 are minima, we have $f \circ \gamma(t) \geq m$. Thus, we have $f \circ \gamma(t) = m$ for every $t \in [a,b]$. □

Proposition 8.4.5. *Let X be a geodesic metric space and let $f: X \to \mathbb{R}$ be a strictly convex function. Then f has at most one minimum in X.*

Proof. Suppose that x_0 and x_1 are two distinct minima of f, let $m = f(x_0) = f(x_1) = \min_{t \in [a,b]} f \circ \gamma(t)$ and let $\gamma: [a,b] \to \mathbb{R}$ be a geodesic path joining x_0 and x_1. Then we have

$$f\left(\gamma\left(\frac{a+b}{2}\right)\right) < \frac{f\big(\gamma(a)\big)}{2} + \frac{f\big(\gamma(a)\big)}{2} = \frac{m}{2} + \frac{m}{2} = m$$

which contradicts the fact that x_0 and x_1 are minima of f. □

We introduced the notion of geodesically convex subsets in uniquely geodesic spaces. The following proposition, which is analogous to Proposition 6.1.23, establishes a relation between convex functions and convex subsets.

Proposition 8.4.6 (Sublevels are geodesically convex). *Let X be a uniquely geodesic space. Then, for every convex function $f: X \to \mathbb{R}$ and for every real number α, the sublevel set*

$$f_\alpha = \{x \in X,\ f(x) \leq \alpha\}$$

and the strict sublevel set

$$f_\alpha^* = \{x \in X,\ f(x) \le \alpha\}$$

are geodesically convex.

Proof. Let x_0 and x_1 be two points in f_α and for every t in $[0, 1]$, let $x_t = (1-t)x_0+tx_1$ (using the natural parametrization of the unique geodesic segment $[x_0, x_1]$ joining x_0 and x_1). By convexity of f, we have

$$f(x_t) \le (1-t)x_0 + tx_1 \le (1-t)\alpha + t\alpha = \alpha.$$

This shows that f_α is convex. For the strict sublevel set, the proof is the same up to replacing the last large inequality by a strict inequality. □

Examples 8.4.7 (Convex functions). In all the following examples, X is a Busemann space.

(i) *Distance function.* For every x_0 in X, the map $x \mapsto |x - x_0|$ is convex. This follows directly from the definition of a Busemann space (see Definition 8.1.1 in which we take $\gamma\colon [a, b] \to X$ to be the constant map $= x_0$.) Furthermore, for every $\alpha > 1$ the map $x \mapsto |x - x_0|^\alpha$ is strictly convex. Indeed, this map is the composition of the convex map $x \mapsto |x-x_0|$ with the map $t \mapsto t^\alpha$ defined on $[0, \infty[$, which is increasing and strictly convex if $\alpha > 1$. Thus, the result follows from Proposition 8.4.3.

We note that if X is an arbitrary metric space, the convexity of the distance function $x \mapsto |x - x_0|$ for all x_0 in X does not imply that X is a Busemann space, since as we saw, any normed vector space satisfies this property, and some of these normed spaces are not Busemann spaces.

(ii) *Distance function defined on a convex subset.* This is slightly more general than Example (i). Let A be a geodesically convex subset of X. By Proposition 8.1.5, A is itself a Busemann space. For any x_0 in X, the map $d_{x_0,A}\colon A \to \mathbb{R}$ defined by $x \mapsto |x - x_0|$ is convex. The reason is the same as for Example (i). Furthermore, for any $\alpha > 1$, the map $d^\alpha_{x_0,A}\colon A \to \mathbb{R}$ defined by $x \mapsto |x - x_0|^\alpha$ is strictly convex. This follows also from Proposition 8.4.3 or from Proposition 6.1.19.

(iii) *Distance to a convex subset.* Let A be a nonempty convex subset of X. Then, the distance function $d_A\colon X \to \mathbb{R}$ is convex. The proof can be done as a slight variation of the proof of Proposition 6.1.4. Important examples of distance functions are distances to a point (Example (i) above) or distances to a geodesic segment or to a geodesic ray (we shall use these functions in Chapter 10).

(iv) *Displacement function.* For every isometry $f\colon X \to X$, the displacement function $d_f\colon X \to \mathbb{R}$ is convex. Indeed, let $\gamma\colon [a, b] \to X$ be a geodesic path. Since f is an isometry, $f \circ \gamma\colon [a, b] \to X$ is also a geodesic path. Thus, the map $t \mapsto |\gamma(t) - f \circ \gamma(t)|$ defined on $[a, b]$ is convex. But this map is precisely the map $d_f \circ \gamma$. This shows that the map $d_f\colon X \to \mathbb{R}$ is convex.

(v) *Positive linear combinations of (strictly) convex functions* Let n be an integer ≥ 1. For every $i = 1, \ldots, n$, let $f_i : X \to \mathbb{R}$ be a (strictly) convex function and let λ_i be a positive real number. Then the function $\sum_{i=1}^{n} \lambda_i f_i$ is (strictly) convex. This follows from the analogous property for convex functions defined on convex subsets of an affine space (Example 6.1.3 (iii).)

Proposition 8.4.8 (Projections). *Let X be a Busemann space and let A be a nonempty closed convex subset of X. Then each point x in X has a unique projection on A.*

Proof. The existence of the projection follows the fact that A is nonempty and closed. For the uniqueness, consider the map $d_{x,A} : A \to \mathbb{R}$ defined by $y \mapsto |x - y|$ for all y in A. We already know that this function is convex (Example 8.4.7 (ii) above). Let us fix a real number $\alpha > 1$. Then, using Proposition 8.4.3 or Example 8.4.7 (ii), the map on A defined by $y \mapsto |x - y|^\alpha$ is strictly convex. Therefore, by Proposition 8.4.5, this map has a unique minimum. A projection of x on A is a minimum of this function. Therefore, the projection is unique. □

A *weighted point* in a space X is a pair (p, m) where p is a point in X and m a positive real number.

We now consider the notion of center of mass of a finite collection of weighted points in a Busemann space. This notion generalizes the classical notion of barycenter in $\mathbb{R}^n$. We start with the following

Proposition 8.4.9. *Let X be a Busemann space and let*

$$D = \{(p_1, m_1), \ldots (p_n, m_n)\}$$

be a nonempty set of weighted points in X. Then, for every $\alpha > 1$, the map on X defined by $x \mapsto |x - p_i|^\alpha$ has a unique minimum.

Proof. From Example 8.4.7 (i) and from the fact that a positive linear combination of strictly convex functions is strictly convex, the map $x \mapsto m_i|x - p_i|^\alpha$ is strictly convex. Furthermore, this map is proper and therefore it attains a minimum. By Proposition 8.4.5, this minimum is unique. □

Definition 8.4.10. Let X be a Busemann space, let $D = \{(p_1, m_1), \ldots (p_n, m_n)\}$ be a nonempty set of n weighted points in X. The *barycenter* of D is the unique minimum of the map $x \mapsto m_i d(x, p_i)^2$. We shall denote by $\mathcal{B}(D)$ the barycenter of the set D.

The barycenter depends continuously on points and weights. It is also clear that if the set D is reduced to a single weighted point, (p, m), then $\mathcal{B}(D) = p$.

Corollary 8.4.11 (Cartan). *Let X be a Busemann space and let G be a finite group of isometries of X. Then, there exists a point p in X that is fixed by every element of G.*

Proof. Take an arbitrary point in X and let $\{p_1, \ldots, p_n\}$ be its orbit under the group G. Let D be this collection of points, each of them equipped with the weight 1 and let $p = \mathcal{B}(D)$ the corresponding barycenter. The set D is invariant by G and therefore the point p is fixed by this group. □

Corollary 8.4.11 is usually called Cartan's Theorem because, in [36] p. 354, Elie Cartan uses the same arguments (uniqueness of the minimum of the sum of the square of the distance functions) to prove an analogous result in the setting of simply connected Riemannian manifolds of nonpositive curvature.

A more general statement is the following (also classical) result:

Proposition 8.4.12. *Let X be a Busemann space and let G be a compact topological group acting by isometries on X. Then there exists a point p in X that is fixed by the whole group G.*

Proof. The compact group G has a Haar measure μ, that is, a nonnegative measure of finite non-zero total mass that is invariant by left and right translations[2]. We choose an arbitrary point x in X and we let μ be the measure on the orbit Gx induced from the Haar measure of x. We consider this measure μ as a measure on X. Then the function $D\colon X \to \mathbb{R}$ defined by

$$D(p) = \int_X d^2(p, x)\, d\mu(x)$$

is proper and strictly convex, as a mean of squares of distance functions, and therefore it has a unique minimum. This minimum is a fixed point for the group G. □

We note finally that there is a useful notion of center of mass of a measure defined on an arbitrary metric space that generalizes the notion of barycenter that we consider here; see [80].

We conclude this chapter with the following result, which is due to Cartan in the case where X is a Riemannian manifold of nonpositive curvature (cf. [36] p. 354).

Theorem 8.4.13 (Cartan). *Let X be a complete locally compact and locally convex length space. Then, the fundamental group of X contains no element of finite order except the identity. As a consequence, either X is simply connected or the fundamental group of X is infinite.*

Proof. Let $\tilde{X}$ be the universal covering of X. We equip $\tilde{X}$ with its canonical length metric (Proposition 3.5.1). Since $\tilde{X}$ is locally isometric to X, it is also locally compact

[2] The existence and uniqueness (up to a multiplication by a scalar factor) of Haar measures on compact groups (and more generally on locally compact groups) was proved by R. Haar (see [58]) with the additional assumption that the group G is separable. For a proof of existence and uniqueness without this assumption, we refer the reader to [143].

and locally convex. Furthermore, by Proposition 3.5.5, $\tilde{X}$ is complete. By the theorem of Hopf–Rinow (Theorem 2.4.6), $\tilde{X}$ is geodesic. Now the fundamental group of X is canonically isomorphic to the group G of deck transformations of the covering $\tilde{X} \to X$. Suppose G contains an element g of finite order. Then, by Corollary 8.4.11, there exists a point p in X that is fixed by g. But any element of G, acting as a deck transformation, has no fixed point, unless it is the identity. Therefore, G has no element of finite order. Finally, if X is not simply connected, then its fundamental group contains an element that is distinct from the identity, and which therefore is of infinite order. □

Notes on Chapter 8

Busemann's definition of nonpositive, of zero and of negative curvatures. Busemann calls a "nonpositively curved space" a G-space in which each point has a neighborhood that is a Busemann space in our sense. We shall call such a space a "locally convex space" and we shall study these spaces in Chapter 9. Condition (vii) of Proposition 8.1.2 is the one that Busemann uses in [26] p. 237 to define nonpositive curvature. In the case where the space satisfies locally Condition (vii) with the large inequality sign $\leq$ replaced by an equality sign, Busemann calls this space a "space of curvature 0". If the condition is locally satisfied with a strict inequality sign $<$ instead of the large inequality sign, for all triples of points x_0, x_1 and x_1' that are not contained in a geodesic segment, then Busemann says that the space is "negatively curved".

Convex functions on simply connected nonpositively curved Riemannian manifolds. The classical examples of Busemann spaces are the simply connected Riemannian manifolds of nonpositive sectional curvature. A theory of convex functions defined on such spaces is developed in the paper [17] by R. Bishop and B. O'Neill. In this paper, the authors give several methods of constructing convex functions and strictly convex functions on Riemannian manifolds of nonpositive sectional curvature or of negative curvature. Some of these examples are in the spirit of the examples in 8.4.7 above. For instance, they prove the following (Theorem 4.1 of [17], p. 12): if M is a complete simply connected Riemannian manifold of nonpositive sectional curvature, then, for every closed convex submanifold S of M, the square of the "distance function to S", that is, the function $d_S^2 \colon M \to \mathbb{R}$ defined by

$$x \mapsto \big(d(x,S)\big)^2$$

is convex. Furthermore, if the sectional curvature of M is negative, then d_S^2 is strictly convex on $M \setminus S$.

We note that in the context of Riemannian manifolds, the square of the distance function has the advantage (over the distance function) of being smooth (in fact, it is C^∞), so that convexity boils down to having positive Hessian, and the techniques of classical differential geometry can be applied to this function. (Distance functions are only continuous in the general case.)

Let us also note that several theorems in the paper [17] have been generalized by Busemann and Phadke in [32] to the context of peakless functions defined on G-spaces. For instance, Busemann and Phadke gave in that paper an example of a surface of revolution in $\mathbb{R}^3$ that does not carry nonconstant peakless functions.

Cartan's Theorem and the barycenter of a measure. There is a useful notion of barycenter of a measure, which generalizes the notion of barycenter of a collection of weighted points that is used in Cartan's Theorem (Corollary 8.4.11 above). The idea for the definition of the barycenter of a measure is used in the proof of Proposition 8.4.12. In the paper [17] by Bishop and O'Neill, the authors prove that if M is a complete simply-connected Riemannian manifold of nonpositive sectional curvature, then for any finite positive measure μ on M with compact support, the map

$$f_\mu : p \mapsto \int_M d^2(p,x)\, d\mu(x)$$

is strictly convex, and therefore it has a unique minimum point in M. This result implies Cartan's theorem and its generalization (Proposition 8.4.12) for simply connected complete Riemannian manifolds of nonpositive curvature.

For a notion of barycenter of a measure defined on an arbitrary metric space, we refer the reader to [80].

Convex functions and nonnegative curvature. The global study of manifolds of nonnegative sectional curvature (that is, the large-scale behaviour of geodesics, the structure of the isometry group and so on) is also an active field since many decades, and the theory of convex subsets and convex functions intervenes as well in that study. In fact, important geometrical and topological information on such manifolds have been deduced from the existence of non-constant convex functions. Let us give a few examples. In the 1930s, Cohn-Vossen obtained a famous result that states that a noncompact complete 2-dimensional manifold of nonnegative curvature is either diffeomorphic to $\mathbb{R}^2$ or is flat (see [40]). Gromoll and Meyer, generalizing Cohn-Vossen's techniques and results to higher dimensions, proved that each noncompact complete manifold of positive sectional curvature is diffeomorphic to $\mathbb{R}^n$. Their proof relies on considering compact convex subsets in such a manifold and proving the existence of (not necessarily smooth) convex functions (see [54]). L. Greene and H. Wu gave a proof of the theorem of Gromoll and Meyer by constructing, on each complete noncompact manifold of positive sectional curvature, a $\mathcal{C}^\infty$ strictly convex function with compact sublevel sets, which allowed the use of Morse theory (see [52] and [53]). Cheeger and Gromoll generalized these convexity techniques so as to be useful in the study of manifolds of nonnegative (and not only positive) curvature (see [38]). For example, they proved that one can define a nonconstant convex function on any complete noncompact manifold with nonnegative sectional curvature. S.-T. Yau proved in [147] that if a complete manifold admits a nonconstant convex function, then its volume is infinite. The last two results combined imply that any complete noncompact manifold with nonnegative sectional curvature has infinite volume.

The Nielsen realization problem and convexity in Teichmüller space. S. Kerckhoff proved in [85] that any finite subgroup of the mapping class group $\mathcal{M}_g$ of a closed oriented surface $S = S_g$ of genus $g \geq 2$, acting on the Teichmüller space $\mathcal{T}_g$, has a fixed point. This result gave a positive answer to the famous *Nielsen realization problem*. We mention here this fact because Kerckhoff's proof is based on a convexity argument that generalizes, in a highly nontrivial manner, the case of genus one (the case where the surface is the two-dimensional torus). The genus one case boils down to the fact that every finite subgroup of $\mathrm{SL}(2, \mathbb{Z})$, acting by isometries on hyperbolic 2-space, has a fixed point, and this result can be proved using Cartan's theorem, cf. Corollary 8.4.11 above. It is interesting to recall here that for some time, it was believed that the Nielsen realization problem was settled, as a consequence of a paper by S. Kravetz [93], in which this author wrongly claimed that the Teichmüller space of a surface of any genus is a Busemann space, and therefore that the solution of the Nielsen problem in any genus followed immediately from Cartan's theorem. After the discovery by M. Linch (see [94]) of a flaw in Kravetz's paper, the Nielsen problem became again an open problem.

Kerckhoff's solution of the Nielsen realization problem, instead of using the convexity of the Teichmüller distance function (which, by a result of Masur [99], is known to be false), uses the convexity of the *geodesic length functions* $l_\alpha \colon \mathcal{T}_g \to \mathbb{R}$ along certain paths in Teichmüller space. These paths are associated to measured geodesic laminations and they generalize the Fenchel–Nielsen twists associated to simple closed curves. They were introduced by Thurston who called them *earthquake paths*. In the case where the surface is the two-torus, the earthquake paths are the horocycles of $\mathbb{H}^2$. To describe briefly Kerckhoff's arguments, we recall that if α is a homotopy class of simple closed curves in the surface S equipped with a hyperbolic structure and considered as an element of Teichmüller space, one defines $l_\alpha(S)$ to be the length of the unique closed geodesic on S that is in the homotopy class α. In this way, one obtains a function $l_\alpha \colon \mathcal{T}_g \to \mathbb{R}$. Kerckhoff proved that for any earthquake path $\gamma \colon \mathbb{R} \to S$ along a measured geodesic lamination μ, we have

$$\frac{d(l_\alpha \circ \gamma(t))}{dt}(0) = \int_\mu \cos\theta \, d\mu$$

where θ is the angle that the simple closed curve α that makes with the lamination μ. To understand this formula, it is useful to first consider the case where the lamination μ is a simple closed curve β. In this case, the formula becomes

$$\frac{d(l_\alpha \circ \gamma(t))}{dt}(0) = \sum_{k=0}^{n} \cos\theta_k$$

where $\theta_1, \dots, \theta_n$ are the angles (from β to α) at the points where α and β intersect transversely. The quantity $\int_\mu \cos\theta \, d\mu$ can be seen as a limit of sequences of weighted quantities of the form $\sum_{k=0}^{n} \cos\theta_k$ associated to a sequence of weighted closed curves converging to μ. The convexity of the length function follows from the fact that a quantity such as $\cos\theta_k$ or $\int_\mu \cos\theta \, d\mu$ is increasing along any earthquake path.

Once the setting and the (extremely interesting) basic results have been established, Kerckhoff's proof is much in the spirit of Cartan's theorem. First, one needs strict convexity instead of convexity and for that purpose, rather than considering a unique length function associated to a homotopy class α, Kerckhoff considers a finite sum of length functions $\sum_{i=1}^{n} l_{\alpha_i} : \mathcal{T}_g \to \mathbb{R}$, where $\{\alpha_1, \ldots, \alpha_n\}$ is a set of homotopy classes of simple closed curves that *fills up the surface* S. This means that any set of curves representing these homotopy classes has nonempty intersection with any homotopically nontrivial simple closed curve on S, and this property implies that the function $\sum_{i=1}^{n} l_{\alpha_i} : \mathcal{T}_g \to \mathbb{R}$ is strictly convex along any earthquake path. Then, given any finite subgroup G of $\mathcal{M}_g$, by replacing the set of homotopy classes by the set of its iterates under G, one can assume that the collection $\{\alpha_1, \ldots, \alpha_n\}$ is invariant by the action of G. Finally, a result of Thurston that says that each pair of points in $\mathcal{T}_g$ can be joined by an earthquake path, together with the invariance of the function $\sum_{i=1}^{n} l_{\alpha_i}$, show that this function has a unique minimum in $\mathcal{T}_g$, and this minimum is a G-invariant point. This is the fixed point that we seek.

We note that the result on the convexity of length functions associated to simple closed curves on the surface has been extended by Kerckhoff to the convexity of length functions associated to geodesic laminations (see [86].)

Let us also note that S. Wolpert, in his paper [144], gives an explicit formula for the second derivative of the length function along Fenchel–Nielsen twist paths, and observing the sign of this second derivative gives another way of seeing the convexity of the length function.

Convexity and the Weil–Petersson metric. In his paper [146], Wolpert worked out a theory of the convexity of the geodesic length functions on Teichmüller space $\mathcal{T}_g$ equipped with its Weil–Petersson metric instead of the Teichmüller metric. We recall that Teichmüller space has a natural complex structure on which the mapping class group acts as a group of biholomorphic mappings, and the Weil–Petersson metric is closely related to that structure. If S is a Riemann surface representing an element of Teichmüller space, then the Weil–Petersson metric is induced by an inner product defined on the space $Q(S)$ of holomorphic quadratic differentials on S, this space $Q(S)$ being naturally identified with the holomorphic cotangent space to Teichmüller space at the point S. Explicitly, a holomorphic quadratic differential on S is an invariant object that has the form $\phi(z)dz^2$ where z is a holomorphic local coordinate and ϕ a holomorphic function in z. Invariance means here invariance under change of coordinates, that is, if w is another holomorphic local coordinate, and if in this local coordinate the holomorphic quadratic differential is written as $\psi(w)dw^2$, then at the overlap between the two charts we have $\phi(z)(dz/dw)^2 = \psi(z)$. Now given two quadratic differentials ϕ and ψ on the Riemann surface S, if $\rho(z)|dz|$ denotes the hyperbolic line element on the surface S equipped with its canonical Poincaré metric, then the Weil–Petersson inner product of ϕ and ψ is defined by

$$\langle \phi, \psi \rangle = \int_S \frac{\phi(z)\overline{\psi}(z)}{(\rho(z))^2}\, dz d\overline{z}.$$

There is a natural duality between the tangent and cotangent spaces at S of Teichmüller space, which is defined by the pairing

$$(\mu, \phi) \mapsto \int_S \mu(z)\phi(z)\, dz d\overline{z},$$

where ϕ is an element of $Q(S)$ and where μ is a measurable Beltrami differential representing a tangent vector to Teichmüller space at S. We recall that a measurable *Beltrami differential* on S is an object which is given in local coordinates as $\mu(z) d\overline{z}/dz$, where μ is a measurable function defined on the range of the coordinate z, with the following invariance property: if w is another local coordinate and if this Beltrami differential is given in this local coordinate as $\nu(w)dw$, then we have, at the overlap between the two charts z and w, $\mu(z)dz/dw = \nu(w)d\overline{z}/d\overline{w}$. We note by the way that this implies that the function $|\mu(z)|$ is well-defined on the whole surface S.

A measurable Beltrami differential on a Riemann surface describes an infinitesimal change in the conformal structure. Using this duality, the Weil–Petersson inner product defines an inner product on the tangent space of Teichmüller space, which gives in particular a Riemannian metric, which is the Weil–Petersson metric. (In fact, we have a Kähler metric.) We note by the way that the Teichmüller metric, as a Finsler metric, is associated to the following norm on cotangent space to $\mathcal{T}_g$:

$$\|\phi\| = \int_S |\phi(z)|\, dz d\overline{z},$$

for any holomorphic quadratic differential ϕ. Here, the norm on tangent space to Teichmüller space is defined using the pairing

$$(\mu, \phi) \mapsto \langle \mu, \phi \rangle = \mathrm{Re} \int_S \mu(z)\phi(z)\, dz d\overline{z}.$$

In fact, the tangent space at S of Teichmüller space is identified with the space of measurable Beltrami differentials quotiented by the subspace of Beltrami differentials μ satisfying $\langle \mu, \phi \rangle = 0$ for every quadratic differential ϕ. In this way, the pairing $(\mu, \phi) \mapsto \langle \mu, \phi \rangle$ induces an isomorphism between tangent and cotangent space to Teichmüller space at S.

We now return to the Weil–Petersson Riemannian metric. This metric has negative sectional curvature, but it is not complete, and therefore the techniques of global geometry do not all apply to it a priori. However, results of Wolpert show that $\mathcal{T}_g$, equipped with the Weil–Petersson metric, behaves fairly like a complete metric space of negative curvature. In fact, the failure of completeness is due to the fact that some geodesics leave the space in finite time, and there are no non-completeness phenomena comparable to the fact that the space $\mathbb{Q}$ of rationals, equipped with its usual metric, is not complete. Wolpert showed that $\mathcal{T}_g$ is a uniquely geodesic space, that the exponential map at each point is distance-increasing and then, that any finite group of isometries has a fixed point. Using these results, he gave a new version of the solution of the Nielsen realization problem. Another interesting result due to Wolpert that is related to our subject here is that the geodesic length functions are convex along Weil–Petersson geodesics. Wolpert also showed that the sublevel sets for the geodesic length functions, that is, the sets of the form

$$\{h \in \mathcal{T}_g \mid l_\alpha(h) < M\},$$

where M is any positive constant and α any homotopy class of simple closed curves, are convex (with Teichmüller space being again equipped with the Weil–Petersson metric). This result is an analog of a result that was obtained by Kerckhoff in [86], that concerns the Teichmüller metric. (We note by the way that although Kerckhoff's paper appeared in print in 1992, a preliminary version was already circulated in 1983).

Thus, the works of Kerckhoff and Wolpert are the bases of two parallel theories of convexity in Teichmüller space.

Chapter 9

Locally convex spaces

Introduction

A locally convex metric space is a metric space in which every point has a neighborhood which, equipped with the induced metric, is a Busemann space. In other words, each point has a neighborhood in which the distance function associated to two geodesic segments contained in that neighborhood is convex.

In each geodesic metric space X and to each point x in X, we consider the space G_x of affinely reparametrized local geodesics $\gamma\colon [0,1] \to X$ satisfying $\gamma(0) = x$. The space G_x is called the *tangent space* of X at x and it is equipped with the topology of uniform convergence. There is a natural map $\exp_x\colon G_x \to X$ that associates to each affinely reparametrized local geodesic $\gamma\colon [0,1] \to X$ starting at x its endpoint $\gamma(1)$. This map is called the *exponential map at* x.

In this chapter, we prove the following result of S. B. Alexander and R. L. Bishop:

Theorem. *Let X be a geodesic complete locally compact and locally convex metric space. Then for every x in X, the exponential map $\exp_x\colon G_x \to X$ is a universal covering map of X.*

This result generalizes a classical theorem which is usually referred to as the "Cartan–Hadamard Theorem" and which gives an analogous description of the tangent space at x in the case where X is a Riemannian manifold of nonpositive curvature (see Cartan's *Leçons sur la géométrie des espaces de Riemann* [36], p. 347).

We also prove the following local-implies-global property:

Theorem. *Every complete geodesic locally compact, locally convex and simply connected metric space is a Busemann space.*

This result is due to M. Gromov (see [55] p. 187, where a first general version is also stated under the name "Cartan–Hadamard Theorem"). The proof that we give here is again due to Alexander and Bishop.

We recall that by the result of Hopf and Rinow (Theorems 1.1.16 and 2.4.6), the assumption that X is geodesic is automatically satisfied if we assume that X, besides being locally compact, is a complete length space.

The plan of this chapter is the following:

Section 1 contains definitions and examples of locally convex metric spaces. We shall linger over the description of surfaces equipped with flat metrics with cone type singularities, with cone angle $> 2\pi$ at each singularity.

In Section 2, we prove the existence and the uniqueness, in a complete locally convex metric space, of an affinely reparametrized local geodesic joining any two points that are close enough to the image of an affinely reparametrized local geodesic, and that the affinely reparametrized local geodesic joining these two points varies continuously with respect to its endpoints.

Section 3 contains the proofs of the two theorems stated above.

9.1 Locally convex spaces

Definition 9.1.1 (Locally convex metric space). A metric space X is said to be *locally convex*[1] if every point x in X possesses a neighborhood $V(x)$ which, equipped with the induced metric, is a Busemann space.

We shall say that such a neighborhood $V(x)$ is a *Busemann neighborhood* of the point x.[2] We recall that from the definition of a Busemann space, it follows that $V(x)$ is a geodesic metric space and that the geodesic segment joining any two points in $V(X)$ is unique.

Examples 9.1.2 (Euclidean and hyperbolic manifolds). A Euclidean or a hyperbolic manifold of any dimension $n \geq 1$ is a locally convex metric space. Indeed, each point x in such a manifold possesses a neighborhood that is isometric to an open ball in Euclidean space $\mathbb{E}^n$ or in hyperbolic space $\mathbb{H}^n$ and, therefore, such a neighborhood is a Busemann neighborhood of x.

Example 9.1.3 (Simplicial graph). A simplicial graph equipped with a simplicial metric is a locally convex metric space, since every point in such a graph has a neighborhood which, equipped with the induced metric, is isometric to a simplicial tree equipped with a simplicial metric, which is a Busemann space (Example 8.1.3 (iii)).

Now, we describe another interesting class of locally convex spaces.

First, we recall the notion of 2-dimensional Euclidean cone. We start with Euclidean cones with cone angle $< 2\pi$. (We already encountered such objects in Example 2.4.16 (ii).) To obtain such a cone, one can start with a sector AOB in a Euclidean disk of center O that is bounded by two rays OA and OB, as in Figure 2.7 of Chapter 2. (The angle $\angle AOB$ can take any value in $]0, 2\pi[$.) Such a sector will be a "fundamental region" for the cone. Gluing together the two sides OA and OB of this sector by an isometry, we obtain a surface V homeomorphic to a closed disk, equipped with a metric that is locally Euclidean except at the image of the point O in V. The space V, or, any (closed or open) ball centered at the image of O in V, is,

[1] In the terminology used by Busemann, such a space would be called a *nonpositively curved space*.

[2] It would have been natural to call $V(x)$ a "convex neighborhood" of x, but we prefer the term "Busemann neighborhood" since we already used the term "convex neighborhood" to denote a neighborhood of a point in a metric space which, equipped with the induced metric, is a geodesic metric space.

by definition, a *Euclidean cone* whose *cone angle* is $\angle AOB$. To obtain a Euclidean cone with cone angle $> 2\pi$, we take several sectors such as AOB and glue them side by side in such a way as the result of the gluing is a connected surface. This surface, or any (closed or open) ball centered at the image of the point O in this surface, is a Euclidean cone whose cone angle is the sum of the angles of the various sectors. For instance, gluing four sectors whose angles are equal to $\pi/2$ gives a Euclidean cone whose cone angle is equal to 2π. Thus, in this case, the resulting cone is isometric to a disk in the Euclidean plane, and the vertex is in fact a regular point. Gluing three sectors whose angles at the vertices are all equal to π gives a Euclidean cone isometric to a ball centered at the armpit of a T-shirt (Figure 9.1).

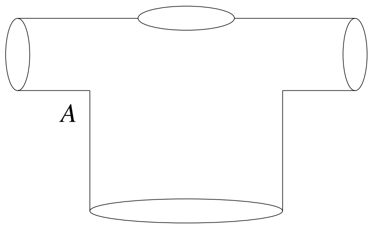


Figure 9.1. At the point A the cone angle is 3π.

Let us consider a surface S equipped with a metric that is locally Euclidean except for an isolated set of points that are called the *singularities*, such that each of these singularities has a neighborhood that is isometric to a Euclidean cone whose vertex corresponds to that singular point. We say that such a point is a *conical point of total angle* α if α is the angle at the vertex of the corresponding Euclidean cone.

We shall call such a structure a *singular flat metric* on the surface S. Such structures appear in several contexts in the geometry of surfaces. For instance, surfaces equipped with singular flat metrics can be obtained by gluing a collection of Euclidean triangles by isometries along edges of equal lengths, and taking the length metric on the resulting surface. Another context where such metrics appear is the geometric theory of holomorphic quadratic differentials on Riemann surfaces. Equivalently, these surfaces appear as metric structures defined by two transverse measured foliations on the topological underlying surface (see for instance [74]). In Chapter 2 (Example 2.4.3 (ii)), we already saw that pairs of transverse measured foliations are useful for describing conformal structures and Teichmüller geodesics on a surface. Let us say a few words on how two transverse measured foliations determine a singular flat metric on a surface. In the complement of the singular points, the two transverse measured foliations define local charts whose range is a subset of the Euclidean plane $\mathbb{E}^2$ and for which the coordinate changes are local Euclidean isometries. This provides a Euclidean structure on the complement of the singular points, and it is easy to see, from the local models of the singular points of the foliations considered (Figure 2.8 of Chapter 2) that in a neighborhood of each singular point, this structure extends as a singular Euclidean structure with a cone type singularity at that point. Thus, the Euclidean structure defined on the complement of the singular points extends to the whole surface as a singular flat metric. In these examples, every cone angle at each singular point is an integer multiple of π.

Now the important fact for us here is the following

Proposition 9.1.4 (Singular flat metric). *Let S be a surface equipped with a singular flat metric such that at each singular point, the cone angle is $> 2\pi$. Then S is a locally convex metric space.*

Proof. Each non-singular point in S possesses a neighborhood that is isometric to an open disk in the Euclidean $\mathbb{E}^2$ and such a neighborhood, equipped with the induced metric, is a Busemann space. Therefore, the only problem is to prove the local convexity at the singular points. Let s be a singular point and let V be a (closed or open) disk centered at s of positive radius, this radius being small enough in order for V to be injectively embedded in S. The disk V is isometric to a Euclidean cone whose angle at the vertex is $\alpha > 2\pi$. Let us show that V is a Busemann neighborhood of s.

We start by proving that V is a geodesic space.

Let x and y be two arbitrary points in V. If one of these two points (say x) is the point s, then the Euclidean segment $[y, s]$ is a geodesic segment that is contained in V. If x and y are both distinct from s, we consider the Euclidean segments $[s, x]$ and $[s, y]$. At the point s, the two segments $[s, x]$ and $[s, y]$ define two angles opposite to each other but whose sum is greater than 2π (the sum is equal to the cone angle at s). We break up the argument into two cases.

First case. One of the angles formed by the two segments $[s, x]$ and $[s, y]$ at the point s is $\leq \pi$. Then these two segments are two sides of a Euclidean triangle sxy that is isometrically embedded in V. (We note that this triangle is degenerate if the angle that we are considering that is made by $[s, x]$ and $[s, y]$ at the point s is $= \pi$.) In this case, the side $[x, y]$ of this triangle is a geodesic segment in V that joins x and y.

Second case. The two angles formed by the two segments $[s, x]$ and $[s, y]$ at the point s are $> \pi$. Then the concatenation of the segments $[x, s]$ and $[s, y]$ is a geodesic segment in S that joins x to y and is contained in V.

Thus, in both cases, the disk V, equipped with the induced metric, is a geodesic metric space. Therefore, V is a convex neighborhood of s.

To show that V is a Busemann neighborhood of s, we prove that Property (vii) of Proposition 8.1.2 is satisfied for any two geodesic segments $[x, y]$ and $[x, z]$ in V. We shall assume for the proof that we are in the generic case where the three points x, y and z are distinct from the singular point s. (In the non-generic cases, the proof is simpler, and, in fact, we can treat these cases as limiting cases of the generic case.)

The arguments that we use are a combination of techniques of proofs for analogous properties in the case of a simplicial tree and in the case of the Euclidean plane. We consider the three Euclidean geodesic segments $[s, x]$, $[s, y]$ and $[s, z]$. It is clear from the description above that each of these segments is the unique segment in V joining its endpoints. We call m and m' the midpoints of $[x, y]$ and $[x, z]$ respectively. The three segments $[s, x]$, $[s, y]$ and $[s, z]$ define naturally three angles at s, whose sum is equal to the cone angle at this point. We distinguish five cases:

First case. The three angles that the three segments $[s, x]$, $[s, y]$ and $[s, z]$ make at the point s are $\geq \pi$. In this case, the concatenation of any two of these segments is a geodesic segment in S that is contained in V: the concatenation of $[x, s]$ and $[s, y]$ is a geodesic segment joining x and y, the concatenation of $[x, s]$ and $[s, z]$ is a geodesic segment joining x and z and the concatenation of $[y, s]$ and $[s, z]$ is a geodesic segment joining y and z. Then the tripod $[s, x] \cup [s, y] \cup [s, z]$ is a convex subset of S (any geodesic segment joining two points on this tripod is contained in the tripod). Thus, in this case, all the computations are made in the tripod $[s, x] \cup [s, y] \cup [s, z]$ equipped with the metric induced from that of S (which is the length metric of the tripod), and the proof of the required property can be done in the same way as that of Example 8.1.3 (iii) that concerns the case of a metric tree.

Second case. The angle that the segments $[s, x]$ and $[s, y]$ make at s is $\geq \pi$, the angle that the segments $[s, x]$ and $[s, z]$ make at s is $\geq \pi$ and the angle that the segments $[s, y]$ and $[s, z]$ make at s is $< \pi$. In this case, the three points s, y and z are the vertices of a Euclidean triangle that is isometrically embedded in V. Furthermore, the union of $[x, s]$ and $[s, y]$ is a geodesic segment joining x and y and the union of $[x, s]$ and $[s, z]$ is a geodesic segment joining x and z. The points m and m' are on the tripod $[s, x] \cup [s, y] \cup [s, z]$. Up to interchanging the names of y and z, we have three subcases:

First subcase. The point m is on $[s, y]$ and the point m' is on $[s, z]$. In this case, the three points s, y and z are the vertices of a Euclidean triangle that is embedded in S, and the distances $|m - m'|$ and $|y - z|$ are measured in the Euclidean triangle syz. Let I (respectively J) be the midpoint of the side $[s, y]$ (respectively $[s, z]$) of this triangle. Then I is on the segment my and J is on the segment $m'z$ (Figure 9.2 (a)). We have $|I - J| = (1/2)|y - z|$. We also have $|m - m'| \leq |I - J|$. To see this, we can first consider the special case where $|s - y| = |s - z|$. In this case, in the triangle syz, mm' is parallel to IJ and therefore the required inequality is satisfied. For the general case, we can suppose without loss of generality that $|s - y| \geq |s - z|$. Then, observing the variation of the distance $|m - m'|$ when $|s - y|$ increases from a value where it is equal to $|s - z|$ until it reaches its actual value, it is easy to see in that case that $|m - m'|$ is always bounded by $|I - J|$. Thus, in any case, we have $|m - m'| \leq (1/2)|y - z|$, as required.

Second subcase. The points m and m' are on $[s, x]$, and we have, on this segment, the following ordered sequence of points: x, m, m', s (Figure 9.2 (b)). In this case, we have

$$|x - m| = (1/2)(|x - s| + |s - y|)$$

and

$$|x - m'| = (1/2)(|x - s| + |s - z|),$$

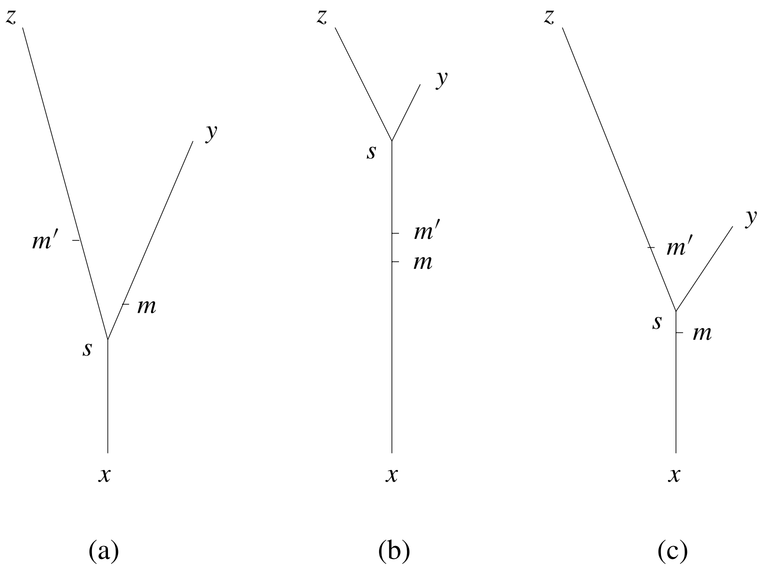


Figure 9.2. The three subcases of Case 2 in the proof of Proposition 9.1.4.

therefore

$$|m - m'| = |x - m'| - |x - m| = (1/2)(|s - z| - |s - y|).$$

From the triangle inequality, we deduce

$$\begin{aligned} |y - z| - 2|m - m'| &= |y - z| - (|s - z| - |s - y|) \\ &= |y - z| - |s - z| + |s - y| \geq 0. \end{aligned}$$

Thus, we obtain $|m - m'| \leq (1/2)|y - z|$ which again is the required inequality.

Third subcase. The point m is on $[s, x]$ and the point m' is on $[s, z]$ (Figure 9.2 (c)). Here, we note that if we bring the point x closer to the point s by a distance a that is small enough (and let us denote by x_a the new position of x), then the new position of m (which is now the midpoint of the segment $[x_a, y]$, and which we denote by m_a) will be closer from s by a distance of $a/2$, and the new position of m' (which is the midpoint of the segment $[x_a, z]$ and which we denote by m'_a) will be closer from z by a distance that is also equal to $a/2$. Thus, for a small enough, we are always in the configuration of the third subcase, and the distance $|m'_a - m_a|$ is equal to $|m' - m|$. In fact, this holds until the value a reaches the limiting value $2|s - m|$, for which the point m_a attains a limiting position (where this point becomes equal to the point s). But this limiting position is also a limiting position for the first subcase, and the equality $|m - m'| \leq (1/2)|y - z|$ is therefore also satisfied in this case.

Third case. The angle that the segments $[s, x]$ and $[s, y]$ make at s is $< \pi$, the angle that the segments $[s, x]$ and $[s, z]$ make at s is $\geq \pi$ and the angle that the segments $[s, y]$ and $[s, z]$ make at s is $\geq \pi$. In this case, the union of segments $[x, s] \cup [s, z]$ and $[y, s] \cup [s, z]$ are geodesic segments, and the three points x, s and y are the vertices of a Euclidean triangle xsy that is embedded in S. The point m is on the side $[x, y]$ of the triangle xsy whereas the point m' is on the union $[x, s] \cup [s, z]$. We distinguish two subcases :

First subcase. The point m' is on $[x, s]$ (Figure 9.3 (a)). Let I be the midpoint of $[x, s]$. We have, on that segment, the sequence of four points x, I, m', s in that order.

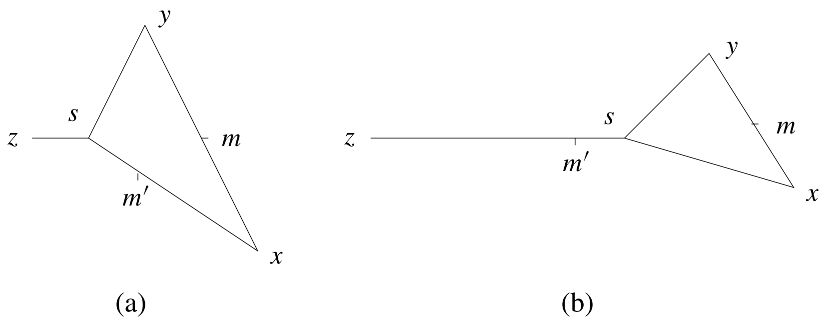


Figure 9.3. The two subcases of Case 3 in the proof of Proposition 9.1.4.

From the triangle inequality, we obtain

$$|m - m'| \leq |m - I| + |I - m'|.$$

Furthermore, we have

$$|I - m'| \leq |x - m'| = (1/2)|s - z|,$$

and in the triangle xys, we have the relation

$$|m - I| = (1/2)|y - s|.$$

Therefore, we obtain

$$|m - m'| \leq (1/2)(|y - s| + |s - z|) = (1/2)|y - z|.$$

Second subcase. The point m' is on $[s, z]$ (Figure 9.3 (b)). In this case, we have

$$|m - m'| = |s - m| + |s - m'|.$$

In the Euclidean triangle xys, we have the relation $|s-m| \leq (1/2)(|s-y|+|s-x|)$. Therefore, we obtain

$$\begin{aligned}|m-m'| &\leq (1/2)(|s-y|+|s-x|)+|s-m'|\\ &= (1/2)(|s-y|+|s-m'|)+(1/2)(|s-x|+|s-m'|)\\ &= (1/2)(|s-y|+|s-m'|)+(1/2)|z-m'|\\ &= (1/2)(|y-s|+|s-m'|+|m'-z|)\\ &= (1/2)|y-z|.\end{aligned}$$

Fourth case. The angle that the segments $[s, x]$ and $[s, y]$ make at s is $< \pi$, the angle that the segments $[s, x]$ and $[s, z]$ make at s is $< \pi$ and the angle that the segments $[s, y]$ and $[s, z]$ make at s is $\geq \pi$. In this case, the union $[y, s] \cup [s, z]$ is a geodesic segment and we have $|y-z| = |y-s|+|s-z|$. The points x, s and y (respectively x, s and z) are the vertices of a Euclidean triangle that is embedded in S. The point m' (respectively m) is on the side $[x, y]$ (respectively $[x, z]$) of that triangle. Let m'' be the midpoint of $[x, s]$. Then we have

$$|m-m'| \leq |m-m''|+|m''-m'| = (1/2)|y-s|+(1/2)|s-z| = (1/2)|y-z|.$$

Fifth case. The angle that the segments $[s, x]$ and $[s, y]$ make at s is $\geq \pi$, the angle that the segments $[s, x]$ and $[s, z]$ make at s is $< \pi$ and the angle that the segments $[s, y]$ and $[s, z]$ make at s is $< \pi$. In this case, the figure formed by the union of the two Euclidean triangles xsz and ysz having the side sz in common is isometrically embedded in S. Here, we distinguish two subcases:

First subcase. The sum of the two angles $\angle ysz$ and $\angle zsx$ is $\leq \pi$. We can assume without loss of generality that x, y and z are close enough to s so that there is a Euclidean triangle with vertices x, y and z that is embedded in S (Figure 9.4). The points m and m' are respectively the midpoints of the sides $[y, x]$ and $[z, x]$ of that triangle, and we have $|m-m'| = (1/2)|y-z|$.

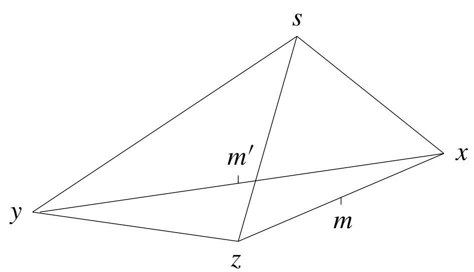


Figure 9.4. The first subcase of Case 5 in the proof of Proposition 9.1.4.

Second subcase. The sum of the two angles $\angle ysz$ and $\angle zsx$ is $> \pi$. (We recall that since we are in the fifth case, this sum is $\leq 2\pi$.) Then the union $[x, s] \cup [s, y]$ is a geodesic segment. We can assume without loss of generality that the distance $|y - z|$ is small enough so that the closed ball of center s and of radius $|s - y|$ contains no other singular point than the point s. Now let us consider the point y' such that the three points x, s and y' are contained in that order on a Euclidean segment and such that $|s - y'| = |s - y|$. Figures 9.5 (a) and 9.5 (b) represent respectively the case where m' is on the segment $[s, x]$ and the case where m is on the segment $[s, y]$. We

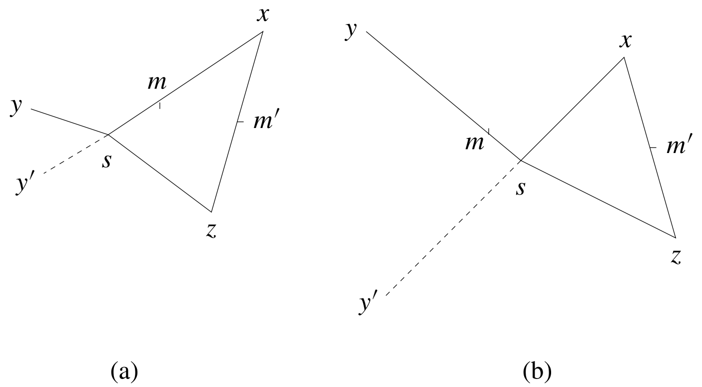


Figure 9.5. The second subcase of Case 5 in the proof of Proposition 9.1.4.

deal with the first case. The other case can be dealt with in a similar manner. In the Euclidean triangle $y'zx$, the points m and m' are the midpoints of the sides $[y', x]$ and $[z, x]$ respectively. Therefore we have

$$|m - m'| = (1/2)|y' - z|.$$

Now let us compare the lengths of the sides $y'z$ and yz in the triangles $y'sz$ and ysz respectively. The value of the angle $\angle y'sz$ is smaller than that of the angle $\angle ysz$, and these angles are adjacent to two congruent sides. We deduce that $|y' - z| \leq |y - z|$. Therefore, we obtain

$$|m - m'| \leq (1/2)|y - z|,$$

which is the desired inequality. This completes the proof of Proposition 9.1.4. □

9.2 Variation of local geodesics

In a locally convex space, it is often useful to work with a Busemann neighborhood that is a closed ball rather than with an arbitrary Busemann neighborhood, and for that purpose we record the following three lemmas:

Lemma 9.2.1. *Let X be a locally convex metric space. For every x in X, there exists a positive real number r_0 such that for all $r \leq r_0$, the* (*open or closed*) *ball centered at x and of radius r is a Busemann neighborhood of x.*

Proof. Let $V(x)$ be a Busemann neighborhood of x. If r is small enough, then the (open or closed) ball of center x and radius r is contained in $V(x)$, and therefore the result follows from Corollary 8.3.2. □

Lemma 9.2.2. *Let X be a locally convex metric space. For every x in X, let*

$$r_x = \sup\{r \in]0, \infty] \text{ such that } B(x, r) \text{ is a Busemann space}\}.$$

Then the map $x \mapsto r_x$ is continuous, and we have either $r_x = \infty$ for all x in X or $r_x < \infty$ for all x in X.

Proof. If $r_{x_0} = \infty$ for some x_0 in X, then X is a Busemann space and therefore $r_x = \infty$ for all x in X. Thus, let us suppose that r_x is finite for all x in X and let us prove that the map $x \mapsto r_x$ is continuous in this case. For every x and y in X and for every r in $]0, \infty[$, we have, by the triangle inequality, $B(y, r - |x - y|) \subset B(x, r)$, which implies $|r_x - r_y| \leq |x - y|$. Thus, the map $x \mapsto r_x$ is 1-Lipschitz and therefore continuous. □

Lemma 9.2.3. *Let X be a locally convex metric space and let $\gamma \colon [0, 1] \to X$ be a path. Then there exists a positive real number r such that for all t in* $[0, 1]$*, the closed ball $B(\gamma(t), r)$ of radius r centered at $\gamma(t)$ is a Busemann neighborhood of $\gamma(t)$.*

Proof. We use the compactness of the geodesic segment $\gamma([0, 1])$ and we apply Lemma 9.2.2. □

The following theorem is due to Alexander and Bishop ([1]). All the important results in the next section follow from it.

We recall that if we denote a local geodesic by $\gamma_{x,y}$, then we mean that this local geodesic joins the points x and y.

Theorem 9.2.4 (Variation of affinely reparametrized local geodesics). *Let X be a complete locally convex metric space and let $\gamma = \gamma_{x,y} \colon [0, 1] \to X$ be an affinely reparametrized local geodesic in X. Let r be a positive real number such that for each t in* $[0, 1]$*, the closed ball $B(\gamma(t), r)$ of radius r centered at $\gamma(t)$ is a Busemann neighborhood for $\gamma(t)$. Then for all p and q in X satisfying $|p - x| \leq r/2$ and $|q - y| \leq r/2$, there exists a unique affinely reparametrized local geodesic $\gamma_{p,q} \colon [0, 1] \to X$ satisfying $|\gamma_{p,q}(t) - \gamma(t)| \leq r/2$ for all t in* $[0, 1]$*. Furthermore, if p' and q' are two other points in X satisfying $|p' - x| \leq r/2$ and $|q' - y| \leq r/2$ and if $\gamma_{p',q'} \colon [0, 1] \to X$ is the unique affinely reparametrized local geodesic satisfying $|\gamma_{p',q'}(t) - \gamma(t)| \leq r/2$ for all t in* $[0, 1]$*, then the map $t \mapsto |\gamma_{p,q}(t) - \gamma_{p',q'}(t)|$, defined on the interval* $[0, 1]$*, is convex.*

In fact, we shall prove the following theorem, which is easily seen to be equivalent to Theorem 9.2.4:

Theorem 9.2.4′ (Equivalent version). *Let X be a complete locally convex metric space, let $\gamma = \gamma_{x,y}\colon [0,1] \to X$ be an affinely reparametrized local geodesic in X and let r be a positive real number such that for each t in $[0,1]$, the closed ball $B(\gamma(t), r)$ of radius r centered at $\gamma(t)$ is a Busemann neighborhood of $\gamma(t)$. Then the following property, which we call $\mathcal{P}(L)$, is true for all $L \geq 0$.*

> ***$\mathcal{P}(L)$**: Let $\overline{p}$ and $\overline{q}$ be two points on $\gamma([0,1])$ and let $\gamma_{\overline{p},\overline{q}}\colon [0,1] \to X$ be an affinely reparametrized local geodesic whose image is contained in $\gamma([0,1])$ and whose length is $\leq L$. Then for all p and q in X satisfying $|p-\overline{p}| \leq r/2$ and $|q-\overline{q}| \leq r/2$, there exists a unique affinely reparametrized local geodesic $\gamma_{p,q}\colon [0,1] \to X$ satisfying $|\gamma_{p,q}(t) - \gamma_{\overline{p},\overline{q}}(t)| \leq r/2$ for all t in $[0,1]$. Furthermore, if p' and q' are two other points in X that satisfy $|p'-\overline{p}| \leq r/2$ and $|q'-\overline{q}| \leq r/2$, and if $\gamma_{p',q'}\colon [0,1] \to X$ is the unique affinely reparametrized local geodesic satisfying $|\gamma_{p',q'}(t) - \gamma_{\overline{p},\overline{q}}(t)| \leq r/2$ for all t in $[0,1]$, then the map $t \mapsto |\gamma_{p,q}(t) - \gamma_{p',q'}(t)|$ is convex.*

Proof. We follow the proof of Alexander and Bishop. First, we note that the second part in Property $\mathcal{P}(L)$ follows from the first one. Indeed, since for all t in $[0,1]$ the points $\gamma_{p,q}(t)$ and $\gamma_{p',q'}(t)$ are both at distance $\leq r/2$ from $\gamma_{\overline{p},\overline{q}}(t)$, then they are in the interior of the closed ball $B(\gamma_{\overline{p},\overline{q}}(t), r)$, which is a Busemann neighborhood of the point $\gamma_{\overline{p},\overline{q}}(t)$. Thus, the map $t \mapsto |\gamma_{p,q}(t) - \gamma_{p',q'}(t)|$ is locally convex. Therefore, by Theorem 6.2.16, this map is convex.

Now we prove the first part of Property $\mathcal{P}(L)$. We break up this proof into two steps. In Step 1, we prove that Property $\mathcal{P}(r)$ (that is, Property $\mathcal{P}(L)$ with $L = r$) is satisfied. In Step 2, we prove that if Property $\mathcal{P}(L)$ is satisfied for some positive L, then $\mathcal{P}(3L/2)$ is also satisfied. It is clear that this will prove Theorem 9.2.4′, and consequently, Theorem 9.2.4.

Step 1. Let us consider two points $\overline{p}$ and $\overline{q}$ in $\gamma([0,1])$, and let $\gamma_{\overline{p},\overline{q}}\colon [0,1] \to X$ be the affinely reparametrized local geodesic whose image is contained in $\gamma([0,1])$. Suppose that $L(\gamma_{\overline{p},\overline{q}}) \leq r$ and let $z = \gamma_{\overline{p},\overline{q}}(1/2)$. Thus, we have in particular $|\overline{p} - \overline{q}| \leq r$, $|\overline{p} - z| \leq r/2$ and $|z - \overline{q}| \leq r/2$. Let $B = B(z, r)$ be the closed ball of radius r centered at z. If p and q satisfy $|p - \overline{p}| \leq r/2$ and $|q - \overline{q}| \leq r/2$, then we have

$$|p - z| \leq |p - \overline{p}| + |\overline{p} - z| \leq r/2 + r/2 = r,$$

and

$$|q - z| \leq |q - \overline{q}| + |\overline{q} - z| \leq r/2 + r/2 = r.$$

Therefore, the points p and q are contained in B. This ball, equipped with the metric induced from that of X, is a geodesic metric space. Let $[p, q]$ be a geodesic segment in B joining p and q. Using Corollary 8.2.2, there exists a unique affinely reparametrized local geodesic $\gamma_{p,q}\colon [0,1] \to X$ joining p and q and this implies in particular that the image $\gamma_{p,q}([0,1])$ is the geodesic segment $[p,q]$. The images of the two affinely reparametrized local geodesics $\gamma_{\overline{p},\overline{q}}$ and $\gamma_{p,q}$ are contained in B, which is a Busemann space, and we have $|\gamma_{\overline{p},\overline{q}}(0) - \gamma_{p,q}(0)| \leq r/2$ and $|\gamma_{\overline{p},\overline{q}}(1) - \gamma_{p,q}(1)| \leq r/2$. By Theorem 6.2.16 (Locally convex functions are convex), the map $t \mapsto |\gamma_{\overline{p},\overline{q}}(t) - \gamma_{p,q}(t)|$ is convex. Thus, by Proposition 6.2.14, we have $|\gamma_{\overline{p},\overline{q}}(t) - \gamma_{p,q}(t)| \leq r/2$ for all t in $[0,1]$. This completes the proof of the Step 1.

Step 2. We assume that Property $\mathcal{P}(L)$ is satisfied for some positive L. Let $\overline{p}$ and $\overline{q}$ be two points on $\gamma([0,1])$, let $\gamma_{\overline{p},\overline{q}}\colon [0,1] \to X$ be the affinely reparametrized local geodesic satisfying $\gamma_{\overline{p},\overline{q}}([0,1]) \subset \gamma([0,1])$ and suppose that $L(\gamma_{\overline{p},\overline{q}}) \leq 3L/2$. Let p and q be two points in X satisfying $|p - \overline{p}| \leq r/2$ and $|q - \overline{q}| \leq r/2$ and let p_0 and q_0 be the points on $\gamma([0,1])$ defined by $q_0 = \gamma_{\overline{p},\overline{q}}(1/3)$ and $p_0 = \gamma_{\overline{p},\overline{q}}(2/3)$. Finally, let $\gamma_{\overline{p},p_0}\colon [0,1] \to X$ be the affinely reparametrized local geodesic satisfying $\gamma_{\overline{p},p_0}([0,1]) \subset \gamma([0,1])$ and let $\gamma_{q_0,\overline{q}}\colon [0,1] \to X$ be the affinely reparametrized local geodesic satisfying $\gamma_{q_0,\overline{q}}([0,1]) \subset \gamma([0,1])$.

We define, by induction, two sequences $(p_i)_{i\geq 0}$ and $(q_i)_{i\geq 0}$ in X) such that for all $i \geq 1$, Conditions (9.2.4.1) to (9.2.4.3) below are satisfied:

(9.2.4.1) There exists a unique affinely reparametrized local geodesic $\gamma_{p,p_{i-1}}\colon [0,1] \to X$ such that the map $t \mapsto |\gamma_{p,p_{i-1}}(t) - \gamma_{\overline{p},p_0}(t)|$ is convex and satisfies $|\gamma_{p,p_{i-1}}(t) - \gamma_{\overline{p},p_0}(t)| \leq r/2$ for all t in $[0,1]$, and there exists a unique affinely reparametrized local geodesic $\gamma_{q_{i-1},q}\colon [0,1] \to X$ such that the map $t \mapsto |\gamma_{q_{i-1},q}(t) - \gamma_{q_0,\overline{q}}(t)|$ is convex and satisfies $|\gamma_{q_{i-1},q}(t) - \gamma_{q_0,\overline{q}}(t)| \leq r/2$ for all t in $[0,1]$.

(9.2.4.2) $p_i = \gamma_{q_{i-1},q}(1/2)$ and $q_i = \gamma_{p,p_{i-1}}(1/2)$.

(9.2.4.3) $|p_i - p_0| \leq r/2$ and $|q_i - q_0| \leq r/2$.

The points p_0 and q_0 are already defined. Let us define the points p_1 and q_1.

The lengths of the two paths $\gamma_{\overline{p},p_0}\colon [0,1] \to X$ and $\gamma_{q_0,\overline{q}}\colon [0,1] \to X$ are $\leq L$. Therefore, property $\mathcal{P}(L)$ implies that there exists a unique affinely reparametrized local geodesic γ_{p,p_0} and a unique affinely reparametrized local geodesic $\gamma_{q_0,q}$ satisfying, for all t in $[0,1]$,

(9.2.4.4) $$|\gamma_{p,p_0}(t) - \gamma_{\overline{p},p_0}(t)| \leq r/2 \text{ and } |\gamma_{q_0,q}(t) - \gamma_{q_0,\overline{q}}(t)| \leq r/2.$$

We set $p_1 = \gamma_{q_0,q}(1/2)$ and $q_1 = \gamma_{p,p_0}(1/2)$ (Figure 9.6). Relations (9.2.4.1) and (9.2.4.2) are therefore satisfied for $i = 1$. The convexity of the functions

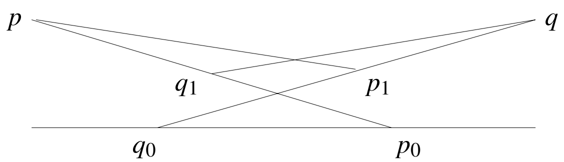


Figure 9.6

$t \mapsto |\gamma_{p,p_0}(t) - \gamma_{\overline{p},}\ p_0(t)|$ and $t \mapsto |\gamma_{q_0,q}(t) - \gamma_{q_0,\overline{q}}(t)|$, together with (9.2.4.4), imply that (9.2.4.3) is satisfied for $i = 1$. Thus, the points p_1 and q_1 satisfy the three required conditions.

Now suppose that the points p_{i-1} and q_{i-1} have been defined for some integer $i \geq 2$ and let us define the points p_i and q_i. We set $p_i = \gamma_{q_{i-1},q}(1/2)$ and $q_i = \gamma_{p,p_{i-1}}(1/2)$. Since $p_0 = \gamma_{q_0,\overline{q}}(1/2)$, we have, by convexity, $|p_i - p_0| \leq r/2$. In the same way, we have $|q_i - q_0| \leq r/2$. By Property $\mathcal{P}(L)$, there exists a unique affinely reparametrized local geodesic $\gamma_{p,p_i}\colon [0,1] \to X$ such that the map $t \mapsto |\gamma_{p,p_i}(t) - \gamma_{\overline{p},p_0}(t)|$ is convex and satisfies $|\gamma_{p,p_i}(t) - \gamma_{\overline{p},p_0}(t)| \leq r/2$ for all t in $[0,1]$, and there exists a unique affinely reparametrized local geodesic $\gamma_{q_i,q}\colon [0,1] \to X$ such that the map $t \mapsto |\gamma_{q_i,q}(t) - \gamma_{q_0,\overline{q}}(t)|$ is convex and satisfies $|\gamma_{q_i,q}(t) - \gamma_{q_0,\overline{q}}(t)| \leq r/2$ for all t in $[0,1]$.

This completes the induction for the existence of the sequences $(p_i)_{i\geq 0}$ and $(q_i)_{i\geq 0}$.

To continue Step 2 of the proof of Theorem 9.2.4′, we need the following

Lemma 9.2.5. *For every integer $i \geq 1$ and for every t in* $[0,1]$, *we have*

(9.2.5.1)
$$|\gamma_{p,p_{i-1}}(t) - \gamma_{p,p_i}(t)| \leq \frac{r}{2^i} \quad \textit{and} \quad |\gamma_{q_{i-1},q}(t) - \gamma_{q_i,q}(t)| \leq \frac{r}{2^i}.$$

Proof. We prove this lemma by induction. We start with the case $i = 1$. We have

$$|\gamma_{p,p_0}(0) - \gamma_{p,p_1}(0)| = |p - p| = 0$$

and

$$|\gamma_{p,p_0}(1) - \gamma_{p,p_1}(1)| = |p_0 - p_1| \leq r/2.$$

Therefore, by convexity of the map $t \mapsto |\gamma_{p,p_0}(t) - \gamma_{p,p_1}(t)|$, we have

$$|\gamma_{p,p_0}(t) - \gamma_{p,p_1}(t)| \leq r/2$$

for all t in $[0,1]$.

An analogous reasoning shows that

$$|\gamma_{q_0,q}(t) - \gamma_{q_1,q}(t)| \leq r/2$$

for all t in $[0,1]$.

This proves (9.2.5.1) for $i = 1$.

Now, let us suppose, as an induction hypothesis, that (9.2.5.1) is satisfied for some integer $i \geq 1$. From (9.2.4.2), we have, for all $i \geq 1$, $\gamma_{p,p_{i-1}}(1/2) = q_i$. The induction hypothesis (applied with $t = 1$) and the convexity of the map $t \mapsto |\gamma_{p,p_{i-1}}(t) - \gamma_{p,p_i}(t)|$ imply therefore (by taking $t = 1/2$)

$$|q_i - q_{i+1}| \leq \frac{1}{2}\frac{r}{2^i} = \frac{r}{2^{i+1}}.$$

In the same manner, we obtain

$$|p_i - p_{i+1}| \leq \frac{1}{2}\frac{r}{2^i} = \frac{r}{2^{i+1}}.$$

Now using the convexity of the map $t \mapsto |\gamma_{q_i,q}(t) - \gamma_{q_{i+1},q}(t)|$, we obtain

$$|\gamma_{q_i,q}(t) - \gamma_{q_{i+1},q}(t)| \leq \frac{r}{2^{i+1}}$$

for all t in [0, 1].

In the same way, we have

$$|\gamma_{p,p_i}(t) - \gamma_{p,p_{i+1}}(t)| \leq \frac{r}{2^{i+1}}$$

for all t in [0, 1].

This completes the induction step for (9.2.5.1), and the proof of Lemma 9.2.5. □

Now we continue the proof of Theorem 9.2.4′.

Relations (9.2.5.1) show that the sequences $\big(\gamma_{p,p_i}(t)\big)_{i\geq 0}$ and $\big(\gamma_{q_i,q}(t)\big)_{i\geq 0}$ are Cauchy sequences, uniformly in t. Therefore, the sequence of paths $(\gamma_{p,p_i})_{i\geq 1}$ converges uniformly to a path γ_{p,p_∞}, where $p_\infty = \lim_{i\to\infty} p_i$. In the same way, the sequence of paths $(\gamma_{q_i,q})_{i\geq 1}$ converges uniformly to a path $\gamma_{q_\infty,q}$, where $q_\infty = \lim_{i\to\infty} q_i$.

By taking limits, we see that the paths γ_{p,p_∞} and $\gamma_{q_\infty,q}$ are affinely reparametrized local geodesics, and that they satisfy respectively

$$|\gamma_{p,p_\infty}(t) - \gamma_{\overline{p},p_0}(t)| \leq r/2$$

and

$$|\gamma_{q_\infty,q}(t) - \gamma_{q_0,\overline{q}}(t)| \leq r/2.$$

Again, by taking limits, we obtain $\gamma_{p,p_\infty}(1/2) = q_\infty$ and $\gamma_{q_\infty,q}(1/2) = p_\infty$.

The map $t \mapsto |\gamma_{p,p_{i-1}}(1/2+t) - \gamma_{q_{i-1},q}(t)|$, defined on [1/2, 1], is convex. Its values for $t = 1/2$ and $t = 1$ are bounded above by $r/2^i$. Thus, we have, for all t in [1/2, 1],

$$|\gamma_{p,p_{i-1}}(1/2+t) - \gamma_{q_{i-1},q}(t)| \leq r/2^i.$$

Hence, for every t in [1/2, 1], the sequences $\big(\gamma_{p,p_{i-1}}(1/2+t)\big)_{i\geq 1}$ and $\big(\gamma_{q_{i-1},q}(t)\big)_{i\geq 1}$ in X converge to the same point as $i \to \infty$. We deduce that the images of the two affinely reparametrized local geodesics γ_{p,p_∞} and $\gamma_{q_\infty,q}$ have in common the image of the unique affinely reparametrized local geodesic $\gamma_{p_\infty,q_\infty} : [0, 1] \to X$ satisfying $|\gamma_{p_\infty,q_\infty}(t) - \gamma_{q_0,p_0}(t)| \leq r/2$ for all t in [0, 1]. Therefore, the union of these two images is the image of an affinely reparametrized local geodesic joining p and q satisfying the required properties. This completes the proof of Theorem 9.2.4′, and therefore that of Theorem 9.2.4. □

We deduce from Theorem 9.2.4 a result on the convexity of neighborhoods of geodesic segments in Busemann spaces. To state this result, we recall the following definition of Busemann's ([28] p. 243) that we already used in special cases (Chapter 5).

Definition 9.2.6 (Capsule). Let X be a metric space, let r be a nonnegative real number and let $[x, y]$ be a geodesic segment in X. Then, the *capsule of radius r and axis* $[x, y]$ is the set

$$C([x, y], r) = \{z \in X, d(z, [x, y]) \le r\}.$$

The notion of capsule generalizes the notion of closed ball, since when the segment $[x, y]$ is reduced to a point, the capsule $C([x, y], r)$ is the closed ball of center x and radius r. Thus, the following proposition generalizes Proposition 8.3.2 which says that in a Busemann space, closed balls are convex.

Proposition 9.2.7 (Capsules in Busemann spaces are convex). *Let X be a complete Busemann space. Then, for every geodesic segment $[x, y]$ and for every nonnegative real number r, the capsule $C([x, y], r)$ is convex.*

Proof. Let $\gamma_{x,y}\colon [0, 1] \to X$ be an affinely reparametrized geodesic whose image is the segment $[x, y]$. For every t in $[0, 1]$ and for every $r > 0$, the open ball $B(\gamma(t), 2r)$ is a Busemann neighborhood of the point $\gamma(t)$ (we use Corollary 8.3.2 which says that any open ball in a Busemann space, equipped with the induced metric, is a Busemann space). By applying Theorem 9.2.4 to any subsegment of $[x, y]$, we see that for every p and q in the capsule $C([x, y], r)$, there exists an affinely reparametrized local geodesic $\gamma_{p,q}\colon [0, 1] \to X$ whose image is contained in this capsule. By the existence of geodesics and uniqueness (up to parametrization) of local geodesics joining two points in a Busemann space (Corollary 8.2.2), the path $\gamma_{p,q}$ is an affinely reparametrized geodesic. Therefore, its image is a geodesic segment contained in the capsule. This proves Proposition 9.2.7. □

9.3 The universal covering of a locally convex metric space

The following definition is due to Alexander and Bishop (see [1]). It generalizes the notion of exponential map in the context of Riemannian geometry, whose domain of definition is the tangent space, at each point of a Riemannian manifold.

Definition 9.3.1 (Tangent space and exponential map).[3] Let X be a geodesic metric space and let x be a point in X. The *tangent space at* x is the set G_x of affinely reparametrized local geodesics $\gamma\colon [0, 1] \to X$ that start at x (that is, satisfying $\gamma(0) = x$). The *exponential map at* x, $\exp_x\colon G_x \to X$, is then defined by setting $\exp_x(\gamma) = \gamma(1)$ for all γ in G_x.

[3] This definition makes sense in any metric space, but of course, if the space is not locally geodesic, the definition should be replaced by a discrete version.

For every x in X, we equip G_x with the metric d defined by

$$d(\gamma_1, \gamma_2) = \sup_{t\in[0,1]} |\gamma_1(t) - \gamma_2(t)|_X$$

for all γ_1 and γ_2 in G_x. We shall refer to this metric as the *metric of uniform convergence* on G_x.

Notice that if X is a locally convex metric space, then X is connected, locally arcwise connected and semi-locally simply connected, and therefore it makes sense to talk about the universal covering of X. We now prove the following

Theorem 9.3.2 (Gromov–Alexander–Bishop: The exponential map is the universal covering map). *Let X be a geodesic, complete, locally compact and locally convex metric space. Then for every x in X, the exponential map* $\exp_x \colon G_x \to X$ *is a universal covering map.*

Proof. Let x be a point in X. We first prove that the exponential map $\exp_x$ is a local isometry. Let γ be an element of G_x and let us prove that there exists a neighborhood $V(\gamma)$ of γ in G_x such that the restriction of $\exp_x$ to $V(\gamma)$ sends this neighborhood isometrically onto a neighborhood of the point $\exp_x(\gamma)$ in X.

Let r be a positive real number such that for every t in $[0, 1]$, the closed ball $B(\gamma(t), r)$ of radius r centered at $\gamma(t)$ is a Busemann neighborhood for $\gamma(t)$ (Lemma 9.2.3) and let $V(\gamma)$ be the set of elements $\eta\colon [0, 1] \to X$ in G_x satisfying $|\eta-\gamma|_{G_x} \leq r/2$. Then $V(\gamma)$ is a neighborhood of γ in G_x. For every η_1 and η_2 in $V(\gamma)$, let $d_{\eta_1,\eta_2}\colon [0, 1] \to \mathbb{R}$ be the map $t \mapsto |\eta_1(t)-\eta_2(t)|$. This map is convex, and it satisfies $d_{\eta_1,\eta_2}(0) = 0$ and $d_{\eta_1,\eta_2}(t) \geq 0$ for all t in $[0, 1]$. Therefore, by Corollary 6.2.12, it is increasing. This implies

$$\begin{aligned} |\eta_1 - \eta_2|_{G_x} &= \sup_{t\in[0,1]} |\eta_1(t) - \eta_2(t)|_X \\ &= |\eta_1(1) - \eta_2(1)|_X \\ &= |\exp_x(\eta_1) - \exp_x(\eta_2)|_X. \end{aligned}$$

Thus, the exponential map restricted to $V(\gamma)$ is distance-preserving. In particular, it is continuous and injective. It is also surjective, since X is a geodesic metric space. Furthermore, Theorem 9.2.4 implies that the set $\exp_x\big(V(\gamma)\big)$ is a neighborhood of $\exp(\gamma)$ in G_x. Thus, the map $\exp_x$ is a local isometry.

Now, let us prove that the map $\exp_x\colon G_x \to X$ is a covering map.

The space X, being locally convex, is locally uniquely geodesic. We have seen that the map $\exp_x$ is a local isometry. We would like to apply Theorem 3.5.4 but we cannot do so because the space G_x, equipped with the metric d of uniform convergence, is not necessarily a length space. (As Alexander and Bishop note in [1], p. 317, an example of a space X that satisfies the hypotheses of Theorem 9.3.2 but for which d is not a length metric, is obtained by taking X to be a circle in the Euclidean plane.) Instead, we apply Theorem 3.5.4 to the metric space (G_x, d_ℓ) where d_ℓ is the length metric associated to d (Definition 2.1.6). To see that this is possible, we show that the metric d_ℓ satisfies Properties (9.3.2.1) and (9.3.2.2) below:

(9.3.2.1) *The metric d_ℓ is complete.*

Indeed, the metric d (which is the metric of uniform convergence) is complete, and we have $d_\ell \geq d$. (This follows immediately from the definition of the length metric associated to a metric). Thus, a Cauchy sequence for d_ℓ is also a Cauchy sequence for d. A Cauchy sequence for the metric d is convergent. Therefore, this sequence converges also for d_ℓ.

(9.3.2.2) *The map* $\exp_x : (G_x, d_\ell) \to X$ *is a local isometry.*

Indeed, since X is locally convex, (G_x, d) is also locally convex. In particular, (G_x, d) is locally a length space. Therefore, the metric d_ℓ coincides locally with d. (In the example we mentioned above where X is a circle, d and d_ℓ do not coincide globally.) In particular, (G_x, d_ℓ) is locally compact and locally uniquely geodesic, and the map $\exp_x : (G_x, d_\ell)$ is, like the map $\exp_x : (G_x, d) \to X$, a local isometry.

Therefore, by Theorem 3.5.4, the map $\exp_x : G_x \to X$ is a covering map.

Finally we prove that for all x in X, the space G_x is simply connected. This will imply that the map $\exp_x : G_x \to X$ is the universal covering map.

We show that the space G_x is contractible. To see this, we use the map $H : [0, 1] \times G_x \to G_x$ where for all u in $[0, 1]$, $H(u, \gamma)$ is the element of G_x defined by $t \mapsto \gamma(ut)$. We have, for all γ in G_x, $H(0, \gamma) = \gamma_0$ where $\gamma_0(t) = x$ for all t in $[0, 1]$, and $H(1, \gamma) = \gamma$. Theorem 9.2.4 implies that any element $\gamma_{x,y}$ in G_x varies continuously as a function of y, with respect to the topology of uniform convergence on G_x, and this proves that the map H is continuous. Therefore, H is a retraction. This completes the proof of Theorem 9.3.2. □

From Theorem 9.3.2, we deduce the following

Corollary 9.3.3. *Let X be a complete geodesic locally convex and locally compact metric space. In each homotopy class of paths in X with fixed endpoints that contains a rectifiable path, there exists a local geodesic, and this local geodesic is unique up to reparametrization.*

Proof. The existence of the local geodesic follows from Proposition 2.4.11. The uniqueness follows from the fact that the exponential map is a universal covering map (Theorem 9.3.2). □

We end this chapter with the following theorem (cf. [1], p. 309 where this theorem is attributed to Gromov).

Theorem 9.3.4 (Gromov). *Let X be a simply connected geodesic, complete, locally compact and locally convex length space. Then X is a Busemann space.*

Proof. Let x be a point in X. Since X is simply connected, Theorem 9.3.2 implies that the covering map $\exp_x : G_x \to X$ is a homeomorphism, for all x in X. In particular, for any two points in X, there exists an affinely reparametrized local geodesic that joins them that is unique up to reparametrization. Since the space X is geodesic (Theorems 2.1.15 and 2.4.6), this affinely reparametrized local geodesic is an affinely reparametrized geodesic. Thus, given two arbitrary points in X, there exists an affinely reparametrized geodesic that joins them, and this affinely reparametrized geodesic is unique up to reparametrization.

Now let us consider three geodesic segments $[x, y]$, $[x, z]$ and $[y, z]$ in X and let m and m' be respectively the midpoints of $[x, y]$ and $[x, z]$. For each t in $[0, 1]$, let $\gamma(t)$ be the point on $[y, z]$ defined by $\gamma(t) = (1 - t)y + tz$ (we are using the natural parametrization of a geodesic segment), and let $\eta_t : [0, 1] \to X$ be the unique affinely reparametrized geodesic satisfying $\eta_t(0) = x$ and $\eta_t(1) = \gamma(t)$. In particular, the image of η_0 is the geodesic segment $[x, y]$ and the image of η_1 is the geodesic segment $[x, z]$.

Let $R = \max\{|x - y|, |x - z|\}$. For every t in $[0, 1]$, the image of η_t is contained in the closed ball $B(x, R)$.

Let us consider again the map on X defined by

$$x \mapsto r_x = \sup\{r \in]0, \infty] \text{ such that } B(x, r) \text{ is convex}\}.$$

By Lemma 9.2.2, this map is continuous. Since the closed ball $B(x, r)$ is compact (Theorem 2.1.15), there exists a real number $r > 0$ such that for all $x \in B(x, R)$, we have $r_x > r$.

Let $(t_i)_{i=0,\dots,n}$ be a subdivision of $[0, 1]$ such that for all $i = 0, \dots, n$, we have $|\gamma(t_i) - \gamma(t_{i+1})| < r/2$. By Theorem 9.2.4, for all $i = 0, 1, \dots, n - 1$, the map $t \mapsto |\eta_{t_i}(t) - \eta_{t_{i+1}}(t)|$ is convex. Thus, we have

$$\begin{aligned} |\eta_{t_i}(1/2) - \eta_{t_{i+1}}(1/2)| &\leq (1/2)|\eta_{t_i}(1) - \eta_{t_{i+1}}(1)| \\ &= (1/2)|\gamma(t_i) - \gamma(t_{i+1})|. \end{aligned}$$

Applying n times the triangle inequality, we obtain:

$$\begin{aligned} |m - m'| &= |\eta_0(1/2) - \eta_1(1/2)| \\ &\leq \sum_{i=0}^{n-1} |\eta_{t_i}(1/2) - \eta_{t_{i+1}}(1/2)| \\ &\leq (1/2) \sum_{i=0}^{n-1} |\eta_{t_i}(1) - \eta_{t_{i+1}}(1)| \\ &= (1/2)|y - z|. \end{aligned}$$

This completes the proof of Theorem 9.3.4. □

Notes on Chapter 9

Theorem 9.2.4 and its proof follow Theorem 2 of [1]. An early version of Corollary 9.3.3 is due to Busemann ([28] p. 288), where the setting is that of straight G-spaces. This result generalizes a result of Hadamard for surfaces equipped with a metric of nonpositive curvature that we mentioned in the introduction (cf. [62], Part III, Theorem 31). Theorem 9.3.2 is Theorem 1 in [1], where this result is attributed to Gromov, and where it is proved without the assumption that X is locally compact. In the proof that we give here, we rely on Theorem 3.5.4, which requires local compactness. A proof of Theorem 9.3.2, without the local compactness condition is contained in [20] (Theorem 3.4 p. 193). Another version of this result is contained in [7].

Convexity of capsules. Busemann introduced in [28] a notion of a "space in which capsules with a small radius are convex". Such a property is satisfied by complete Busemann spaces, and in fact, in a complete Busemann space, all capsules are convex (Proposition 9.2.7 above). In [28], Theorem 41.6, Busemann shows that a Riemannian manifold has nonpositive curvature if and only if in that space capsules of small radii are convex. Busemann shows in [28] §36.16 that in a G-space, this property on capsules is equivalent to the fact that each point has a neighborhood such that for each geodesic segment T and for each geodesic segment $[x_0, x_1]$ in that neighborhood, the function $f_T(t) = \mathrm{dist}(x_t, T)$ defined on [0, 1], is peakless. Here, for all t in [0, 1], x_t is as usual the point on $[x_0, x_1]$ satisfying $|x_0 - x_t| = t|x_0 - x_1|$. Of course, in the case where X is a Busemann space, the function f_T is convex. For the definition of a peekless function, see Definition 6.2.10.

The Cartan–Hadamard Theorem. In [36], Notes III & IV, Cartan proves that if M is a proper Riemannian manifold of dimension n, then the set $\mathcal{E}$ of local geodesics starting at some fixed point x_0 of M is a simply connected covering of M (and therefore it is the universal covering of M), the covering map from $\mathcal{E}$ to M being the one that assigns to each local geodesic starting at x_0 its endpoint. The main point in Cartan's argument is the fact that each point in M can be joined to x_0 by a local geodesic (that is, that the map $\mathcal{E} \to M$ is surjective); cf. [36] p. 346. On p. 360 of the same book, Cartan has another proof of this fact, based on the theorem of Hopf–Rinow (for which he gives a proof), that is, the fact that for any two points in the Riemannian manifold M, there is a geodesic segment joining them whose length is equal to the distance between these points. Cartan also proves that in the case where M is nonpositively curved, $\mathcal{E}$ is homeomorphic to $\mathbb{R}^n$. His arguments are based on Hadamard's ideas contained in [62], which justifies the name "Cartan–Hadamard Theorem". As an application of that result, Cartan considers the case of a manifold X homeomorphic to $S^2 \times]0, 1[$. This space cannot be equipped with a Riemannian metric of nonpositive curvature that is proper (that is, where the two ends are at infinity), since this space is simply connected but not contractible ([36] p. 348).

Chapter 10

Asymptotic rays and the visual boundary

Introduction

Given two geodesic rays $r_1\colon [0,\infty[\to X$ and $r_2\colon [0,\infty[\to X$ in a metric space X, we say that they are *asymptotic* if there exists a real number α such that for all $t \geq 0$, we have

$$|r_1(t) - r_2(t)| \leq \alpha.$$

Equivalently, two geodesic rays in a metric space X are asymptotic if and only if their images are at finite Hausdorff distance.

Asymptoticity of geodesic rays is one possible (and may be the most useful) generalization to arbitrary metric spaces of parallelism of geodesic rays in Euclidean space.

A first version of this notion of asymptoticity, in the setting of surfaces of nonpositive curvature, has been considered by Hadamard in his paper [62]. If X is a Busemann space, two geodesic rays starting at the same point are asymptotic if and only if they are equal.

In this chapter, for each geodesic metric space X and for each point p in X, we consider the set $\mathcal{R}_pX$ of geodesic rays in X starting at p, and we equip it with the topology of uniform convergence on compact sets. We then associate to the point p a space ∂_pX, which is better adapted to play the role of a "boundary" of X than the space $\mathcal{R}_pX$. This is the quotient space of $\mathcal{R}_p(X)$ by the equivalence relation that identifies two geodesic rays starting at p whenever these rays are asymptotic. The space ∂_pX is equipped with the quotient topology induced from that of $\mathcal{R}_p(X)$ and it is called the *visual boundary* of X at p. One can try to make the relations between the families $(\mathcal{R}_pX)_{p\in X}$ and $(\partial_pX)_{p\in X}$ and the family of tangent spaces associated to a geodesic metric space, which we encountered in Chapter 9.

In the case where X is a Busemann space, each equivalence class of geodesic rays starting at any given point is reduced to one element, and therefore there is a canonical identification $\mathcal{R}_pX \simeq \partial_pX$, for every p in X. In the case where X is a proper Busemann space, ∂_pX is compact and there is a canonical bijection between the spaces ∂_pX and ∂_qX, for any p and q in X. A study of the boundary spaces ∂_pX has been started by P. Hotchkiss in [65], where he proves that the canonical map $\partial_pX \to \partial_qX$ is a homeomorphism.

The outline of this chapter is as follows.

In Section 1, we consider asymptotic geodesic rays in metric spaces and we study in some detail the case of Busemann spaces.

In Section 2, we consider the space $\mathcal{R}_p X$ of geodesic rays in a metric space X starting at a point p and we define the visual boundary $\partial_p X$ at p. In the case where X is a Busemann space, we show that $\partial_p X \simeq \mathcal{R}_p X$ and that for any p and q in X, there is a natural bijection $\Phi_{p,q} \colon \partial_p X \to \partial_q X$. We define a topology on the union $X \cup \partial_p X$. We end this section with a few remarks on the topology of the spaces $\mathcal{R}_p X$ and $\partial_p X$.

10.1 Asymptotic rays

Let us start by recalling that a *geodesic ray* in a metric space X is a distance-preserving map $r \colon [0, \infty[\to X$. We say that $r(0)$ is the *origin* of r, and that r *starts at* $r(0)$.

We also recall that the *diameter* of a metric space X is the element of $[0, \infty[\cup \{\infty\}$ defined as

$$\mathrm{diam}(X) = \sup\{|x - y|,\ x \text{ in } y \in X\}.$$

It is clear that if X has finite diameter, then there is no geodesic ray in X. Conversely, we have the following

Proposition 10.1.1. *Let X be a complete geodesic metric space of infinite diameter. Then every point in X is the origin of some geodesic ray.*

Proof. Since $\mathrm{diam}(X) = \infty$, we can find, for every nonnegative integer n, two points y_n and z_n in X satisfying $|y_n - z_n| \geq n$. Let x be any point in X. From the triangle inequality, we have either $|x - y_n| \geq n/2$ or $|x - z_n| \geq n/2$. Let us set, for each nonnegative integer n,

$$x_n = \begin{cases} y_n & \text{if } |x - y_n| \geq n/2, \\ z_n & \text{if } |x - y_n| < n/2. \end{cases}$$

We have $|x - x_n| \to \infty$ as $n \to \infty$. Since X is geodesic, we can find, for each $n \geq 0$, a geodesic path $\gamma_n^* \colon [0, |x - x_n|] \to X$ joining x and x_n. We extend γ_n^* to a map $\gamma_n \colon [0, \infty[\to X$ by setting

$$\gamma_n(t) = \begin{cases} \gamma_n^*(t) & \text{if } 0 \leq t \leq |x - x_n|, \\ x_n & \text{if } |x - x_n| \leq t < \infty. \end{cases}$$

The sequence of maps (γ_n) is equicontinuous, since each of these maps is 1-Lipschitz. Furthermore, for each $t \geq 0$, the set $\{\gamma_n(t),\ n \geq 0\}$ is contained in the closed ball of center x and radius t, and therefore it is bounded. Since $\mathbb{R}$ is separable and X complete, Ascoli's theorem (Theorem 1.4.9) implies that the sequence (γ_n) has a subsequence that converges uniformly on compact subsets of $\mathbb{R}$. From the continuity of the distance function, it is easy to see, as in the proof of Proposition 2.3.1, that the limit of this subsequence is a geodesic ray starting at x. □

Definition 10.1.2 (Asymptotic geodesic rays). Let X be a metric space and let $r_1\colon [0,\infty[\to X$ and $r_2\colon [0,\infty[\to X$ be two geodesic rays. We say that r_1 *is asymptotic to* r_2 if there exists a real number α such that $|r_1(t)-r_2(t)|\leq\alpha$ for all $t\geq 0$.

It is clear that the relation of being asymptotic is an equivalence relation on the set of geodesic rays.

If $r\colon [0,\infty[\to X$ and $r'\colon [0,\infty[\to X$ are two geodesic rays in X, than we say that r' is a *subray of* r if there exists a real number t_0 such that $r'(t)=r(t_0+t)$ for all $t\geq 0$. It is clear that if r' is a subray of r, then r and r' are asymptotic.

Examples 10.1.3 (Asymptotic geodesic rays).

(i) Let us first give an example of a space in which there are several asymptotic rays starting at some point. Consider the space $\mathbb{E}^3$ equipped with a coordinate system x,y,z and let $X\subset\mathbb{E}^3$ be the subset obtained by rotating around the z-axis the subset of the xz-plane defined as the union of the half-line $\{x=-1, z\geq 0\}$ with the affine segment joining the point $(0,-1)$ to the point $(-1,0)$ (see Figure 10.1). We equip X with the length metric associated to the metric induced from its inclusion in $\mathbb{E}^3$. In this length space, the union of the half-line $\{x=-1, z\geq 0\}$ with the segment joining the point $(0,-1)$ to the points $(-1,0)$ is the image of a geodesic ray r starting at $(0,-1)$, and any ray obtained by composing the ray r with a rotation around the z-axis is asymptotic to r.

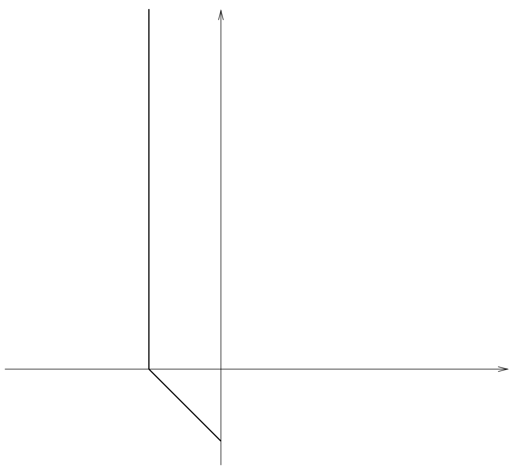


Figure 10.1. The surface X in $\mathbb{E}^3$ referred to in Example 10.1.3 (i) is obtained by rotating around the vertical axis the curve drawn here.

(ii) If $X=\mathbb{R}$, then each point in X is the origin of exactly two geodesic rays, and these two rays are not asymptotic

(iii) The same property holds if X is the Euclidean cylindre $S^1 \times \mathbb{R}$: each point in X is the origin of exactly two geodesic rays, and these rays are not asymptotic.

(iv) *Euclidean space.* If $r_1 \colon [0,\infty[\to \mathbb{E}^n$ and $r_2 \colon [0,\infty[\to \mathbb{E}^n$ are two geodesic rays in Euclidean n-space $\mathbb{E}^n$, then r_1 and r_2 are asymptotic if and only if one of the following two properties holds:

• r_2 is a subray of r_1 or r_1 is a subray of r_2;

• r_1 and r_2 have disjoint images, these images are contained in a 2-dimensional affine plane, and in that plane, the images are contained in images of parallel Euclidean straight lines and have the same direction.

(v) *Hyperbolic space.* In hyperbolic space $\mathbb{H}^n$, two geodesic rays $r_1 \colon [0,\infty[\to \mathbb{H}^n$ and $r_2 \colon [0,\infty[\to \mathbb{H}^n$ are asymptotic if and only if the limits $r_1(\infty)$ and $r_2(\infty)$ are equal as points of the boundary S^{n-1}. (This is best visualized in the open ball model B^n of $\mathbb{H}^n$.)

(vi) $\mathbb{R}$-*tree.* Let T be an $\mathbb{R}$-tree. Then, two geodesic rays $r_1 \colon [0,\infty[\to T$ and $r_2 \colon [0,\infty[\to T$ are asymptotic if and only if their images coincide up to a compact segment, that is, if and only if there exists $t_1 \geq 0$ and $t_2 \geq 0$ such that $r_1([t_1,\infty[) = r_2([t_2,\infty[)$.

On the set of geodesic rays, we can compare the relation of being asymptotic to the relation of having images at finite Hausdorff distance. We start with a few general observations on subsets that are at finite Hausdorff distance.

From the definition of the Hausdorff distance $d_{\mathcal{H}}$ (cf. Chapter 4), it follows that if A and B are subsets of a metric space X, then $d_{\mathcal{H}}(A,B) < \infty$ if and only if there exists $\epsilon > 0$ such that $B \subset N(A,\epsilon)$ and $A \subset N(B,\epsilon)$. It is clear that the relation $d_{\mathcal{H}}(A,B) < \infty$ is an equivalence relation between subsets of X, and that any two bounded subsets of X are equivalent with respect to this relation.

Proposition 10.1.4. *Let X be a metric space and let $r_1 \colon [0,\infty[\to X$ and $r_2 \colon [0,\infty[\to X$ be two geodesic rays in X. Then, r_1 and r_2 are asymptotic if and only if their images are at finite Hausdorff distance.*

Proof. We prove the nontrivial part of the statement, i.e., if the images are at finite Hausdorff distance, then the geodesic rays are asymptotic. Let $\kappa = d_{\mathcal{H}}(\mathrm{Im}(r_1),\mathrm{Im}(r_2))$ and let us show that if $\kappa < \infty$, then $|r_1(t) - r_2(t)| \leq 2\kappa + |r_1(0) - r_2(0)|$ for all $t \geq 0$.

For each $t \geq 0$, consider the projection $p_{\mathrm{Im}(r_2)}(r_1(t))$ of $r_1(t)$ on the closed set $\mathrm{Im}(r_2)$. Then, $p_{\mathrm{Im}(r_2)}(r_1(t)) = r_2(t')$ for some $t' \geq 0$ and we have

$$\begin{aligned} t' = |r_2(t') - r_2(0)| &\leq |r_2(t') - r_1(t)| + |r_1(t) - r_1(0)| + |r_1(0) - r_2(0)| \\ &\leq \kappa + t + |r_1(0) - r_2(0)|. \end{aligned}$$

On the other hand, we have

$$\begin{aligned} t &= |r_1(t) - r_1(0)| \\ &\leq |r_1(t) - r_2(t')| + |r_2(t') - r_2(0)| + |r_2(0) - r_1(0)| \\ &\leq \kappa + t' + |r_2(0) - r_1(0)|, \end{aligned}$$

which gives $t' \geq t - \kappa - |r_2(0) - r_1(0)|$.

Thus, we obtain

$$-\kappa - |r_1(0) - r_2(0)| \leq t' - t \leq \kappa + |r_1(0) - r_2(0)|$$

or, equivalently,

$$|t' - t| \leq \kappa + |r_1(0) - r_2(0)|.$$

Therefore we have, for all $t \geq 0$,

$$\begin{aligned} |r_1(t) - r_2(t)| &\leq |r_1(t) - r_2(t')| + |r_2(t') - r_2(t)| \\ &\leq \kappa + |t' - t| \\ &\leq 2\kappa + |r_1(0) - r_2(0)|. \end{aligned}$$

This completes the proof of Proposition 10.1.4. □

We are particularly interested in images of geodesic rays in Busemann spaces, and concerning such rays we need the following:

Proposition 10.1.5. *Let X be a Busemann space, let p be a point in X and let $r_1\colon [0,\infty[\to X$ and $r_2\colon [0,\infty[\to X$ be two distinct geodesic rays starting at p. Then, the map $d_{r_1,r_2}\colon [0,\infty[\to \mathbb{R}$ defined by*

$$d_{r_1,r_2}(t) = |r_1(t) - r_2(t)|$$

is increasing and satisfies $\lim_{t\to\infty} d_{r_1,r_2}(t) = \infty$.

Proof. Since X is a Busemann space, for any fixed $T \geq 0$, the restriction of d_{r_1,r_2} to the interval $[0, T]$ is convex. Furthermore, this map takes only nonnegative values, and it takes the value 0 at $t = 0$. Therefore, it is increasing (Corollary 6.2.12). This implies that the map d_{r_1,r_2} is also increasing and convex. Since the geodesic rays r_1 and r_2 are distinct and since they start at the same point, the map d_{r_1,r_2} is not constant. Therefore, by Proposition 6.2.3, we have $\lim_{t\to\infty} d_{r_1,r_2}(t) = \infty$. □

From the preceding results, we deduce the following

Proposition 10.1.6. *Let X be a Busemann space and let $r_1\colon [0,\infty[\to X$ and $r_2\colon [0,\infty[\to X$ be two geodesic rays starting at the same point. Then, the following three properties are equivalent:*

(i) $r_1 = r_2$;

(ii) r_1 *and* r_2 *are asymptotic;*

(iii) $d_{\mathcal{H}}(\mathrm{Im}(r_1), \mathrm{Im}(r_2)) < \infty$.

Proof. Implication (i) ⇒ (ii) is trivial. By Proposition 10.1.4, we have (ii) $\iff$ (iii). Finally, (ii) ⇒ (i) follows from Proposition 10.1.5. □

We already mentioned that the relation of being asymptotic (or, equivalently, the relation of having their images at finite Hausdorff distance) on the set of geodesic rays is a possible generalization the relation of being parallel between geodesic rays in Euclidean space. Therefore, the next proposition shows that, in some sense, for any geodesic ray in a proper Busemann space, there exists a unique "parallel" geodesic ray starting at any point.

Proposition 10.1.7. *Let X be a proper Busemann space, let q be a point in X and let $r\colon [0, \infty[\to X$ be a geodesic ray. Then, there exists a unique geodesic ray r' starting at q that is asymptotic to r (or, equivalently, whose image is at finite Hausdorff distance from that of r).*

Proof. Uniqueness follows from Proposition 10.1.6 and from the transitivity of the relation of being at finite Hausdorff distance. For existence, we use the same construction as in the proof of Proposition 10.1.1. For all $n \geq 0$, we let $\gamma_n\colon [0, |q - r(n)|] \to X$ be the geodesic path joining q to $r(n)$ and $r_n\colon [0, \infty[\to X$ the map defined by

$$r_n(t) = \begin{cases} \gamma_n(t) & \text{if } 0 \leq t \leq |q - r(n)|, \\ \gamma_n(|q - r(n)|) & \text{if } t \geq |q - r(n)|. \end{cases}$$

Since the map r_n is 1-Lipschitz, the sequence $(r_n)_{n\geq 0}$ is equicontinuous. Furthermore, for all $t \geq 0$, we have $|q - r_n(t)| \leq t$, therefore the set $\{r_n(t) | n \geq 0\}$ is bounded. Since X is geodesic and proper, it is complete and we can use Ascoli's theorem to conclude that the sequence $(r_n)_{n\geq 0}$ has a subsequence $(r_{n_i})_{i\geq 0}$ that converges uniformly on compact subsets to a map $r'\colon [0, \infty[\to X$ which is a geodesic ray starting at q.

We now prove that $\mathrm{Im}(r'([0, \infty[))$ is at finite Hausdorff distance from the set $\mathrm{Im}(r([0, \infty[))$.

Let $p = r(0)$ and let us consider the two geodesic segments $[p, r(n)]$ and $[q, r(n)]$. These are the images of $r_{|[0,n]}$ and $r_{n|[0,|q-r(n)|]}$ respectively. Using the natural parameters associated to these two segments (cf. Chapter 2, §3), we set, for all t in $[0, 1]$,

$$x_t = (1 - t)p + tr(n)$$

and

$$y_t = (1 - t)q + tr(n).$$

Since X is a Busemann space, we have

$$|x_t - y_t| \leq |x_0 - y_0| = |p - q|.$$

Thus, for every point x_t on $[p, r(n)]$, there exists a point y_t on $[q, r(n)]$ satisfying $|x_t - y_t| \leq |p - q|$.

Assume that m and n are positive integers satisfying $m \geq n$, and let x_t be a point on $[p, r(n)]$. Since $[p, r(n)] \subset [p, r(m)]$, we can find a point $y_{t,m}$ on $[q, r(m)] = r_m([0, |q - r(m)|])$ satisfying $|x_t - y_{t,m}| \leq |p - q|$. Thus, x_t is in $N(r_m, |p - q|)$ for all $m \geq n$. As r_m converges to r' uniformly on compact sets, we obtain $x_t \in N\big(\mathrm{Im}(r'), |p - q|\big)$. Thus, $\mathrm{Im}(r) \subset N\big(\mathrm{Im}(r'), |p - q|\big)$. An analogous argument shows that $\mathrm{Im}(r') \subset N\big(\mathrm{Im}(r), |p - q|\big)$. Finally, we obtain $d_{\mathcal{H}}\big(\mathrm{Im}(r), \mathrm{Im}(r')\big) \leq |p - q|$, which completes the proof of Proposition 10.1.7. □

The idea for the construction of the geodesic ray r' asymptotic to r is already contained in the paper [62] by Hadamard.

10.2 The visual boundary

Let X be a metric space and for each p in X, let $\mathcal{R}_p(X)$ be the set of geodesic rays starting at p, equipped with the topology of uniform convergence on compact sets.

Definition 10.2.1 (Visual boundary). Let p be a point in X. The *visual boundary at* p of X, denoted by $\partial_p X$, is the set of equivalence classes of asymptotic geodesic rays starting at p. The space $\partial_p X$ is equipped with the quotient of the topology of uniform convergence on compact sets.

Of course, if X is a bounded space (that is, if $\mathrm{diam}(X) < \infty$), then $\mathcal{R}_p(X) \simeq \partial_p X = \emptyset$.

Example 10.2.2 (Visual boundary).

(i) *Euclidean and hyperbolic space.* If $X = \mathbb{E}^n$, then for every p in X, the visual boundary $\partial_p X$ coincides with the space of geodesic rays $\mathcal{R}_p(X)$ and is homeomorphic to the sphere S^{n-1}. The same holds in the case where $X = \mathbb{H}^n$ (this is best visualized in the conformal model B^n). More generally, for complete metric spaces satisfying the existence and uniqueness of prolongation of geodesics at each point, the visual boundary at any point is homeomorphic to a metric sphere around that point.

(ii) *Homogeneous tree.* If X is a homogeneous simplicial metric tree of degree $n \geq 2$, then for each p in X, we have $\mathcal{R}_p(X) \simeq \partial_p X$ and this space is totally disconnected. For all $n \geq 3$, it is a Cantor set.

(iii) *Euclidean cone.* Let X be a singular two-dimensional Euclidean plane, with the singular set reduced to a single point s that is a cone point of angle $\alpha \in]0, \infty[$. (To avoid taking subcases, we also include the case $\alpha = 2\pi$, where X is non-singular.) More precisely, let us consider the following:

• for $\alpha < 2\pi$, the space X is obtained from the Euclidean plane $X = \mathbb{E}^2$ by starting with the region $V_\alpha \subset \mathbb{E}^2$ defined in polar coordinates by

$$V_\alpha = \{(r, \theta) | r \in [0, \infty[\text{ and } 0 \leq \theta \leq \alpha\},$$

gluing the two boundary rays $\{\theta = 0\}$ and $\{\theta = \alpha\}$ by a Euclidean isometry and taking the length metric on the resulting quotient space;

• for $\alpha = 2\pi$, the space X is the Euclidean plane;

• for $\alpha > 2\pi$, the space X is obtained by gluing as many regions of the form $V_{\alpha'}$ (defined as above for $\alpha' < 2\pi$) as needed, in the way we did it in the description before Proposition 9.1.4 where we defined the local model for a Euclidean cone point of angle $\alpha > 2\pi$, and taking again the length metric on the resulting space.

It is easy to see that for any $\alpha \in]0, \infty[$, the visual boundary $\partial_s X$ at the singular point s is homeomorphic to a circle. However, for a point p distinct from s, a quick analysis of the behaviour of the geodesic rays starting at p shows that there are two cases:

• for $\alpha \geq 2\pi$, $\mathcal{R}_p(X) \simeq \partial_p X$ is homeomorphic to a circle;

• for $\alpha < 2\pi$, $\mathcal{R}_p(X) \simeq \partial_p X$ is homeomorphic to an open interval.

One can see the property stated for $\alpha < 2\pi$ by looking at Figure 10.2: there is no geodesic ray starting at the point p and whose tangent vector at p is directed upwards.

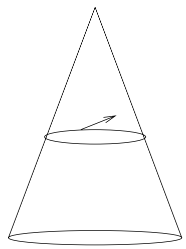


Figure 10.2. There is no geodesic ray based at the vector drawn if the vector direction is upwards (Example 10.2.2 (iii)).

Thus, in particular, we have here an example of a complete geodesic metric space X in which the spaces $\partial_p X$ and $\partial_s X$ are not homeomorphic for some p and s in X. However, in the case where X is a Busemann space (for instance, in the case where X is the Euclidean cone considered above with cone angle $\geq 2\pi$), the visual boundaries $\partial_p X$ for various $p \in X$ are pairwise homeomorphic (see [65]).

(iv) *Euclidean cylindre.* If $X = S^1 \times \mathbb{R}$ is a Euclidean cylindre, then for each point p, $\partial_p X$ consists in two points (cf. Example 10.1.3 (ii)).

(v) *Funnel-cusp.* Let X be a "funnel-cusp" in hyperbolic two-space, that is, the surface homeomorphic to a cylindre that is obtained as the quotient of the upper half-space model H^2 of hyperbolic plane $\mathbb{H}^2$ by a map of the form $z \mapsto z + c$ with $c \in \mathbb{R}$ (Figure 10.3). Then, for any x in X, we have $\partial_p X \simeq \mathcal{R}_p$, and this space is disconnected. It consists in the union of a point (corresponding to the unique geodesic ray starting at x and that tends to the cusp) with an open interval (corresponding to geodesics that converge to points on the boundary of the funnel.)

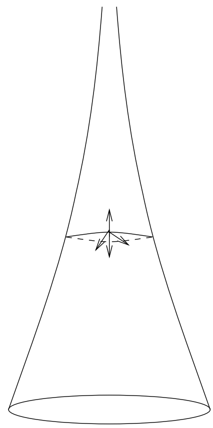


Figure 10.3. A funnel-cusp. At each point, there is one geodesic ray starting there and converging to the cusp, and a continuum of geodesic rays converging to points on the boundary of the funnel (Example 10.2.2 (v)).

(vi) For $n \geq 2$, let $X = \mathbb{E}^n$ or $\mathbb{H}^n$, let p be a point in X and let $X' = X \setminus \{p\}$ equipped with the subspace metric (which is also the intrinsic metric of the subspace). Then, the visual boundary at any point of X' is homeomorphic to an $(n-1)$-dimensional sphere with a point removed.

Proposition 10.2.3. *Let X be a proper Busemann space. Then, the visual boundary at any point is compact and for all p and q in X, there exists a natural one-to-one correspondence $\Phi_{p,q} \colon \partial_p X \to \partial_q X$.*

Proof. The compactness of the visual boundary follows from Ascoli's Theorem (Theorem 1.4.9). Of course, $\Phi_{p,q}$ is the map provided by Proposition 10.1.7, which assigns to each geodesic ray $r \colon [0, \infty[\to X$ starting at p the unique geodesic ray $r' \colon [0, \infty[\to X$ starting at q and satisfying $d_{\mathcal{H}}(r, r') < \infty$. This map is bijective since it has an inverse map, which is the map $\Phi_{q,p} \colon \partial_q X \to \partial_p X$ □

We recall that if X is a Busemann space, then, for each p in X, the space $\partial_p X \simeq \mathcal{R}_p$ is equipped with the topology of uniform convergence on compact sets. This means that a sequence of geodesic rays (r_n) in X converges to a geodesic ray r if and only if for every compact subset K of $[0, \infty[$, the restriction of r_n to K converges uniformly to the restriction of r to K.

For each point p in a Busemann space X, we now describe a natural topology on the union $\overline{X}_p = X \cup \partial_p X$.

Consider first the set X_p of maps $r\colon [0, \infty[\to X$ such that there exists $t_r \in [0, \infty[$ satisfying the following two properties:

- the restriction $r_{|[0,t_r]}$ is a geodesic path starting at p

and

- $r_{|[t_r,\infty]}$ is a constant map.

The map $X_p \to X$ which sends each element $r\colon [0, \infty[\to X$ of X_p to the point $r(t_r)$ is one-to-one. (This follows from the fact that X is uniquely geodesic.)

We now define the space $\overline{X}_p$ to be the set of maps $r\colon [0, \infty[\to X$ which are either elements of X_p, or geodesic rays starting at p. We equip $\overline{X}_p$ with the topology of uniform convergence on compact sets.

Proposition 10.2.4 (Topology on the union $X \cup \partial_p X$). *Let X be a proper Busemann space. Then, the space $\overline{X}_p = X \cup \partial_p X$, equipped with the topology of uniform convergence on compact sets, is compact. This topology on the union $\overline{X}_p = X \cup \partial_p X$ is obtained by putting a point in $\partial_p X$ as an endpoint to each geodesic ray r in X starting at p; this endpoint is the ray r itself.*

Proof. Compactness follows from Ascoli's Theorem. The two natural inclusions $X \simeq X_p \subset \overline{X}_p$ and $\partial_p X \subset \overline{X}_p$ are homeomorphisms onto their images, and therefore the space $\overline{X}_p$ has a topology that is compatible with the topologies of its two constituent subspaces. It is clear from the definition of this topology that for every geodesic ray $r\colon [0, \infty[\to X$ starting at x, the point $r \in \partial_p X$ is the limit of the family of points $r(t)$ as $t \to \infty$. □

If $f\colon X \to X$ is an isometry, then f sends any geodesic ray starting at a point p in X to a geodesic ray starting at $f(p)$. Therefore, there is a well-defined induced map

$$\partial f_p\colon \partial_p X \to \partial_{f(p)} X.$$

Likewise, f induces a well-defined map $X_p \to X_{f(p)}$, and an extension

$$f_p\colon \overline{X}_p \to \overline{X}_{f(p)},$$

which is bijective.

Let us recall a few facts about the topology of uniform convergence on compact sets and its relation to other topologies. For the details, we refer the reader to Munkres's book [114], §46. To avoid unnecessary notations, we limit ourselves to spaces of maps from $[0, \infty[$ to a metric space X instead of considering maps between two arbitrary spaces.

Rather than on the set $\partial_p X$ itself, the topology of uniform convergence on compact sets is initially defined on the set $X^{[0,\infty[}$ of all maps from $[0, \infty[$ to X. A basis for this topology consists of the family $\{B(K, \epsilon, f)\}$, where K is an arbitrary compact subset of $[0, \infty[$, ϵ an arbitrary positive real number and $f : [0, \infty[\to X$ an arbitrary map, and where

$$B(K, \epsilon, f) = \{g : [0, \infty[\to X \text{ such that } \sup_{x \in K} |f(x) - g(x)| < \epsilon\}.$$

The topology of uniform convergence of compact sets on $\partial_p X$ is the restriction of the topology of $X^{[0,\infty[}$ to the subset $\partial_p X \subset X^{[0,\infty[}$.

We now recall another useful topology, the *compact-open topology*, which is defined on the set of all maps between two topological spaces (no metric is involved). Again, to save notations, we restrict the exposition to the space of maps from $[0, \infty[$ to X rather than to maps between arbitrary spaces.

For each compact subset K of $[0, \infty[$ and for each open subset U of X, we set

$$N(K, U) = \{r : [0, \infty[\to X \text{ such that } r(K) \subset U\}.$$

The family $\{N(K, U)\}$, where K varies over the set of compact subsets of $[0, \infty[$ and U varies over the set of open subsets of X, is a sub-basis for the compact-open topology on the set of maps $X^{[0,\infty[}$.

Again, the compact-open topology on $\partial_p X$ is the restriction to this subspace of the compact-open topology on the space $X^{[0,\infty[}$.

There is a general result which says that the topology of uniform convergence on compact sets and the compact-open topology have the same restriction on the set $\mathcal{C}([0, \infty[, X)$ of continuous maps from $[0, \infty[$ to X. We refer the reader to [114], Theorem 4.6.8, p. 285. Since $\partial_p X$ is a subspace of $\mathcal{C}([0, \infty[, X)$, the two topologies that we consider here on $\partial_p X$ coincide. We record this fact as a proposition:

Proposition 10.2.5. *The topology of uniform convergence on compact sets on $\partial_p X$ coincides with the compact-open topology on this space.*

In the paper [65], P. Hotchkiss proves that for any proper Busemann space X and for any p and q in X, the map $\Phi_{p,q} : \partial_p X \to \partial_q X$ of Proposition 10.2.4 is a homeomorphism.

Notes on Chapter 10

Asymptotic geodesic rays. Asymptoticity relations on the sets of geodesic lines and of geodesic rays in metric spaces are studied by Busemann in [28], §23. We already mentioned that such ideas can be traced back to work by Hadamard, who studied asymptotic geodesic lines in surfaces of nonpositive curvature. Elie Cartan, in [36], Note III, studied asymptotic geodesics in Riemannian manifolds of nonpositive curvature, using the fact that in such spaces, the distance function from a variable point on the image of a geodesic to the image of another geodesic is convex. Asymptoticity is studied extensively in the paper [45] by Eberlein and O'Neill.

Visual boundary. The definition of the visual boundary can be made in the setting of an arbitrary geodesic metric space. The first general such definition is due to Eberlein and O'Neill (cf. [45]). One can define versions of that boundary for non-geodesic metric spaces by studying discrete geodesics.

The visual boundary of Teichmüller space. In [84], Kerckhoff showed that a compactification of the Teichmüller space $\mathcal{T}_g$ (equipped with the Teichmüller metric) of a closed surface of genus $g \geq 2$ by a visual boundary, obtained by choosing a point p in $\mathcal{T}_g$ and attaching to each geodesic ray starting at p its endpoint, depends on the choice of the point p. More precisely, Kerckhoff proved that the natural map which one obtains between the compactified spaces $\mathcal{T}_g \cup \partial_p X$ and $\mathcal{T}_g \cup \partial_q X$ is discontinuous for certain points p and q in $\mathcal{T}_g$. The study of the visual boundary of Teichmüller space is a fascinating subject. In the paper [96], J. McCarthy and the author proved that for any $g \geq 2$ and for any p in $\mathcal{T}_g$, the visual boundary $\partial_p \mathcal{T}_g$ is a non-Hausdorff space. In genus one, Teichmüller space is isometric to Hyperbolic space $\mathbb{H}^2$ and the compactification of $\mathbb{H}^2$ by the visual boundary does not depend on the choice of the basepoint.

Chapter 11
Isometries

Introduction

Let X be a metric space and let $f : X \to X$ be a map. The *minimal displacement* of f is defined as

$$\lambda(f) = \inf_{x \in X} |x - f(x)|,$$

and the *minimal set* of f is defined as

$$\mathrm{Min}(f) = \{x \in X,\ |x - f(x)| = \lambda(f)\}.$$

From these definitions, we see that if the fixed point set $\mathrm{Fix}(f)$ is not empty, then $\mathrm{Min}(f) = \mathrm{Fix}(f)$. Thus, in some sense, information on the minimal set is complementary to that provided by the fixed point set.

The minimal displacement and the minimal set of an isometry are basic tools for analyzing its global behaviour. In particular, they are used in the classification of isometries $f : X \to X$ into the following three types:

- f is parabolic if $\mathrm{Min}(f) = \emptyset$,
- f is elliptic if $\mathrm{Min}(f) \neq \emptyset$ and $\lambda(f) = 0$
- f is hyperbolic if $\mathrm{Min}(f) \neq \emptyset$ and $\lambda(f) > 0$.

This classification has been used by now by several authors (see for instance the paper [134] by Thurston and the books [8] by Ballmann, Gromov and Schroeder and [20] by Bridson and Haefliger), and it generalizes the well-known classification of isometries of hyperbolic space $\mathbb{H}^2$ that was carried on in the 19th century by F. Klein in his study of surfaces equipped with metrics of constant curvature (see [88]).

An important sub-class of the class of hyperbolic isometries is the class of axial isometries. An isometry $f : X \to X$ is *axial* if f has no fixed point and if there exists a geodesic line whose image is setwise invariant by f. Axial isometries were studied extensively by Busemann in [28]. In the case where X is a Busemann space, an isometry of X is axial if and only if it is hyperbolic.

The study of axial isometries and of their minimal displacements provides information about periodic geodesics. For instance, if $\tilde{X} \to X$ is the universal covering map of a metric space X, then a deck transformation of this covering is axial if and only if the corresponding element of the fundamental group of X contains in its free homotopy class a periodic geodesic and, in this case, the length of the periodic geodesic is equal to the minimal displacement of f.

The outline of this chapter is the following.

In Section 1, we introduce the minimal displacement and the minimal set of a self-map of a metric space X and we study the classification of isometries of X into three types: parabolic, hyperbolic and elliptic.

In Section 2, we introduce the axial isometries of X and we study the relation between axial and hyperbolic isometries and between axes and minimal sets.

In Section 3, we study some relations between the periodic geodesics in a length space (in particular, in a Busemann space) and the axial isometries of its universal cover.

In Section 4, we specialize to the case of Busemann spaces. We prove that for any isometry of such a space, its minimal set is a convex subset. We prove also that an isometry of a Busemann space is hyperbolic if and only if it is axial.

In Section 5, we study the notions of parallel geodesic lines (or, equivalently, parallel straight lines) in metric spaces. We prove that two oriented axes of the same axial isometry are parallel. Again, in this context, there are more precise statements in the case of Busemann spaces.

11.1 Minimal displacement, minimal sets and the isometry types

Definition 11.1.1 (Minimal displacement and minimal set). Let X be a metric space and let $f\colon X \to X$ be a map. The *minimal displacement of* f, denoted by $\lambda(f)$, is the greatest lower bound of the displacement function of f (Definition 3.1.3).

In symbols, we have

$$\lambda(f) = \inf_{x\in X} |x - f(x)|.$$

The *minimal set of* f, denoted by $\mathrm{Min}(f)$, is the subset of X defined as

$$\mathrm{Min}(f) = \{x \in X,\ |x - f(x)| = \lambda(f)\}.$$

We start with some examples of minimal displacements and minimal sets of self-maps of some familiar spaces.

Examples 11.1.2 (Minimal displacements and minimal sets).

(i) If H^2 is the upper-half plane model of hyperbolic plane, and if $f\colon H^2 \to H^2$ is the map of defined by $z \mapsto z + 1$, then $\lambda(f) = 0$ and $\mathrm{Min}(f) = \emptyset$.

(ii) If $f\colon \mathbb{E}^n \to \mathbb{E}^n$ is the map $x \mapsto -x$, then $\lambda(f) = 0$ and $\mathrm{Min}(f) = \{0\}$.

(iii) If $f\colon S^n \to S^n$ is defined by $x \mapsto -x$, then $\lambda(f) = 2$ (the diameter of S^n) and $\mathrm{Min}(f) = S^n$.

(iv) If $f\colon \mathbb{E}^n \to \mathbb{E}^n$ is a translation, then $\lambda(f)$ is the norm of the translation vector and $\mathrm{Min}(f) = \mathbb{E}^n$.

It follows from Definition 11.1.1 that if $\mathrm{Fix}(f) \neq \emptyset$, then $\lambda(f) = 0$ and $\mathrm{Min}(f) = \mathrm{Fix}(f)$. Thus, one can think of the minimal displacement of a map as a kind of measure for how far the map is from having a fixed point (although in general $\lambda(f) = 0$ does not imply that f has a fixed point: consider again the map $z \mapsto z+1$ of the upper-half space model of $\mathbb{H}^2$). For instance, it is clear that if X is compact, then $\lambda(f) = 0$ if and only if f has a fixed point.

Proposition 11.1.3. *For any metric space X and for any maps $f: X \to X$ and $g: X \to X$, we have $\lambda(g \circ f) \leq \lambda(f) + \lambda(g)$. In particular, $\lambda(f^n) \leq n\lambda(f)$ for every integer n.*

Proof. For all x in X, we have

$$\begin{aligned} d_{g\circ f}(x) &= |x - g \circ f(x)| \\ &\leq |x - f(x)| + |f(x) - g \circ f(x)|. \end{aligned}$$

Taking the infimum over all x in X, we obtain $\lambda(g \circ f) \leq \lambda(f) + \lambda(g)$. The rest of the proposition follows easily. □

Proposition 11.1.4. *Let X be a metric space, let $f: X \to X$ be a map and let $g: X \to X$ be an isometry. Then,*[1]

(i) *$\lambda(gfg^{-1}) = \lambda(f)$ and $\mathrm{Min}(gfg^{-1}) = g\big(\mathrm{Min}(f)\big)$;*

(ii) *if f commutes with g, $\lambda(f) = \lambda(g)$ and $\mathrm{Min}(f) = \mathrm{Min}(g)$.*

Proof. Let us first prove (i). Since g is an isometry, we have

$$\lambda(gfg^{-1}) = \inf_{x\in X} |x - gfg^{-1}(x)| = \inf_{x\in X} |g^{-1}(x) - fg^{-1}(x)|.$$

Letting y denote the point $g^{-1}(x)$, we obtain

$$\lambda(gfg^{-1}) = \inf_{y\in X} |y - f(y)|,$$

which shows that $\lambda(gfg^{-1}) = \lambda(f)$.

Let x be a point in X. Then,

$$\begin{aligned} x \in \mathrm{Min}(gfg^{-1}) &\iff |x - gfg^{-1}(x)| = \inf_{y\in X} |y - gfg^{-1}(y)| \\ &\iff |g^{-1}(x) - fg^{-1}(x)| = \inf_{y\in X} |g^{-1}(y) - fg^{-1}(y)| \\ &\iff |g^{-1}(x) - fg^{-1}(x)| = \inf_{z\in X} |z - f(z)| \\ &\iff g^{-1}(x) \in \mathrm{Min}(f) \\ &\iff x \in g(\mathrm{Min}(f)). \end{aligned}$$

[1] A remark about notation: when writing the composition of three or more maps, we delete the composition sign $\circ$.

This proves that $\mathrm{Min}(gfg^{-1}) = g\big(\mathrm{Min}(f)\big)$, which completes the proof of (i). Property (ii) follows easily from (i) and from Proposition 11.1.2 (i). □

Most of the results that we prove in the rest of this section about maps $f : X \to X$ concern the case where f is an isometry.

Proposition 11.1.5. *Let X be a metric space and let $f : X \to X$ be an isometry. Then*

(i) $\lambda(f) = \lambda(f^{-1})$ *and* $\mathrm{Min}(f) = \mathrm{Min}(f^{-1})$*;*

(ii) $\mathrm{Min}(f)$ *and* $\mathrm{Fix}(f)$ *are closed invariant subsets of* X*.*

Proof. Let us first prove (i). Using the fact that f is an isometry, we can write

$$\lambda(f^{-1}) = \inf_{x\in X} |x - f^{-1}(x)| = \inf_{x\in X} |f(x) - x| = \lambda(f).$$

Furthermore, a point x in X is in $\mathrm{Min}(f^{-1})$ if and only if

$$|x - f^{-1}(x)| = \lambda(f^{-1}) = \lambda(f).$$

Since f is an isometry, this is equivalent to $|f(x) - x| = \lambda(f)$, that is, to $x \in \mathrm{Min}(f)$. Thus, we have $\mathrm{Min}(f^{-1}) = \mathrm{Min}(f)$. This proves (i).

We prove the statement in (ii) about $\mathrm{Min}(f)$. The result for $\mathrm{Fix}(f)$ will follow automatically since either $\mathrm{Fix}(f) = \emptyset$ or $\mathrm{Fix}(f) = \mathrm{Min}(f)$.

The fact that $\mathrm{Min}(f)$ is closed follows from the continuity of f. We show invariance. Let x be in $\mathrm{Min}(f)$. Then, since f is an isometry, we have

$$|f(x) - f^2(x)| = |x - f(x)| = \lambda(f).$$

This shows that $f(x)$ is in $\mathrm{Min}(f)$. The same kind of argument shows that $f^{-1}(x)$ is in $\mathrm{Min}(f)$. We conclude that $f\big(\mathrm{Min}(f)\big) = \mathrm{Min}(f)$, which completes the proof of Proposition 11.1.5. □

We note that the f-invariance of $\mathrm{Min}(f)$ does not imply that $\mathrm{Min}(f) = \mathrm{Min}(f^2)$. For instance, if $f : \mathbb{E}^n \to \mathbb{E}^n$ is the involution $x \mapsto -x$, then $\mathrm{Min}(f)$ consists of the origin of $\mathbb{E}^n$, whereas f^2 is the identity and therefore $\mathrm{Min}(f^2) = \mathbb{E}^n$.

We have the following general classification of isometries of a metric space X in terms of the invariants $\lambda(f)$ and $\mathrm{Min}(f)$.

Definition 11.1.6 (The three isometry types). Let X be a metric space and let $f : X \to X$ be an isometry. Then f is said to be

(i) *Parabolic* if $\mathrm{Min}(f) = \emptyset$.

(ii) *Elliptic* if $\mathrm{Min}(f) \neq \emptyset$ and $\lambda(f) = 0$. (Thus, f is elliptic if and only if $\mathrm{Fix}(f) \neq \emptyset$).

(iii) *Hyperbolic* if $\mathrm{Min}(f) \neq \emptyset$ and $\lambda(f) > 0$.

The adjectives *parabolic*, *elliptic* and *hyperbolic* were used by F. Klein in his classification of isometries of the hyperbolic plane (see [88]).

We recall that two isometries f and g of X are said to be *conjugate* if there exists an isometry $h\colon X \to X$ such that $g = hfh^{-1}$. It is plain that conjugacy is an equivalence relation on the set of isometries.

Example 11.1.7. The following is a family of examples which shows in some sense that one can pass continuously from one isometry type to another. Let B_r be the closed ball in n-dimensional Euclidean space $\mathbb{E}^n$ centered at the origin and of radius $r > 0$ and let $\overset{\circ}{B}_r$ be its interior. Then, the map $x \mapsto -x$, if it is defined on $\mathbb{E}^n \setminus B_r$, is parabolic; if it is defined on $\mathbb{E}^n \setminus \overset{\circ}{B}_r$, it is hyperbolic, and if it is defined on $\mathbb{E}^n$, it is elliptic.

Proposition 11.1.8 (Classification of isometries). *For any metric space X, the isometry classes described in Definition* 11.1.6 *are pairwise disjoint. Furthermore, the inverse of any isometry is an isometry of the same type and two conjugate isometries have the same type.*

Proof. The fact that the three isometry classes are pairwise disjoint is clear from the definition, and the fact that an isometry and its inverse have the same type follows from Proposition 11.1.5. Now let $f\colon X \to X$ and $g\colon X \to X$ be two isometries that are conjugate, and let $h\colon X \to X$ be an isometry satisfying $g = hfh^{-1}$. Then, by Proposition 11.1.4 (i), we have $\lambda(f) = \lambda(g)$ and $\mathrm{Min}(f) = h(\mathrm{Min}(g))$. The last inequality shows that $\mathrm{Min}(f)$ is non-empty if and only if $\mathrm{Min}(g)$ is non-empty. This implies that f and g have the same type. □

It is clear that if f is elliptic, then f^n is elliptic for every integer n. However, it is not true in general that for f non-elliptic, f^n (for $n \neq 0$) is of the same type as f. To see this, we again take f to be the involution $x \mapsto -x$ of the sphere S^n; then f is hyperbolic whereas f^2 is the identity map and therefore elliptic. We shall see that if X is a Busemann space and if f is hyperbolic, then f^n is hyperbolic for every $n \neq 0$ (cf. Corollary 11.4.3 and 11.2.7 below).

Examples 11.1.9 (Classification of isometries). The first two examples are elementary and well-known. They are basic examples because in these cases each isometry class can be completely described up to conjugacy.

(i) *Isometries of the Euclidean plane.* We know from elementary plane Euclidean geometry that an orientation-preserving isometry of the Euclidean plane $\mathbb{E}^2$ is of one of the following two types:

- a rotation, and in this case the isometry is elliptic;
- a translation by a nonzero vector, and in this case it is hyperbolic.

We also recall that an orientation-reversing isometry of $\mathbb{E}^2$ can be decomposed into a translation followed by a symmetry with respect to a straight line that is parallel to the translation vector. Thus, an orientation-reversing isometry f of $\mathbb{E}^2$ is again either hyperbolic (this is the case where the translation vector is nonzero) or elliptic (and in this case $\mathrm{Fix}(f)$ is a straight line).

(ii) *Isometries of* $\mathbb{H}^n$. Let B^n be the conformal ball model of $\mathbb{H}^n$ and let $f: B^n \to B^n$ be an orientation-preserving isometry. It is well-known that f extends continuously to a homeomorpism of the closed ball $\overline{B^n} = B^n \cup S^{n-1}$. This can be deduced from the fact that any isometry of B^n is a Möbius transformation, that is, a composition of restrictions to the ball B^n of a finite number of inversions in Euclidean spheres or reflections in extended hyperplanes of the compactified space $\mathbb{E}^n \cup \{\infty\}$. Therefore, by the Brouwer fixed point theorem, the extension $\bar{f}: \overline{B^n} \to \overline{B^n}$ of f has a fixed point. Then:

• f is parabolic if and only if $\bar{f}$ has no fixed point in B^n and has exactly one fixed point in S^{n-1};

• f is elliptic if and only if $\bar{f}$ has a fixed point in B^n;

• f is hyperbolic if and only if $\bar{f}$ has no fixed point in B^n and has exactly two fixed points in S^{n-1}.

There are particularly simple descriptions of the actions of isometries of the hyperbolic plane $\mathbb{H}^2$, up to conjugacy. For instance:

• in the upper-half plane model H^2, a parabolic isometry is conjugate to a Euclidean translation that preserves this plane, that is, to a map of the form $(x_1, x_2) \mapsto (x_1, x_2+c)$ with c a nonzero real number;

• in the conformal disk model B^2, an elliptic isometry is conjugate to a Euclidean rotation about the Euclidean center of the disk;

• in the upper-half plane model H^2, a hyperbolic isometry is conjugate to a Euclidean homothety, that is, to a map of the form $(x_1, x_2) \mapsto \lambda(x_1, x_2)$ with $\lambda \neq 1$.

There exist descriptions of the same kind as well as algebraic characterizations of each isometry type for hyperbolic spaces of any dimension. For dimensions 2 and 3, we refer the reader to Thurston's analyses in Chapter 2 of [137] and Chapter 5 of [134]. The examples above were certainly the main motivation for the work on the classification of isometries of other spaces of various kinds (nonpositively curved spaces, Teichmüller spaces, symmetric spaces, Gromov-hyperbolic spaces and so on).

(iii) *Isometries of an* $\mathbb{R}$*-tree.* An isometry of an $\mathbb{R}$-tree is either elliptic or hyperbolic. This is proved by Morgan and Shalen in [112] in the setting of Λ-trees (which are objects more general than $\mathbb{R}$-trees). For a proof in the special case of $\mathbb{R}$-trees, we refer to the paper [42] by Culler and Shalen. In fact, in Section 1.3 of that paper, Culler and Shalen prove that given any isometry f of an $\mathbb{R}$-tree T, its minimal set $\mathrm{Min}(f)$ is a closed nonempty invariant subtree of T, and that either $\lambda(f) = 0$ and in that case $\mathrm{Min}(f) = \mathrm{Fix}(g)$, or $\lambda(f) > 0$ and in that case $\mathrm{Min}(f)$ is isometric to a real line and the action of f on $\mathrm{Min}(f)$ is conjugate to a translation of $\mathbb{R}$ by the factor $\lambda(f)$. Furthermore, in the last case, for any point p in T, its image $f(p)$ satisfies

$$|p - f(p)| = \lambda(f) + 2d(p, \mathrm{Min}(f)).$$

In Figure 11.1, p' and p'' denote respectively the projections of p and $f(p)$ on the axis of f, and $pp'p''f(p)$ is the geodesic segment joining p to $f(p)$.

In the special case of simplicial trees, this classification is contained in Serre's book [128], Chapter I, Proposition 2.4.

Figure 11.1. The horizontal line is the axis of f (Example 11.1.9 (iii)).

(iv) *Isometries of the Teichmüller metric.* Consider again the Teichmüller space $\mathcal{T}_g$ of a closed oriented surface S_g of genus $g \geq 2$. By a theorem of H. L. Royden (see [125]), the isometry group of $\mathcal{T}_g$ equipped with the Teichmüller metric is the *modular group* (also called the *mapping class group*) $\mathcal{M}_g$ of S_g, that is, the group of isotopy classes of self-homeomorphisms of S_g. Such an isotopy class is usually called a *mapping class*. In the paper [13], L. Bers worked out a classification of the elements of $\mathcal{M}_g$ in terms of their action on $\mathcal{T}_g$ equipped with the Teichmüller metric. Again, this classification is done in terms of the minimal set and the displacement function of mapping classes acting by isometries on $\mathcal{T}_g$. Bers's definitions of an elliptic and of a hyperbolic element of $\mathcal{M}_g$ are the same as those in Definition 11.1.6 above. However, in the case where the minimal set of a mapping class f is empty, Bers distinguishes two cases:

• the *parabolic* case, which (in Bers's definition) corresponds to $\mathrm{Min}(f) = \emptyset$ and $\lambda(f) = 0$;

• the *pseudo-hyperbolic* case, which corresponds to $\mathrm{Min}(f) = \emptyset$ and $\lambda(f) = 0$.

Bers's classification of mapping classes is related to Thurston's famous classification (and, in fact, Bers's classification is based on Thurston's). We recall that in Thurston's theory of surfaces and their diffeomorphisms (see [136]), mapping classes are classified into three types:

• *finite order*, which is the case where the mapping class can be represented by a periodic homeomorphism;

• *reducible*, which is the case where the mapping class has a representative which preserves a collection of disjoint simple loops (such a collection is called a "reducing curve system") which are not homotopic to a point and which are pairwise non-homotopic;

• *pseudo-Anosov*, which is the case where the mapping class has a representative which preserves two transverse measured foliations, multiplying the transverse measure of one of these foliations by a factor $\lambda \neq 1$ and the transverse measure of the other foliation by the factor $1/\lambda$.

The relation between Bers's and Thurston's classification is as follows. A mapping class is of finite order if and only if it is elliptic, and it is pseudo-Anosov if and only if it is hyperbolic. A reducible non-finite order mapping class can be either parabolic or pseudo-hyperbolic in the sense of Bers, depending on the action of the reducible map on the components of the surface cut-off along a reducing curve system.

In the case of a closed surface of genus 1 (that is, the case of the two-torus), the Teichmüller space, equipped with its Teichmüller metric, is isometric to the hyperbolic plane $\mathbb{H}^2$ and the mapping class group is simply the group $\mathrm{GL}(2, \mathbb{Z})$ acting on $\mathbb{H}^2$ (realized as the upper half-plane model of hyperbolic 2-space) by fractional linear transformations.

(v) *Isometries of the Weil–Petersson metric.* H. Masur and M. Wolf proved in the paper [100] that the isometry group of Teichmüller space equipped with the Weil–Petersson metric is, like in the case of the Teichmüller metric, the mapping class group. In the paper [43], G. Daskalopoulos and R. Wentworth worked out a classification of isometries of Teichmüller space, equipped with the Weil–Petersson metric, in terms of the minimal displacement of such an isometry. This classification parallels Bers's classification in the case of the Teichmüller metric. Daskalopoulos and Wentworth introduced the following terminology.

• A mapping class is *pseudo-periodic* if it is either of finite order or reducible and finite order on its components (that is, on the surface cut-off along a reducing curve system).

• A mapping class is *strictly pseudo-periodic* if it is pseudo-periodic but not periodic.

Daskalopoulos and Wentworth proved the following:

• A mapping class is semi-simple if and only if it is periodic or pseudo-Anosov. The periodic (respectively pseudo-Anosov) case occurs when the minimal displacement is zero (respectively nonzero).

• A mapping class is not semi-simple if and only if it is strictly pseudo-periodic or reducible but not pseudo-periodic. The strictly pseudo-periodic (respectively reducible but not pseudo-periodic) case occurs when the minimal displacement is zero (respectively nonzero).

The arguments in the paper [43] are based on the existence of invariant geodesics for the action of pseudo-Anosov mapping classes (that is, the authors prove that pseudo-Anosov mapping classes are axial isometries for the Weil–Petersson metric), and on the fact that if γ and γ' are axes of independent pseudo-Anosov mapping classes, then these axes diverge in the sense that the map $(t, s) \mapsto d(\gamma(t), \gamma'(t))$ is proper.

The union of classes (ii) and (iii) in Definition 11.1.6 is the class of isometries whose displacement function assumes a minimum. Such isometries appear for instance as deck transformations of covering spaces. This class of isometries is also particularly interesting in the study of group actions on metric spaces. For instance, any element of a group acting by isometries properly discontinuously and cocompactly belongs to this class. In fact, the elements in this class have a special name, and we record the following definition:

Definition 11.1.10 (Semi-simple isometry). Let X be a metric space. We say that an isometry $f: X \to X$ is *semi-simple* if $\mathrm{Min}(f) \neq \emptyset$, or, equivalently, if f is either elliptic or hyperbolic.

As we noted in Example 11.1.9 (iii), every isometry of an $\mathbb{R}$-tree is semi-simple.

Now we turn to isometries of Busemann spaces. We have the following

Proposition 11.1.11 ($\mathrm{Min}(f)$ and $\mathrm{Fix}(f)$ are convex). *Let X be a Busemann space and let $f: X \to X$ be an isometry. Then* $\mathrm{Fix}(f)$ *and* $\mathrm{Min}(f)$ *are invariant closed convex subsets of X.*

Proof. The fact that $\mathrm{Fix}(f)$ and $\mathrm{Min}(f)$ are closed and invariant is contained in Proposition 11.1.5. Let us prove that $\mathrm{Min}(f)$ is convex. Again, the result for $\mathrm{Fix}(f)$ will follow from the result for $\mathrm{Min}(f)$ since if $\mathrm{Fix}(f)$ is not the empty set, then it is equal to $\mathrm{Min}(f)$.

Let x and y be two points in $\mathrm{Min}(f)$ and let $\gamma: [a,b] \to X$ be a geodesic path joining them. We have $d_f \circ \gamma(a) = d_f \circ \gamma(b) = \lambda(f)$ and it follows from the definition of the map d_f that for any t in $[a,b]$, we have $d_f \circ \gamma(t) \geq \lambda(f)$. We saw in Example 8.4.5 (iv) that the map $d_f \circ \gamma: [a,b] \to \mathbb{R}$ is convex and therefore, by Proposition 6.2.14, $d_f \circ \gamma(t) = \lambda(f)$ for every t in $[a,b]$. Thus, the image of γ is contained in $\mathrm{Min}(f)$, which proves that $\mathrm{Min}(f)$ is convex. This completes the proof of Proposition 11.1.11. □

We note that the convexity of $\mathrm{Fix}(f)$ in a general uniquely geodesic metric space also follows from Corollary 3.2.6.

11.2 Axial isometries

Most of the results of this section are due to Busemann (cf. [28], §6), although he presents them under slightly different hypotheses.

Definition 11.2.1 (Axial isometry). Let X be a metric space and let $f: X \to X$ be an isometry. Then f is said to be *axial* if f has no fixed point and if there exists a straight line, which is called an *axis of f*, that is setwise invariant by f.

It is clear from this definition that if an isometry $f : X \to X$ is axial then its inverse is also axial, and that f and its inverse have the same axes.

The simplest example of an axial isometry is a non-trivial translation of $\mathbb{R}$. More generally, if X is any metric space, then any self-map of $X \times \mathbb{R}$ (equipped with a product metric) that is the identity on the first factor and a non-trivial translation on the second factor is axial. Any hyperbolic isometry of hyperbolic space $\mathbb{H}^n$ is axial. We also saw that any hyperbolic isometry of an $\mathbb{R}$-tree is axial (see Example 11.1.9 (iii) above).

In some cases, the axis of an axial isometry is unique. This holds for instance in the case of a hyperbolic isometry of $\mathbb{H}^n$, or of a hyperbolic isometry of an $\mathbb{R}$-tree (Example 11.1.9 (iii)). But there are cases where the axis is not unique, like the case of a non-trivial translation of $\mathbb{E}^n$, where every straight line which is parallel to the translation vector is an axis.

Proposition 11.2.2. *Let X be a metric space, let $f : X \to X$ be an axial isometry and let ℓ be an axis of f. Then, for every x and x' on ℓ, we have $|x - f(x)| = |x' - f(x')|$. Furthermore, the elements of the sequence $(f^n(x))_{n\in\mathbb{Z}}$ are pairwise distinct, and they follow each other on the line ℓ according to the order induced by the index $n \in \mathbb{Z}$. Finally, the orientation of ℓ for which a point on this axis is translated in the positive direction does not depend on the choice of this point.*

Proof. The axis ℓ, equipped with the metric induced from that of X, is isometric to the real line $\mathbb{R}$ equipped with its usual metric, and the map induced on this axis is conjugate to a translation of $\mathbb{R}$. Since f has no fixed point, this translation of $\mathbb{R}$ is non-trivial and consequently it (and therefore f) satisfies all the required properties. □

Using Proposition 11.2.2, we can make the following definition:

Definition 11.2.3 (The canonical orientation of an axis). Let X be a metric space, let $f : X \to X$ be an axial isometry and let ℓ be an axis of f. Then the *canonical orientation of ℓ* is the orientation provided by Proposition 11.2.2, that is, the one for which any point on ℓ is translated in the positive direction under the action of f.

It is clear that ℓ, equipped with its anti-canonical orientation, is an axis for f^{-1} equipped with its canonical orientation with respect to the map f^{-1}.

Proposition 11.2.4. *Let X be a metric space, let $f : X \to X$ be an isometry and suppose that there exists a straight line ℓ on which the induced action of f is a non-trivial translation. Then for every x on ℓ and for every integer n, we have*

$$|x - f^n(x)| = |n| \, .|x - f(x)|.$$

Proof. This also follows from the fact that the action induced by f on ℓ is conjugate to a non-trivial translation of $\mathbb{R}$. □

Proposition 11.2.5. *Let X be a metric space and let $f: X \to X$ be an isometry. Suppose that there exists a straight line ℓ on which the induced action of f is a non-trivial translation. (In other words, suppose that this induced action is conjugate to a non-trivial translation of $\mathbb{R}$.) Then for every x in X, we have $|x - f^n(x)| \to \infty$ as $n \to \infty$.*

Proof. By Proposition 11.2.4, for any x on ℓ, we have $|x - f^n(x)| \to \infty$ as $n \to \infty$. Now if y is an arbitrary point in X, we choose a point x on ℓ and we write, using the triangle inequality,

$$\begin{aligned} |y - f^n(y)| &\geq -|y - x| + |x - f^n(x)| - |f^n(x) - f^n(y)| \\ &= |x - f^n(x)| - 2|y - x|, \end{aligned}$$

which shows that $|y - f^n(y)| \to \infty$ as $n \to \infty$. □

Corollary 11.2.6. *Let X be a metric space and let $f: X \to X$ be an isometry. Then, the following two properties are equivalent:*

(i) *f is axial;*

(ii) *there exists a straight line ℓ on which the induced action of f is a non-trivial translation.*

Proof. We have (i) ⇒ (ii) by Proposition 11.2.2 and its proof. Conversely, if (ii) is satisfied, then, by Proposition 12.2.5, f has no fixed point, which shows that f is axial. □

Corollary 11.2.7. *If $f: X \to X$ is an axial isometry, then, for every nonzero integer n, the isometry f^n is axial.*

Proof. This follows from the characterization of axial isometries in Corollary 11.2.6 (ii). □

It is easy to see that if f is axial, then for every $n \neq 0$, any axis of f is an axis of f^n. But the converse is not true, that is, there may be axes of f^n that are not axes of f, as one can see by taking $f: \mathbb{E}^2 \to \mathbb{E}^2$ to be a translation by a non-zero vector v followed by a reflection along a straight line ℓ that is parallel to v. In this case, ℓ is the only axis of f, whereas f^2 is a translation of $\mathbb{E}^2$ of vector $2v$, which implies that each straight line parallel to v is an axis of f^2.

Proposition 11.2.8 (Axes are contained in the minimal set). *Let X be a metric space and let $f: X \to X$ be an axial isometry. Then every axis of f is contained in* $\mathrm{Min}(f)$.

Proof. Let ℓ be an axis of f and let n be an integer ≥ 2. By Proposition 11.2.4, we have, for every $x \in \ell$, $|x - f^n(x)| = n|x - f(x)|$. If z is an arbitrary point in X, we have $|f^i(z) - f^{i+1}(z)| = |z - f(z)|$ for all $i \in \mathbb{Z}$. We deduce that

$$\begin{aligned} n|x - f(x)| &= |x - f^n(x)| \\ &\leq |x - z| + |z - f(z)| + |f(z) - f^2(z)| \\ &\qquad + \cdots + |f^{n-1}(z) - f^n(z)| + |f^n(z) - f^n(x)| \\ &= 2|x - z| + n|z - f(z)|. \end{aligned}$$

Dividing by n, we get $|x - f(x)| \leq |z - f(z)| + (2/n)|x - z|$. Letting $n \to \infty$, we obtain $|x - f(x)| \leq |z - f(z)|$ for all z in X, which implies $|x - f(x)| = \lambda(f)$. Thus, x is in $\mathrm{Min}(f)$. This proves Proposition 11.2.8. □

Corollary 11.2.9 (Axial implies hyperbolic). *Let X be a metric space and let $f : X \to X$ be an axial isometry. Then f is of hyperbolic type.*

Proof. By Proposition 11.2.8, the minimal set of f is not empty, which implies that f is not parabolic. Since f has no fixed point, it is not elliptic. Therefore, by Proposition 11.1.8, f is hyperbolic. □

Thus, an axial isometry has a nonempty minimal set.

Proposition 11.2.10. *Let X be a metric space, let $f : X \to X$ be an axial isometry with axis ℓ and let $\gamma : \mathbb{R} \to X$ be a geodesic line whose image is ℓ and such that the canonical orientation of ℓ as an axis of f coincides with the positive orientation of $\mathbb{R}$ (transported by the map γ). Then, for every $t \in \mathbb{R}$, we have $f \circ \gamma(t) = \gamma(t + \alpha)$ with $\alpha = \lambda(f)$. Furthermore, $\lambda(f)$ is the smallest value of α for which the equation $f \circ \gamma(t) = \gamma(t + \alpha)$ holds.*

Proof. Again, this follows from the fact that the action induced by f on the axis ℓ (equipped with the metric induced from that of X) is conjugate to a translation of $\mathbb{R}$ by some constant α. Since ℓ is contained in $\mathrm{Min}(f)$ (Proposition 11.2.8), we obtain $\alpha = \lambda(f)$, and the proposition follows easily from this fact. □

Proposition 11.2.11. *Let X be a metric space, let $f : X \to X$ be an axial isometry with axis ℓ and let $g : X \to X$ be an arbitrary isometry. Then gfg^{-1} is an axial isometry having $g(\ell)$ as an axis and $\lambda(f)$ as minimal displacement.*

Proof. Suppose that $gfg^{-1}(x)$ had some fixed point x. Then $gfg^{-1}(x) = x$, which implies $fg^{-1}(x) = g^{-1}(x)$. Therefore $g^{-1}(x)$ would be a fixed point of f. Since f has no fixed point, we conclude that gfg^{-1} has no fixed point. Furthermore, we have $gfg^{-1}(g(\ell)) = g(f(\ell)) = g(\ell)$, which implies that gfg^{-1} is axial and that $g(\ell)$ is an axis for this isometry. Finally, to find the minimal displacement of gfg^{-1}, we take an arbitrary point x on $g(\ell)$. By Proposition 11.2.8, $g(\ell)$ is contained in the minimal set of gfg^{-1}, therefore we have

$$\lambda(gfg^{-1}) = |x - gfg^{-1}(x)| = |g^{-1}(x) - fg^{-1}(x)|.$$

Since $g^{-1}(x)$ is on ℓ, we have $|g^{-1}(x) - fg^{-1}(x)| = \lambda(f)$, which implies $\lambda(gfg^{-1}) = \lambda(f)$. This proves Proposition 11.2.11. □

Proposition 11.2.12. *Let X be a uniquely geodesic metric space and let $f: X \to X$ be a fixed-point free isometry. Suppose that there exists a point x in X and a straight line ℓ in X such that for every n in $\mathbb{Z}$, the point $f^n(x)$ lies on ℓ. Then f is an axial isometry and ℓ is an axis of f.*

Proof. Since the space X is uniquely geodesic, for every x and y in X, f sends the unique geodesic segment joining x and y to the unique geodesic segment joining $f(x)$ and $f(y)$. Suppose that the sequence $\big(f^n(x)\big)_{n\in\mathbb{Z}}$ lies on a straight line ℓ. Then the restriction of f to ℓ is conjugate to a translation of $\mathbb{R}$, and there exists a positive constant c such that for any integer n, the point $f^n(x)$ is situated at distance $|n|c$ from x. Since for every integer n, the map f sends the segment $[f^n(x), f^{n+1}(x)]$ to the segment $[f^{n+1}(x), f^{n+2}(x)]$, we deduce that f preserves ℓ. Thus, f is an axial isometry and ℓ is an axis of f. □

11.3 Periodic geodesics

Definition 11.3.1 (Periodic geodesic). Let X be a metric space. A *periodic geodesic* in X is a map $c: \mathbb{R} \to X$ that is periodic and locally distance-preserving. In other words, c is a periodic geodesic if there exists $\alpha > 0$ (called a *period* of c) such that $c(t+\alpha) = c(t)$ for all t in $\mathbb{R}$ and if for any such t, there is a closed interval $I(t) \subset \mathbb{R}$ containing t in its interior such that the restriction of c to $I(t)$ is a geodesic path. We shall call a *periodic geodesic of period* α a pair (c, α) where $c: \mathbb{R} \to X$ is a periodic geodesic and α a period of the map c. (Note that we do not assume in this definition that α is the smallest period of c). The *length* of a periodic geodesic of period α is then the length of the restriction of c to the interval $[0, \alpha]$.

We recall that a *loop* in a topological space is a path whose two endpoints coincide, and that the *free homotopy class* of a loop is, by definition, its homotopy class with no basepoint fixed.

We call the *homotopy class* of a periodic geodesic c of period α the free homotopy class of the loop $c_{|[0,\alpha]}$.

We now give a criterion for a loop in a proper length space to be geodesic, which is analogous to the criterion for geodesic paths in homotopy classes with fixed endpoints that we saw in Chapter 2 (proof of Proposition 2.4.11):

Proposition 11.3.2. *Let X be a proper length space in which each point is the center of a simply connected ball of positive radius, and let γ be a loop in X whose length α is smallest in its free homotopy class. Then there exists a periodic geodesic $c\colon \mathbb{R} \to X$ whose length is equal to α and such that the map $c_{|[0,\alpha]}$ coincides with γ, up to a reparametrization of this loop γ.*

Proof. We omit the proof since it is a straightforward adaptation of the proof of Proposition 2.4.11. □

Example 11.3.3 (Funnel-cusp). A standard example in classical hyperbolic geometry of a space X with a non-trivial free homotopy class of loops containing no periodic geodesic is the cylinder obtained as a quotient of hyperbolic space $\mathbb{H}^2$ by a parabolic isometry. In the upper half-plane model H^2, this cylinder (which we called a funnel-cusp) is the quotient of H^2 by a map of the form $z \mapsto z + \zeta$ ($\zeta > 0$). If two points x_1 and x_2 are on the same horizontal line in this space, situated at (Euclidean) distance y above the x-axis and whose Euclidean mutual distance is a, then the hyperbolic length of the Euclidean segment $[x_1, x_2]$ is equal to a/y. Letting y tend to infinity, the family of images of the segments $[x_1, x_2]$ in the cylinder gives a family of loops in the same non-trivial homotopy class whose lengths are unbounded from below.

It should be noted, in relation to Proposition 11.3.2, that the fact that the length of a loop is smallest in its free homotopy class is not a necessary condition for that loop to be a periodic geodesic. In fact, we know that on the contrary, on a sphere, the geodesics are the circles of largest radius. We can also give examples of non-simply connected surfaces; in the example described in Figure 11.2 which represents a surface homeomorphic to a cylinder that is embedded in $\mathbb{R}^3$, the two circles that have small radii are periodic geodesics, but the central circle of large radius is also a periodic geodesic.

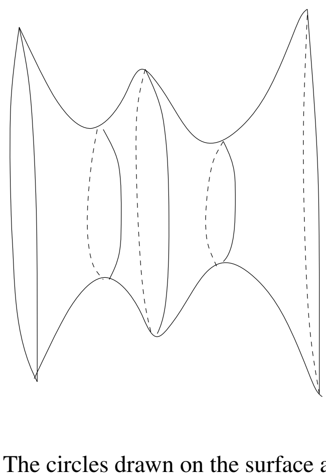


Figure 11.2. The circles drawn on the surface are geodesics.

Proposition 11.3.4. *Let $p\colon \tilde{X} \to X$ be a covering map between length spaces, with p being a local isometry, let $f\colon \tilde{X} \to \tilde{X}$ be a deck transformation that is axial and let $\gamma\colon \mathbb{R} \to \tilde{X}$ be a geodesic line whose image is an axis ℓ of f. Then $p \circ \gamma$ is a periodic geodesic in X and the minimal displacement $\lambda(f)$ is a period for this periodic geodesic.*

Proof. By Proposition 3.4.8, $p\circ\gamma$ is a local geodesic. Since f is a deck transformation, we have $p\circ\gamma(t) = p\circ f(\gamma(t))$ for every t in $\mathbb{R}$ and by Proposition 11.2.10, $f(\gamma(t)) = \gamma(t+\alpha)$ where $\alpha = \lambda(f)$. This implies that the map $p\circ\gamma$ is a periodic geodesic with period $\lambda(f)$. □

In the case where X is a locally convex space, we have the following converse:

Proposition 11.3.5. *Let X be a locally convex space, let $c\colon \mathbb{R} \to X$ be a periodic geodesic of length $\alpha > 0$ and let $p\colon \tilde{X} \to X$ be the universal covering equipped with the pull-back length metric. Then, there exists a deck transformation $f\colon \tilde{X} \to \tilde{X}$ that is an axial isometry with minimal displacement $\lambda(f) = \alpha$, and a geodesic line $\ell\colon \mathbb{R} \to \tilde{X}$ whose image is an axis of f and such that $p \circ \gamma = c$.*

Proof. Let us set $x = c(0)$, let $\tilde{x}$ be a point in the fibre $p^{-1}(x)$ and let $c_\alpha\colon [0,\alpha] \to X$ be the restriction of c to $[0,\alpha]$. Then c_α is a local geodesic and by Proposition 3.4.11, we can lift it to a local geodesic $\tilde{c}_\alpha\colon [0,\alpha] \to \tilde{X}$ starting at $\tilde{x}$. The point $\tilde{c}_\alpha(\alpha)$ is in the fibre $p^{-1}(x)$ and it is distinct from $\tilde{x}$, since $\tilde{X}$, being the universal cover of a locally convex space, is a Busemann space and since in such a space there is a unique local geodesic joining any two points (in particular, if the two points coincide, then this local geodesic is the constant map). By the classical theory of covering spaces, there is a deck transformation $f\colon \tilde{X} \to \tilde{X}$ that sends $\tilde{x}$ to $\tilde{c}_\alpha(\alpha)$. The image of the segment $[\tilde{x}, \tilde{c}_\alpha(\alpha)]$ by f is the lift of $\tilde{c}_\alpha$ starting at $\tilde{c}_\alpha(\alpha)$. Let us call $\tilde{c}'_\alpha(\alpha)\colon [0,\alpha] \to \tilde{X}$ this lift. Since $c\colon \mathbb{R} \to X$ is a local geodesic, the concatenation $\tilde{c}_\alpha * \tilde{c}'_\alpha\colon [0,2\alpha] \to \tilde{X}$ is a local geodesic in $\tilde{X}$. The composed map $p\circ(\tilde{c}_\alpha * \tilde{c}'_\alpha)$ winds twice around the image of $\tilde{c}_\alpha$ in X. Since $\tilde{X}$ is a Busemann space, the map $\tilde{c}_\alpha * \tilde{c}'_\alpha$, being a local geodesic in $\tilde{X}$, is a geodesic. By an easy induction using the same construction, we obtain, by concatenating bi-infinitely many lifts of c_α, a geodesic line $\tilde{c}\colon \mathbb{R} \to \tilde{X}$ whose image is invariant by f, on which f acts as a translation by the factor α and satisfying $p\circ\tilde{c} = c$. This completes the proof of Proposition 11.3.5. □

A periodic geodesic $c\colon \mathbb{R} \to X$ of period α, equipped with an orientation (in general, we shall take the orientation induced from the positive orientation of $\mathbb{R}$) defines an oriented loop in X and therefore a conjugacy class in the fundamental group $\pi_1(X)$.

Regarding Proposition 11.3.5, it is worthwhile to note that the deck transformation f that is obtained from the periodic geodesic c equipped with the orientation induced from the positive orientation of $\mathbb{R}$, is the deck transformation which by the general theory of coverings is associated to a loop based at the basepoint of X, and that this deck transformation represents an element of the fundamental group of X, *via* the canonical isomorphism between $\pi_1(X)$ and the group of deck transformations of the universal covering. Of course, in order to have a well-defined isomorphism between the fundamental group and the group of deck transformations, one has to make the choice of a basepoint. Still, without such a choice, the free conjugacy class of an oriented loop determines a well-defined conjugacy class of deck transformations. Using this remark, from Propositions 11.3.4 and 11.3.5 we deduce the following:

Proposition 11.3.6. *Let X be a locally convex space, let $p\colon \tilde{X} \to X$ be its universal covering, with $\tilde{X}$ equipped with the pull-back length metric, and let η be a free homotopy class of loops in X. Then, the conjugacy class of deck transformations determined by η is axial if and only if there exists a periodic geodesic in the free homotopy class η.*

Proof. By Proposition 11.3.5, the deck transformation associated to a periodic geodesic is axial, and by Proposition 11.3.4, an axial transformation determines a periodic geodesic in the corresponding free homotopy class in the fundamental group. This proves Proposition 11.3.6. □

11.4 Axial isometries of Busemann spaces

We begin this section with a useful lemma:

Lemma 11.4.1. *Let X be a Busemann space, let $f\colon X \to X$ be an isometry without fixed point and suppose that there exists a point x in X satisfying $|x - f(x)| = \lambda(f)$. Then the point $f(x)$ lies between x and $f^2(x)$.*

Proof. Let z be the midpoint of the segment $[x, f(x)]$. Since f is an isometry, $f(z)$ is the midpoint of the segment $[f(x), f^2(x)]$. We claim that

$$|z - f(x)| + |f(x) - f(z)| = |z - f(z)|. \qquad (11.4.1.1)$$

Indeed, if not, then we would have

$$\begin{aligned} |z - f(x)| + |f(x) - f(z)| &> |z - f(z)| \\ &\geq |x - f(x)| \\ &= |z - f(x)| + |z - x| \\ &= |z - f(x)| + |f(x) - f(z)|, \end{aligned}$$

which is a contradiction. This proves (11.4.1.1), which implies that $f(x)$ lies between z and $f(z)$. Since z lies between x and $f(x)$, since $f(x)$ lies between z and $f(z)$ and since X is a Busemann space, Proposition 8.2.4 implies that $f(x)$ lies between x and $f(z)$. By the same proposition, since $f(x)$ lies between x and $f(z)$ and since $f(z)$

lies between $f(x)$ and $f^2(x)$, we conclude that $f(x)$ lies between x and $f^2(x)$. This proves Lemma 11.4.1. □

Proposition 11.4.2. *Let X be a Busemann space and let $f: X \to X$ be a hyperbolic isometry. Then f is axial and for every x in* $\mathrm{Min}(f)$*, all the elements of the sequence $\big(f^n(x)\big)_{n\in\mathbb{Z}}$ lie on an axis of f.*

Proof. By Lemma 11.4.1, the point $f(x)$ belongs to $[x, f^2(x)]$, the unique geodesic segment joining x to $f^2(x)$, and this segment is the union of geodesic segments $[x, f(x)] \cup [f(x), f^2(x)]$. Since x is in $\mathrm{Min}(f)$, we have

$$\begin{aligned}|x - f(x)| &= |f(x) - f^2(x)|\\ &= \inf_{z\in X} |z - f(z)|\\ &= \inf_{f(z)\in X} |f(z) - f^2(z)|\\ &= \lambda(f).\end{aligned}$$

By Lemma 11.4.1 applied to the point $f(x)$, we conclude that $f^2(x)$ lies between $f(x)$ and $f^3(x)$. By the existence of geodesics and by the uniqueness of local geodesics joining two arbitrary points in the Busemann space X, the union of geodesic segments

$$[x, f(x)] \cup [f(x), f^2(x)] \cup [f^2(x), f^3(x)],$$

which is the image of a local geodesic (and therefore of a geodesic), is a geodesic segment. The four points $x, f(x), f^2(x), f^3(x)$ are contained in that order on that geodesic segment. By an induction that uses the same reasoning, we obtain that the union $\bigcup_{n\in\mathbb{Z}}[x, f(x)]$ is a straight line ℓ in X which contains the point $f^n(x)$ for every $n \in \mathbb{Z}$. By Proposition 11.2.12, f is an axial isometry and ℓ is an axis of f. □

Corollary 11.4.3 (In a Busemann space, axial is equivalent to hyperbolic). *Let X be a Busemann space and let $f: X \to X$ be an isometry. Then f is axial if and only if it is hyperbolic.*

Proof. This follows from Corollary 11.2.9 and Proposition 11.4.2. □

Corollary 11.4.4. *Let $f: X \to X$ be a hyperbolic isometry of a Busemann space X. Then, for any nonzero integer n, the isometry f^n is hyperbolic.*

Proof. The proof follows from Corollary 11.4.3 and Corollary 11.2.7. □

We already noted that the conclusion of Corollary 11.4.4 is not true in a geodesic metric space without further assumptions. For instance, the isometry of the sphere S^n defined by $x \mapsto -x$ is hyperbolic, but its square is the identity map, which is not hyperbolic.

Proposition 11.4.5 (Axes are disjoint). *Let X be a Busemann space and let $f: X \to X$ be an axial isometry. Then, the axes of f are disjoint*

Proof. The proof of Proposition 11.4.2 implies that for any point x on an axis of f, the orbit of f completely determines the axis containing x. Thus, if a point belongs to two axes, these two axes coincide. □

Proposition 11.4.6. *Let X be Busemann space and let $: X \to X$ be an axial (or, equivalently, a hyperbolic) isometry. Then* $\mathrm{Min}(f)$ *is the disjoint union of the axes of f, and f acts on* $\mathrm{Min}(f)$ *as a translation by the quantity $\lambda(f)$.*

Proof. The second part of Proposition 11.4.2 implies that $\mathrm{Min}(f)$ is contained in the union of the axes of f, Proposition 11.2.8 says that any axis of f is contained in $\mathrm{Min}(f)$ and Proposition 11.4.5 says that the axes of f are disjoint. □

11.5 Parallel lines

We now introduce a parallelism relation between geodesic lines that is analogous to the asymptoticity relation between geodesic rays that we introduced in Chapter 10.

Definition 11.5.1 (Parallel geodesic lines). Let X be a metric space and let $\gamma: \mathbb{R} \to X$ and $\gamma': \mathbb{R} \to X$ be two geodesic lines. Then γ is said to be *parallel to* γ' if there exists a real number α such that $|\gamma(t) - \gamma'(t)| \leq \alpha$ for every t in $\mathbb{R}$.

It is clear that this defines an equivalence relation between geodesic lines in X. It is also clear that if a geodesic line $\gamma': \mathbb{R} \to X$ is obtained from a geodesic line $\gamma: \mathbb{R} \to X$ by a change of parameter, that is, if there exists a real number c satisfying $\gamma'(t) = \gamma(c + t)$ for all $t \in \mathbb{R}$, then γ and γ' are parallel.

Examples 11.5.2 (Parallel geodesic lines).

(i) *Euclidean space.* In n-dimensional Euclidean space $\mathbb{E}^n$, two geodesic lines are parallel if and only if their images are contained in a 2-dimensional plane and if in this plane the two images are parallel in the usual sense of Euclidean plane geometry and the orientations induced from the positive orientations of $\mathbb{R}$ coincide.

(ii) *Hyperbolic space.* In n-dimensional hyperbolic space $\mathbb{H}^n$, two geodesic lines $\gamma: \mathbb{R} \to \mathbb{H}^n$ and $\gamma': \mathbb{R} \to \mathbb{H}^n$ are parallel if and only if γ' is obtained from γ by a change of parameter. This amounts to saying that the images of γ and γ' coincide and that the orientations of these images that are induced by the positive orientation of $\mathbb{R}$ coincide.

(iii) $\mathbb{R}$*-tree.* Likewise, in an $\mathbb{R}$-tree, two geodesic lines are parallel if and only if their images, equipped with the orientations induced by the positive orientations of $\mathbb{R}$, coincide.

For any geodesic line $\gamma\colon \mathbb{R} \to X$, we define the *geodesic line opposite to* γ as the geodesic line $\overline{\gamma}\colon [0, \infty[\to X$ defined by $\overline{\gamma}(t) = \gamma(-t)$ for every t in $\mathbb{R}$.

The following relation between parallelism of geodesic lines and asymptoticity of geodesic rays that we considered in Chapter 10 follows immediately from the definitions:

Proposition 11.5.3. *Let X be a metric space, let $\gamma\colon \mathbb{R} \to X$ and $\gamma'\colon \mathbb{R} \to X$ be two geodesic lines in X and let $\overline{\gamma}\colon [0, \infty[\to X$ and $\overline{\gamma}'\colon [0, \infty[\to X$ be the geodesic lines that are opposite respectively to γ and γ'. Then the geodesic lines γ and γ' are parallel if and only if the geodesic rays $\gamma_{|[0,\infty[}$ and $\gamma'_{|[0,+\infty[}$ are asymptotic and the geodesic rays $\overline{\gamma}_{|[0,+\infty[}$ and $\overline{\gamma}'_{|[0,+\infty[}$ are asymptotic.* □

An *oriented straight line* in a metric X is a straight line that is equipped with an orientation (This makes sense since a straight line is an embedded image of $\mathbb{R}$). If ℓ^+ denotes a straight line equipped with an orientation, then ℓ^- will denote the same straight line equipped with the opposite orientation.

We note that the image of any geodesic ray, being an embedded image of $[0, \infty[$, is equipped with a natural orientation, which is the one induced by the positive orientation of $[0, \infty[$.

Finally, we make the following definition:

Definition 11.5.4 (Parallel oriented straight lines). Two oriented straight lines ℓ_1^+ and ℓ_2^+ in a metric space X are said to be *parallel* if given two geodesic lines $g_1\colon \mathbb{R} \to X$ and $g_2\colon \mathbb{R} \to X$ whose images are respectively ℓ_1^+ and ℓ_2^+, with the orientations on ℓ_1^+ and ℓ_2^+ induced by the positive orientation of $\mathbb{R}$ *via* g_1 and g_2, the geodesic lines g_1 and g_2 are parallel.

From the remark following Definition 11.5.1, we can see that this definition does not depend on the choice of the geodesic lines g_1 and g_2 that parametrize ℓ_1 and ℓ_2. Furthermore, it is clear that ℓ_1^+ and ℓ_2^+ are parallel if and only if ℓ_1^- and ℓ_2^- are parallel.

Proposition 11.5.5. *Let X be a metric space, let $f\colon X \to X$ be an axial isometry and let ℓ_1^+ and ℓ_2^+ be two axes of f that are equipped with their canonical orientations. Then ℓ_1^+ and ℓ_2^+ are parallel.*

Proof. Let $\gamma_1\colon \mathbb{R} \to X$ and $\gamma_2\colon \mathbb{R} \to X$ be two geodesic lines whose images are the axes ℓ_1 and ℓ_2 respectively, such that the orientations on ℓ_1^+ and ℓ_2^+ are induced by the positive orientation of $\mathbb{R}$. Let us set $c = \lambda(f)$. For any x_1 on ℓ_1 and x_2 on ℓ_2, we have, by Propositions 11.2.4 and 11.2.8, for every integer n,

$$|x_1 - f^n(x_1)| = c|n| = |x_2 - f^n(x_2)| = \lambda(f^n).$$

Since $f^n(\gamma_1(0)) = \gamma_1(nc)$ and $f^n(\gamma_2(0)) = \gamma_2(nc)$, we obtain

$$|\gamma_1(nc) - \gamma_2(nc)| = |f^n(\gamma_1(0)) - f^n(\gamma_2(0))| = |\gamma_1(0) - \gamma_2(0)|.$$

Therefore, for every integer n and for every t in $[nc, (n+1)c]$, we have

$$\begin{aligned}|\gamma_1(t) - \gamma_2(t)| &\leq |\gamma_1(t) - \gamma_1(nc)| \\ &\quad + |\gamma_1(nc) - \gamma_2(nc)| + |\gamma_2(nc) - \gamma_2(t)| \\ &= |\gamma_1(0) - \gamma_2(0)| + 2(t - nc) \\ &\leq |\gamma_1(0) - \gamma_2(0)| + 2c,\end{aligned}$$

which shows that the geodesic lines γ_1 and γ_2 are parallel. Therefore the oriented axes ℓ_1^+ and ℓ_2^+ are parallel. □

Proposition 11.5.6. *Let X be a Busemann space and let $\gamma : \mathbb{R} \to X$ and $\gamma' : \mathbb{R} \to X$ be two parallel geodesic lines. Then the map $t \mapsto |\gamma(t) - \gamma'(t)|$, defined on $\mathbb{R}$, is constant.*

Proof. The map $t \mapsto |\gamma(t) - \gamma'(t)|$ is locally convex and therefore convex (cf. Theorem 6.2.16). A map defined on $\mathbb{R}$ that is convex and bounded is constant (Proposition 6.2.3). □

Notes on Chapter 11

The classification of isometries of Riemannian manifolds of nonpositive curvature. One version of the classification of isometries of Riemannian manifolds of nonpositive sectional curvature, involving the minimal set and the displacement function, is contained in the paper [17] by Bishop and O'Neill. More precisely, the authors prove the following (Proposition 4.2 p. 13 of [17]): Let M be a complete simply connected manifold of nonpositive curvature, let $f : M \to M$ be an isometry and let $d_f^2 : M \to \mathbb{R}$ be the square of the displacement function:

$$d_f^2(x) = |x - f(x)|^2.$$

(We recall that in the Riemannian case, the square of the displacement function is smooth and therefore it is often more useful than the displacement function itself which in general is not smooth.) Then, d_f^2 is convex, and exactly one of the following holds:

- f has a fixed point (that is, f is elliptic) and in this case the fixed point set of f is a closed convex submanifold of M;
- f is axial and in this case the minimal set of f is the union of the axes of f;
- the minimal set of f is empty (that is, f is parabolic).

Chapter 12

Busemann functions, co-rays and horospheres

Introduction

The three notions that are in the title of this chapter were introduced by Busemann. Let us briefly review the definitions. If X is a metric space and if $r\colon [0,\infty[\to X$ is a geodesic ray in X, then the *Busemann function* associated to r is the function $B_r\colon X\to\mathbb{R}$ defined for x in X by

$$B_r(x)=\lim_{t\to\infty}(|x-r(t)|-t).$$

A *horosphere* is a level set of a Busemann function.

Finally, given a geodesic ray $r\colon [0,\infty[\to X$ and a point x in X, a *co-ray to r starting at x* is a geodesic ray that is the limit of a sequence of geodesic paths $(\gamma_n)_{n\geq 0}$, such that for all $n\geq 0$, γ_n joins x_n to $r(t_n)$, where (x_n) is a sequence of points in X converging to x and (t_n) is a sequence of nonnegative numbers tending to infinity.

Thus, in some sense, saying that a geodesic ray r' starting at x is a co-ray to r is a way of saying that r' joins x to the point at infinity of r.

We shall limit ourselves to some elementary aspects of that theory, but we warn the reader that there are interesting results concerning co-rays that take place in spaces that admit the uniqueness of prolongation of geodesics. In fact, unlike the other notions introduced in this book, in order to treat efficiently the notion of co-rays, one has to do it in the setting of Busemann G-spaces (see the notes at the end of this chapter).

The plan of this chapter is the following.

In Section 1 we introduce Busemann functions and we study their basic properties.

Section 2 is a brief introduction to Busemann's theory of co-rays.

Section 3 concerns horospheres. We shall see that co-rays are in some sense orthogonal trajectories to horospheres. We end this chapter with a short list of problems.

12.1 Busemann functions

We wish to introduce the Busemann function associated to a geodesic ray, and we start with the following

Lemma 12.1.1. *Let X be a metric space and let $r\colon [0,\infty[\to X$ be a geodesic ray in X. Then, for every point x in X, the map $t\mapsto (|x-r(t)|-t)$ is nonincreasing and bounded from below.*

Proof. Using the triangle inequality, we have, for every t in $\mathbb{R}$,

$$|x - r(t)| - t = |x - r(t)| - |r(0) - r(t)| \geq -|r(0) - x|.$$

Thus, the map $t \mapsto (|x - r(t)| - t)$ is bounded from below. Now if t_1 and t_2 are two real numbers satisfying $0 \leq t_1 \leq t_2$, we have

$$|x - r(t_1)| \geq |r(t_1) - r(t_2)| - |x - r(t_2)|,$$

which implies

$$\begin{aligned} |x - r(t_1)| - t_1 &\geq |x - r(t_2)| - t_1 - |r(t_1) - r(t_2)| \\ &= |x - r(t_2)| - t_1 - (t_2 - t_1) \\ &= |x - r(t_2)| - t_2. \end{aligned}$$

This completes the proof of Lemma 12.1.1 □

From Lemma 12.1.1, we deduce the following

Corollary 12.1.2. *Let X be a metric space and let $r\colon [0, \infty[\to X$ be a geodesic ray in X. Then, for every x in X, $\lim_{t\to\infty}(|x - r(t)| - t)$ exists and is finite.* □

Now we can make the following

Definition 12.1.3 (Busemann function). Let X be a metric space and let $r\colon [0, \infty[\to X$ be a geodesic ray. The *Busemann function associated to r* is the map $B_r\colon X \to \mathbb{R}$ defined for x in X by

$$B_r(x) = \lim_{t\to\infty} (|x - r(t)| - t).$$

Proposition 12.1.4. *For any metric space X and for any geodesic ray $r\colon [0, \infty[\to X$, the associated Busemann function $B_r\colon X \to \mathbb{R}$ satisfies the following properties:*

(i) *for every $t \in [0, \infty[$, we have $B_r(r(t)) = -t$;*

(ii) *for every x and y in X, we have*

$$B_r(x) - B_r(y) = \lim_{t\to\infty} (|x - r(t)| - |y - r(t)|);$$

(iii) *B_r is 1-Lipschitz.*

Proof. For every $t \geq 0$ and for every $t' \geq t$, we have

$$B_r(r(t)) = \lim_{t'\to\infty} (|r(t) - r(t')| - t') = \lim_{t'\to\infty} (t' - t - t') = -t,$$

which proves (i).

For every x and y in X, we have

$$\begin{aligned} B_r(x) - B_r(y) &= \lim_{t\to\infty} \big((|x - r(t)| - t) - (|y - r(t)| + t)\big) \\ &= \lim_{t\to\infty} (|x - r(t)| - |y - r(t)|), \end{aligned}$$

which proves (ii).

For every $t \geq 0$, we have $|x - r(t)| - |y - r(t)| \leq |x - y|$. Therefore, we obtain, using (ii), $B_r(x) - B_r(y) \leq |x - y|$. By symmetry, we also have $B_r(y) - B_r(x) \leq |x - y|$, which shows that $|B_r(y) - B_r(x)| \leq |x - y|$. This proves (iii), which completes the proof of Proposition 12.1.3. □

Proposition 12.1.5 (Convexity of Busemann functions). *Let X be a Busemann space. For any geodesic ray $r\colon [0, \infty[\to X$, the associated Busemann function $B_r\colon X \to \mathbb{R}$ is convex.*

Proof. In a Busemann space, for all $t \in [0, \infty[$, the function $x \mapsto |x - r(t)|$ is convex. Therefore, the function $x \mapsto (|x - r(t)| - t)$ is also convex and by passing to the limit, the function $x \mapsto B_r(x) = \lim_{t\to\infty}(|x - r(t)| - t)$ is convex. □

The Busemann function associated to a geodesic ray r and to a subray of r are equal up to an additive constant. More precisely, we have the following:

Proposition 12.1.6. *If r_0 is a subray of r and if $r_0(0) = r(t_0)$, then, for all $x \in X$, we have $B_{r_0}(x) = B_r(x) + t_0$.*

Proof. We have

$$\begin{aligned} B_{r_0}(x) &= \lim_{t\to\infty} (|x - r_0(t)| - t) \\ &= \lim_{t\to\infty} \big(|x - r(t + t_0)| - t\big) \\ &= \lim_{t\to\infty} (|x - r(t + t_0)| - (t + t_0)) + t_0 \\ &= B_r(x) + t_0. \end{aligned}$$

□

Definition 12.1.7 (Distance at infinity between geodesic rays). Let X be a metric space and let $r\colon [0, \infty[\to X$ and $r'\colon [0, \infty[\to X$ be two geodesic rays. Then, the *distance at infinity* between r and r' is the element in $[0, \infty[\cup\{\infty\}$ defined as

$$\delta_\infty(r, r') = \liminf_{t,t'\to\infty} |r(t) - r'(t)|.$$

In particular, the distance at infinity between two asymptotic rays is finite. We note however that the fact that $\delta_\infty(r, r') = 0$ does not imply that r and r' are asymptotic, as we can see from the following example.

Example 12.1.8. Let ℓ be the graph of a positive function $f : [0, \infty[\to \mathbb{R}$ having infinitely many minima, at points t_n ($n \geq 0$), with $f(t_n)$ decreasing and tending to 0 as $n \to \infty$, and infinitely many maxima, at points t'_n ($n \geq 0$), with $f(t'_n)$ increasing and tending to ∞ as $n \to \infty$ (Figure 12.1). Let S be the surface in $\mathbb{R}^3$ obtained by rotating ℓ around the horizontal axis of Figure 12.1. For any real number θ, if ℓ_θ denotes the image of the curve ℓ by a rotation of angle θ around the horizontal axis, then ℓ and ℓ_θ are images of geodesic rays that are not asymptotic but whose distance at infinity equals zero.

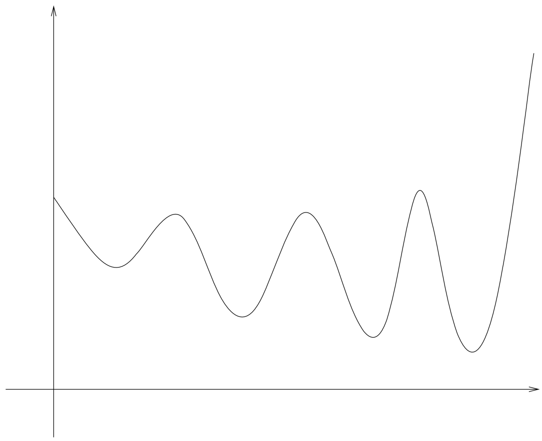


Figure 12.1. The graph ℓ of Example 12.8.1.

Proposition 12.1.9. *Let X be a metric space, let $r : [0, \infty[\to X$ and $r' : [0, \infty[\to X$ be two geodesic rays and let x and x' be two points in X. Then, we have*

$$|B_r(x) + B_{r'}(x') - B_r(x') - B_{r'}(x)| \leq 2\delta_\infty(r, r').$$

Proof. Let us take two sequences of real numbers (t_n) and (t'_n) that tend to ∞ as $n \to \infty$ and such that

$$\delta_\infty(r, r') = \lim_{n\to\infty} |r(t_n) - r'(t'_n)|.$$

We have, for all $n \geq 0$,

$$|x - r(t_n)| - |x - r'(t'{}_n)| \leq |r(t_n) - r'(t'{}_n)|$$

and

$$|x' - r(t_n)| - |x' - r'(t'{}_n)| \leq |r(t_n) - r'(t'{}_n)|,$$

which implies

$$\big||x - r(t_n)| - |x - r'(t'_n)| - (|x' - r(t_n)| - |x' - r'(t'_n)|)\big| \leq 2|r(t_n) - r'(t'_n)|.$$

As $n \to \infty$, the left hand side converges to $|B_r(x) + B_{r'}(x') - B_r(x') - B_{r'}(x)|$ while the right hand side converges to $2\delta_\infty(r, r')$. This proves Proposition 12.1.9. □

Corollary 12.1.10. *Let X be a metric space and $r\colon [0, \infty[\to X$ and $r'\colon [0, \infty[\to X$ be two geodesic rays in X. If $\delta_\infty(r, r') = 0$, then $B_r - B_{r'}$ is a constant map.*

Proof. If $\delta_\infty(r, r') = 0$, then, Proposition 12.1.9 shows that for all x and x' in X, we have

$$B_r(x) - B_{r'}(x) = B_r(x') - B_{r'}(x'),$$

which shows that the map $B_r - B_{r'}$ is constant. □

12.2 Co-rays

Definition 12.2.1 (Geodesic paths converging to a geodesic ray). Let X be a metric space. For any nonnegative integer n, let $\gamma_n\colon [0, a_n] \to X$ be a geodesic path. We say that γ_n *converges to a geodesic ray* $r\colon [0, \infty[\to X$ if $a_n \to \infty$ as $n \to \infty$ and if for every $T \in [0, \infty[$ the sequence of maps $(\gamma_{n|[0,T]})$, which is defined for n large enough, converges uniformly to the map $r_{|[0,T]}$.

For example, the sequence (γ_n) of geodesics that we defined in Proposition 10.1.7 above converges to the geodesic ray r.

Proposition 12.2.2. *Let X be a metric space, let $r\colon [0, \infty[\to X$ be a geodesic ray and let $\gamma_n\colon [0, a_n] \to X$ $(n \geq 0)$ be a sequence of geodesic paths in X. Then, the following two properties are equivalent:*

(i) *the sequence γ_n converges to r as $n \to \infty$;*

(ii) *for every $n \geq 0$, if $\gamma'_n\colon [0, \infty[\to X$ be the map defined for $n \geq 0$ by*

$$\gamma'_n = \begin{cases} \gamma_n(t) & \text{if } 0 \leq t \leq a_n, \\ a_n & \text{if } t \geq a_n, \end{cases}$$

then the sequence (γ'_n) converges to r with respect to the topology of uniform convergence on compact sets of $[0, \infty[$.

Proof. The result follows easily from the definitions. □

Definition 12.2.3 (Co-ray). Let X be a metric space, let $r\colon [0,\infty[\to X$ be a geodesic ray and let x be a point in X. A geodesic ray $r'\colon [0,\infty[\to X$ is said to be a *co-ray to r starting at x* if there exists a sequence of real numbers $(t_n)_{n\geq 0}$ tending to infinity as $n\to\infty$, and for each $n\geq 0$, a geodesic path $\gamma_n\colon [0,a_n]\to X$ satisfying $\gamma_n(a_n)=r(t_n)$, such that $\gamma_n(0)\to x$ and such that the sequence (γ_n) of geodesic paths converges to the geodesic ray r' as $n\to\infty$

It is clear from this definition that any geodesic ray r is a co-ray to itself. We shall see below that if r' is a co-ray to r, then any subray r_0 of r' is a co-ray to r starting at $r_0(0)$.

It is also not true that a co-ray to a given ray r is determined by its initial point, as is shown in Example 12.2.6 (i) below. Likewise, it is not true in the general case that being a co-ray is a symmetric relation; see Example 12.2.6 (iv) below.

Proposition 12.2.4. *Let X be a complete geodesic metric space. Then, for every geodesic ray $r\colon [0,\infty[\to X$ and for every point x in X, there exists a co-ray to r starting at x.*

Proof. We take any sequence $(t_n)_{n\geq 0}$ of real numbers that tends to infinity, and for each $n\geq 0$ a geodesic γ_n joining x to $r(t_n)$. By the proof of Proposition 10.1.7, the sequence γ_n converges to a geodesic ray r', which, by definition, is a co-ray to r. □

Proposition 12.2.5. *Let X be a metric space, let $r\colon [0,\infty[\to X$ be a geodesic ray and let $r'\colon [0,\infty[\to X$ be a co-ray to r. Then, any subray of r' is a co-ray to r.*

Proof. Let r_0 be a subray of r' and let $r_0(0)=r(T)$. Let (x_n) be a sequence of points converging to $r'(0)$, let (t_n) be a sequence of nonnegative real numbers tending to infinity and for each $n\geq 0$ let $\gamma_n\colon [0,|x_n-r(t_n)|[\to X$ be a geodesic path joining x_n to $r(t_n)$ such that the sequence (γ_n) converges to r' as $n\to\infty$.

The point $x'_n=\gamma_n(T)$ is defined for n large enough, and since γ_n converges to r', the sequence (x'_n) converges to $r'(T)=r_0(0)$ as $n\to\infty$. This implies that the sequence of geodesic paths $\gamma'_n\colon [0,|x'_n-r(t_n)|[\to X$, defined for n large enough by

$$\gamma'_n(t)=\gamma_n(|x_n-x'_n|+t),$$

converges to the geodesic ray r_0, which shows that r_0 is a co-ray to r. □

Examples 12.2.6 (Co-rays).

(i) *Non-uniqueness of co-rays.* It is possible to have more than one co-ray to a geodesic ray starting at a given point; it is easy to construct such examples in non-uniquely geodesic spaces. For instance, in the simplicial graph represented in Figure 12.2 (the "bi-infinite horizontal ladder"), if r is a geodesic ray whose image is entirely contained in the upper level, then, for any point x on the lower level, any geodesic ray starting at x whose image coincides with the image of r except for

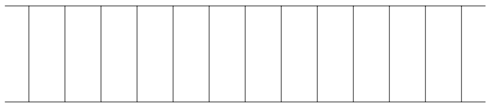


Figure 12.2. A space in which there are several co-rays to the same ray that start at the same point (Example 12.2.6 (i)).

a compact subsegment is a co-ray to r starting at x. Thus, it is clear that there are infinitely many such co-rays starting at x. One can construct in the same way examples of different co-rays starting at the same point in the space $\mathbb{R}^n$ equipped with the ℓ^1 norm.

(ii) *Co-rays in* $\mathbb{E}^n$. It is easy to see, from Euclidean geometry, that if $r\colon [0, \infty[\to \mathbb{E}^n$ is an geodesic ray and if x is a point in $\mathbb{E}^n$, then there is a unique co-ray to r starting at x. In the case where x is on the image of the Euclidean line spanned by r, this co-ray is the Euclidean geodesic ray that starts at x and whose image eventually coincides with the image of r. In the case where x is not on the Euclidean line spanned by r, then the co-ray is the unique geodesic ray starting at x whose image is in the plane spanned by x and the image of r, that is parallel to r in the usual sense of plane Euclidean geometry and that has the same direction as r.

(iii) *Co-rays in* $\mathbb{H}^n$. Let $X = \mathbb{H}^n$. Then, for each geodesic ray $r\colon [0, \infty[\to \mathbb{H}^n$ and for each point x in $\mathbb{H}^n$, there is a unique co-ray to r starting at x. This co-ray is the unique geodesic ray starting at x and converging to the same point on the boundary sphere of $\mathbb{H}^n$. (Again, this is best visualized in the conformal ball model B^n of $\mathbb{H}^n$; see Example 2.4.3 (i) of Chapter 2).

(iv) *Non-symmetry of the co-ray relation.* Let us show that the relation "r_1 is a co-ray to r_2" is not symmetric. Let X be the planar square grid (Figure 12.3), that is, the homogeneous simplicial graph of degree 4, equipped with the length metric for which each edge has length one. Let x, y and z be three successive vertices on the same horizontal, as in Figure 12.3, let $r_1\colon [0, \infty[\to X$ be a vertical geodesic, starting at x and pointing upwards (we use the drawing in Figure 12.3) and let $r_2\colon [0, \infty[\to X$ be the geodesic ray starting at y and obtained by concatenating the geodesic path joining the adjacent vertices y and z with the vertical ray pointing upwards and starting at z (again, we use Figure 12.3). It is clear that r_1 is a co-ray to r_2, but r_2 is not a co-ray to r_1.

The following proposition establishes a relation between co-rays and Busemann functions:

Proposition 12.2.7. *Let X be a metric space, let $r_1\colon [0, \infty[\to X$ be a geodesic ray and let $r_2\colon [0, \infty[\to X$ be a co-ray to r_1. Then, for all t' and t'' in* $[0, \infty[$, *we have*

$$B_{r_1}(r_2(t'')) - B_{r_1}(r_2(t')) = t' - t''.$$

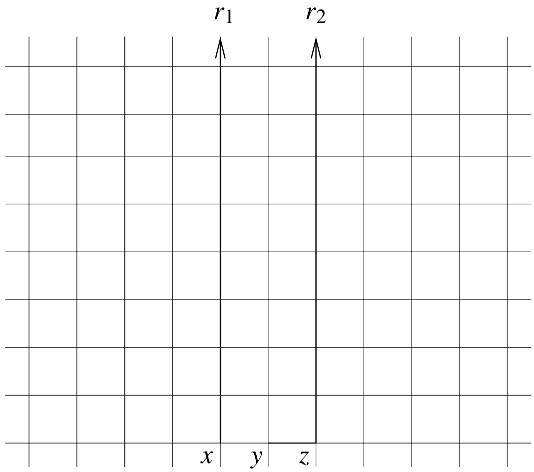


Figure 12.3. r_1 is the vertical geodesic starting at x, r_2 starts at y, passes through z and continues vertically. r_2 is a co-ray to r_1 but r_1 is not a co-ray to r_2 (Example 12.2.6 (iv)).

Proof. Without loss of generality, we can assume that $t'' < t'$. Let us take, as in Definition 12.2.3, two sequences (t_n) and (a_n) of nonnegative real numbers and a sequence $\gamma_n \colon [0, a_n] \to X$ $(n = 0, 1, \dots)$ of geodesic paths satisfying the following:

- $\gamma_n(0) \to r_2(0)$ as $n \to \infty$;
- $\gamma_n(a_n) = r_1(t_n)$ for all $n \geq 0$;
- $r_2(t) = \lim_{n\to\infty} \gamma_n(t)$ for all $t \geq 0$.

By the triangle inequality, we have, for all n large enough,

$$|\gamma_n(t') - r_2(t')| \geq |\gamma_n(t') - r_1(t_n)| - |r_2(t') - r_1(t_n)|$$

and

$$|\gamma_n(t'') - r_2(t'')| \geq |r_2(t'') - r_1(t_n)| - |\gamma_n(t'') - r_1(t_n)|.$$

Adding the two inequalities, we obtain

$$\begin{aligned} |\gamma_n(t') - r_2(t')| + |\gamma_n(t'') - r_2(t'')| &\geq |r_2(t'') - r_1(t_n)| - |r_2(t') - r_1(t_n)| \\ &\quad + \big(|\gamma_n(t') - r_1(t_n)| - |\gamma_n(t'') - r_1(t_n)|\big) \\ &= |r_2(t'') - r_1(t_n)| - |r_2(t') - r_1(t_n)| \\ &\quad - (t' - t''). \end{aligned}$$

In the same way, using the triangle inequality, we can show that

$$\begin{aligned} &|\gamma_n(t') - r_2(t')| + |\gamma_n(t'') - r_2(t'')| \\ &\qquad \leq (t' - t'') - |r_2(t'') - r_1(t_n)| - |r_2(t') - r_1(t_n)|. \end{aligned}$$

Thus, we obtain

$$|\gamma_n(t') - r_2(t')| + |\gamma_n(t'') - r_2(t'')| \\ \leq \big||r_2(t'') - r_1(t_n)| - |r_2(t') - r_1(t_n)| - (t' - t'')\big|.$$

As $n \to \infty$, we have $|\gamma_n(t') - r_2(t')| \to 0$ and $|\gamma_n(t'') - r_2(t'')| \to 0$. By Proposition 12.1.4 (ii), we have

$$\lim_{n\to\infty} (|r_2(t'') - r_1(t_n)| - |r_2(t') - r_1(t_n)|) = B_{r_1}(r_2(t'')) - B_{r_1}(r_2(t')).$$

This proves that

$$B_{r_1}(r_2(t'')) - B_{r_1}(r_2(t')) = t' - t'',$$

which is the desired equality. □

12.3 Horospheres

Definition 12.3.1 (Horosphere). Let X be a metric space, let $r\colon [0, \infty[\to X$ be a geodesic ray and let B_r be the associated Busemann function. A *horosphere with central ray r* is a level set of the function B_r.[1]

If p is an arbitrary point in X, we denote by $S_\infty(r, p)$ the horosphere with central ray r that contains p. Thus, we have

$$S_\infty(r, p) = \{x \in X \text{ such that } B_r(x) = B_r(p)\}.$$

We start with a few elementary properties of horospheres.

Proposition 12.3.2. *If r and r' are two geodesic rays in a metric space satisfying $\delta_\infty(r, r') = 0$, then these two rays have the same associated horospheres.*

Proof. By Corollary 12.1.10, if $\delta_\infty(r, r') = 0$, then $B_r - B_{r'}$ is a constant function. □

Proposition 12.3.3. *Let X be a metric space. Then, the following properties hold:*

(i) *Horosphere are closed subsets of X.*

(ii) *Let $r\colon [0, \infty[\to X$ be a geodesic ray and let x be a point in X. If r_0 is a subray of r, then the two functions B_r and B_{r_0} have the same level sets. More precisely, we have, for all x in X, $S_\infty(r_0, x) = S_\infty(r, x)$.*

[1] For horospheres, Busemann uses the terminology "limit spheres". This is because under appropriate hypotheses on the ambient space, horospheres are indeed limits of spheres.

Proof. The fact that a horosphere is closed follows from the continuity of the function $B_r : X \to \mathbb{R}$ (Proposition 12.1.4 (iii)). To prove (ii), we recall that by Proposition 12.1.6, for each x in X, we have $B_{r_0}(x) = B_r(x) + t_0$. Thus, the maps B_r and B_{r_0} have the same level sets, and furthermore, for all y in X, we have $B_r(y) = B_r(x) \iff B_{r_0}(y) = B_{r_0}(x)$. Therefore, $S_\infty(r_0, x) = S_\infty(r, x)$. This proves Proposition 12.3.3. □

To state the next results, we need the following notation:

For x and y in a metric space X, $S(x, y)$ be the sphere of center x and passing by y. In other words,

$$S(x, y) = \{z \in X \text{ such that } |z - x| = |z - y|.\}$$

Proposition 12.3.4 (Upper limit of spheres contained in horosphere). *Let X be a metric space, let $r : [0, \infty[\to X$ be a geodesic ray and let x be a point in X. Let $(t_n)_{n\geq 0}$ be a sequence of nonnegative real numbers and $(x_n)_{n\geq 0}$ a sequence of points in X such that $t_n \to \infty$ and $x_n \to x$ as $n \to \infty$. Then,*

$$\limsup S\big(r(t_n), x_n\big) \subset S_\infty(r, x).$$

Proof. If y is in $\limsup S\big(r(t_n), x_n\big)$, then, up to replacing the sequence $(r(t_n), x_n)_{n\geq 0}$ by a subsequence, we can find, for each $n \geq 0$, a point y_n in $S\big(r(t_n), x_n\big)$ such that $y_n \to y$ as $n \to \infty$. By Proposition 12.1.4 (ii), we have

$$|B_r(x) - B_r(y)| = \lim_{t_n \to \infty} \big||x - r(t_n)| - |y - r(t_n)|\big|.$$

From the triangle inequality, we obtain

$$|x - r(t_n)| - |y - r(t_n)| \leq \big(|x_n - r(t_n)| - |y_n - r(t_n)|\big) + |x - x_n| + |y - y_n|.$$

Since $y_n \in S\big(r(t_n), x_n\big)$, we have $|x_n - r(t_n)| = |y_n - r(t_n)|$. Therefore,

$$\lim_{t_n \to \infty} \big(|x - r(t_n)| - |y - r(t_n)|\big) \leq \lim_{t_n \to \infty} (|x - x_n| + |y - y_n|) = 0.$$

This implies that y is in $S_\infty(r, x)$, which proves Proposition 12.3.4. □

Following Busemann, we make the following definition:

Definition 12.3.5 (Open horoball). Let X be a metric space, let $r : [0, \infty[\to X$ be a geodesic ray and let x be a point in X. We call the subset of X defined as

$$B_\infty(r, x) = \{y \in X \text{ such that } B_r(y) < B_r(x)\}$$

the *open horoball bounded by the horosphere* $S_\infty(r, x)$. Of course, this is in analogy with the notion of the interior of a sphere $S(x, y)$ of center x and passing by y, which is the open ball $B(x, |x - y|)$ of center x and radius $|x - y|$.

Proposition 12.3.6. *Let $r\colon [0,\infty[\to X$ be a geodesic ray in X. Then, for every $t_1 \geq 0$, the open ball $B\big(r(t_1), t_1\big)$ is contained in the open horoball $B_\infty\big(r, r(0)\big)$.*

Proof. By Proposition 12.1.4 (ii), we have, for all x in X,

$$\begin{aligned} B_r(r(0)) - B_r(x) &= \lim_{t\to\infty} (|r(0) - r(t)| - |x - r(t)|) \\ &= \lim_{t\to\infty} (t_1 - |x - r(t)|). \end{aligned}$$

Since the map $t \mapsto t - |x - r(t)|$ is non-decreasing (Lemma 12.1.1), we obtain, taking $t \geq t_1$,

$$\begin{aligned} B_r(r(0)) - B_r(x) &\geq t - |x - r(t)| \\ &\geq t_1 - |x - r(t_1)|. \end{aligned}$$

Now for $x \in B\big(r(t_1), t_1\big)$, we have $t_1 > |x - r(t_1)|$, which implies $B_r(r(0)) > B_r(x)$, that is, $x \in B_\infty\big(r, r(0)\big)$. □

Lemma 12.3.7. *Let $r\colon [0,\infty[\to X$ be a geodesic ray and let t_1 and t_2 be two real numbers satisfying $0 \leq t_1 \leq t_2$. Then, $B(r(t_1), t_1) \subset B(r(t_2), t_2)$.*

Proof. By the triangle inequality, we have, for all x in $B(r(t_1), t_1)$,

$$\begin{aligned} |x - r(t_2)| &\leq |x - r(t_1)| + r(t_1) - r(t_2)| \\ &= t_1 + t_2 - t_1 = t_2, \end{aligned}$$

that is, $x \in B(r(t_2), t_2)$. □

Corollary 12.3.8. *Let $r\colon [0,\infty[\to X$ be a geodesic ray and let $(t_n)_{n\geq 0}$ be a sequence of nonnegative real numbers tending to infinity. Then, the closed limit of the sequence of subspaces $B\big(r(t_n), t_n\big)$, $n \geq 0$, exists, and it is contained in the horoball $B_\infty(r, r(0))$.*

Proof. The closed limit exists since, by Lemma 12.3.7, the sequence of subsets $B\big(r(t_n), t_n\big)$, $n \geq 0$, is increasing. By Proposition 12.3.6, this closed limit is contained in $B_\infty(r, r(0))$. □

Proposition 12.3.9. *Let r be a geodesic ray in X and let p and q be two points in X. Then, the following two properties are equivalent:*

(i) $B_r(p) - B_r(q) = |p - q| > 0$;

(ii) $q \in B_\infty(r, p)$ *and q is a projection of p on the horosphere* $S_\infty(r, q)$.

Proof. Suppose that $B_r(p) - B_r(q) = |p - q| > 0$. From the definition of an open horoball, since $B_r(p) > B_r(q)$, we have $q \in B_\infty(r, p)$. Let us show that q is a projection of p on $S_\infty(r, q)$. Let x be in $S_\infty(r, q)$. Then, we have $B_r(x) = B_r(q)$. Since B_r is 1-Lipschitz, we have

$$B_r(p) - B_r(q) = B_r(p) - B_r(x) \leq |p - x|.$$

Therefore, we have $|p - q| \leq |p - x|$. This shows that q is a projection of p on $S_\infty(r, q)$. This proves (i) ⇒ (ii).

Let us now prove (ii) ⇒ (i). Suppose that q is a projection of p on $S_\infty(r, q)$ and that $q \in B_\infty(r, p)$. Let $t_0 = B_r(p) - B_r(q)$. We have $t_0 > 0$ since $q \in B_\infty(r, p)$. By Proposition 12.2.4, there exists a co-ray $s\colon [0, \infty[\to X$ from p to r. Let $m = s(t_0)$. By Proposition 12.2.7, we have

$$B_r(p) - B_r(s(t_0)) = B_r(s(0)) - B_r(s(t_0)) = t_0.$$

Therefore, $B_r(s(t_0)) = B_r(q)$, which implies $s(t_0) \in S_\infty(r, q)$. Now since q is a projection of p on $S_\infty(r, q)$, we have

$$t_0 = |s(0) - s(t_0)| = |p - s(t_0))| \geq |p - q|.$$

On the other hand, by Proposition 12.1.4 (iii), we have $t_0 = B_r(p) - B_r(q) \leq |p - q|$. Summing up, we have $t_0 = |p - q|$. This completes the proof of Proposition 12.3.9. □

Corollary 12.3.10. *Let r be a geodesic ray in X and let s be a co-ray to r. Then, for any $t > 0$, the point $s(t)$ is a projection of $r(0)$ on $S_\infty(r, s(t))$.*

Proof. By Proposition 12.2.7, we have

$$B_r(s(0)) - B_r(s(t)) = s(t) - s(0) = t > 0.$$

Thus, the conclusion follows from Proposition 12.3.9. □

In conclusion, let us mention a few problems.

It would be most interesting to develop the theory of co-rays and limit spheres in Teichmüller space, equipped with its various metrics: the Teichmüller metric, the Weil–Petersson metric and Thurston's "stretch metric",. The last metric is not as much popular as the other two; it has been introduced by Thurston in the preprint [135]. Let us briefly recall its definition. Given two hyperbolic metrics g_1 and g_2 on the surface S, one first defines their "distance" as

$$\lambda(g_1, g_2) = \inf_f \log\{\mathrm{lip}(f)\},$$

where the infimum is taken over all Lipschitz maps $f\colon (S, g_1) \to (S, g_2)$ that are isotopic to the identity and where

$$\mathrm{lip}(f) = \sup_{x,y\in S,\ x\neq y} \left\{ \frac{d_{g_2}(f(x), f(y))}{d_{g_1}(x, y)} \right\}$$

is the Lipschitz constant of the map f. It is easy to see that the map λ induces a map defined on pairs of isotopy classes of metrics, that is, a map on $\mathcal{T}_g \times \mathcal{T}_g$. This "stretch metric" on $\mathcal{T}_g$ is not symmetric, and it is interesting to note here that precisely, Busemann developed his theory (for instance in his paper [25]) for non-symmetric metrics. Particularly interesting points to study would be the following:

• the visual boundary of Teichmüller space equipped with Thurston's stretch metric;[2]

• the relation between horospheres and the symplectic (Weil–Petersson) geometry of Teichmüller space;

• the relation of co-rays and horospheres with measured foliations and Thurston's boundary of Teichmüller space;

• the relation between Thurston's earthquake paths and horospheres.

The last questions is interesting for all three metrics on Teichmüller space. In fact, the whole Busemann geometry of Teichmüller space is still to be explored. We already mentioned the works of Kravetz, Masur and Kerckhoff in [93], [84] and [99] that concern the Teichmüller metric. Other related material is contained in the paper [95] by J. McCarthy and the author.

Notes on Chapter 12

The theories of co-rays, Busemann functions and horospheres in G-spaces are developed by Busemann in [24], Chapter III, [25], Chapter III and [28], §22. In their earliest version (see [25]), Busemann functions were called λ-functions, and in later versions, Busemann calls them α-functions.

There are important results on co-rays and horospheres in metric spaces that use Busemann's hypothesis on the uniqueness of prolongation of geodesics, a hypothesis that we did not make in this book. For instance, in a G-space in the sense of Busemann (see the definition in the Notes to Chapter 2), the converse of Proposition 12.2.7 is also true. In other words, the equality

$$B_{r_1}(r_2(t'')) - B_{r_1}(r_2(t')) = t' - t''$$

holds for all t' and t'' in $[0, \infty[$ if and only if r_2 is a co-ray to r_1. This is proved by Busemann in [28], pp. 134 and 136. Another interesting result is the fact that if X is a G-space in which every geodesic path can be extended to a geodesic line, then the inclusion

$$\limsup S\big(r(t_n), x_n\big) \subset S_\infty(r, x)$$

[2] In the papers [116] and [117], there is some information on the visual boundary of Teichmüller space with respect to the stretch metric. In these papers, we describe the limits of certain stretch geodesics as points on Thurston's boundary of Teichmüller space.

in Proposition 12.3.4 above becomes an equality, with the limit sup being an honest limit:

$$\lim S\big(r(t_n), x_n\big) = S_\infty(r, x).$$

(see [28], Corollary 22.5).

References

[1] S. B. Alexander and R. L. Bishop, The Hadamard–Cartan Theorem in locally convex metric spaces, *Enseign. Math.* 36 (1990), 309–320. 219, 224, 225, 226, 228

[2] A. D. Alexandrov, A theorem on triangles in a metric space and some of its maps, *Trudy Mat. Inst. Steklov.* 38 (1951), 5–23. 7

[3] A. D. Alexandrov, Selected works: Part I, *Selected scientific papers*, ed. by Yu. G. Reshetnyak and S. S. Kutateladze, Gordon and Breach, Amsterdam 1996. 7, 8

[4] A. D. Alexandrov and V. A. Zalgaller, *Intrinsic geometry of surfaces*, Transl. Math. Monogr. 15, Amer. Math. Soc., Providence, RI, 1967. 7

[5] C. Arzelà, Funzioni di linee, *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur.* (4) 5 (1883), 342–348. 27

[6] G. Ascoli, Le curve limiti di une varietà data di curve, *Atti Accad. Lincei Mem. Cl. Sci. Fis. Mat. Natur.* (3) 18 (1883), 521–586. 27

[7] W. Ballmann, *Lectures on spaces of nonpositive curvature*, DMV Sem. 25, Birkhäuser Verlag, Basel 1995. 77, 228

[8] W. Ballman, M. Gromov and V. Schröder, Manifolds of nonpositive curvature, Progr. Math. 61, Birkhäuser, Boston 1985. 241

[9] S. Banach, Sur les opérations dans les ensembles abstraits and leurs applications aux équations intégrales, *Fund. Math.* 3 (1922), 133–180. 101

[10] T. J. Barth, Some counterexamples concerning intrinsic distances, *Proc. Amer. Math. Soc.* 66 (1977), 49–53. 52

[11] A. F. Beardon, The geometry of discrete groups, Grad. Texts in Math. 91, Springer-Verlag, New York 1983. 39

[12] E. Beltrami, Risoluzione del problema: riportare i punti di una superficie sopra un piano in modo che le linee geodetiche vengano rappresentate da linee rette, Ann. Mat. Pura Appl. (1) 7 (1865) 185–204; reprinted in *Opere Matematiche*, Tomo primo, Ulrico Hoepli, Milano 1902, 262–280. 158

[13] L. Bers, An extremal problem for quasiconformal mappings and a theorem by Thurston, *Acta Math.* 141 (1978), 73–98. 247

[14] R. Benedetti and C. Petronio, *Lectures on hyperbolic geometry*, Universitext, Springer-Verlag, Berlin 1992. 39

[15] L. Bers, Quasiconformal mappings and Teichmüller's theorem, in *Analytic functions* (Princeton, 1957), ed. by R. Nevanlinna et al., Princeton Math. Ser. 24, Princeton Univ. Press, Princeton, NJ, 1960, 89–119. 60

[16] B. Beauzamy, *Introduction to Banach spaces and their geometry*, 2nd ed., North-Holland Math. Stud. 68, Amsterdam 1985. 186

[17] R. L. Bishop and B. O'Neill, Manifolds of negative curvature, *Trans. Amer. Math. Soc.* 145 (1969), 1–49. 204, 205, 260

[18] L. M. Blumenthal, *Theory and applications of distance geometry*, Oxford University Press, Oxford 1953; reprinted as corr. 2nd ed. by Chelsea Publ. Co., New York 1970. 10

[19] J. M. Borwein and R. C. O'Brien, Cancellation characterizes convexity, *Nanta Math.* 11 (1978), 100–102. 131

[20] M. Bridson and A. Haefliger, *Metric spaces of non-positive curvature*, Grundlehren Math. Wiss. 319, Springer-Verlag, Berlin 1999. 228, 241

[21] L. E. J. Brouwer, Über eineindeutige, stetige Transformationen von Flächen in sich, *Math. Ann.* 69 (1910), 176–180, 592, ibid. 79 (1919), 403; reprinted as pp. 244–249 in *Collected Works*, vol. 2, ed. by H. Freudenthal, North-Holland, Amsterdam 1976. 101

[22] H. Brunn, Über das durch eine beliebige endliche Figur bestimmte Eigebilde, in *Festschrift Luwig Boltzmann* gewidmet zum sechzigsten Geburtstag 20. Februar 1904, Johann Ambrosius Barth, Leipzig 1904, 94–104. 137

[23] H. Brunn, Über Kerneigebiete, *Math. Ann.* 73 (1913), 436–440. 136

[24] H. Busemann, *Metric methods in Finsler spaces and in the foundations of geometry*, Ann. of Math. Stud. 8, Princeton University Press, Princeton 1942. 147, 273

[25] H. Busemann, Local metric geometry, *Trans. Amer. Math. Soc.* 56 (1944), 200–274. 101, 108, 110, 273

[26] H. Busemann, Spaces with nonpositive curvature, *Acta Math.* 80 (1948), 259–310. 7, 204

[27] H. Busemann, The foundations of Minkowskian geometry, *Comment. Math. Helv.* 24 (1950), 156–187. 147

[28] H. Busemann, *The geometry of geodesics*, Pure Appl. Math. 6, Academic Press Inc., New York 1955. 7, 10, 50, 69, 77, 101, 116, 121, 123, 124, 147, 157, 158, 173, 176, 191, 224, 228, 239, 241, 249, 273, 274

[29] H. Busemann, Length-preserving maps, *Pacific J. Math.* 14 (1964), 457–477. 82, 83, 87, 88

[30]] H. Busemann, *Recent synthetic differential geometry*, Ergeb. Math. Grenzgeb. (2) 54, Springer-Verlag, Berlin 1970. 62, 88, 158

[31] H. Busemann and P. J. Kelly, *Projective geometry and projective metrics*, Pure Appl. Math. 3, Academic Press Inc., New York 1953. 157

[32] H. Busemann and B. B. Phadke, Peakless and monotone functions on G-spaces, *Tsukuba J. Math.* 7 (1983), 105–135. 205

[33] C. Carathéodory, Über das Schwarzsche Lemma bei analytischen Funktionen von zwei komplexen Veränderlichen, *Math. Ann.* 97 (1926), 76–98; reprinted as pp. 132–159 in *Gesammelte Mathematische Schriften*, Bd. IV, C. H. Beck, München 1956. 41

[34] C. Carathéodory, Über die Geometrie der analytischen Abbildungen, die durch analytische Funktionen von zwei komplexen Veränderlichen vermittelt werden, Abh. Math. Sem. Univ. Hamburg. 6 (1928), 96–145; reprinted as pp. 167–227 in *Gesammelte Mathematische Schriften*, Bd. IV, C. H. Beck, München 1956. 41

[35] E. Cartan, *La théorie des groupes continus et la géométrie*, Encyclop. Sc. math., édition française d'après l'article de Fano, 1915; reprinted in *Œuvres complètes*, partie III, vol. 2, Géométrie différentielle (Suite), Gauthier-Villars, Paris 1955, 1727–1861. 79

[36] E. Cartan, *Leçons sur la géométrie des espaces de Riemann*, Second augmented edition, Gauthier-Villars, Paris 1963 (1st ed. 1928). 4, 27, 77, 203, 210, 228, 240

[37] A. Cauchy, *Sur les polygones et les polyèdres*, *Journal de l'École Polytechnique*, XVI[e] Cahier, Tome IX, p. 87 (1813), reprinted as pp. 26–38 in *Œuvres complètes*, II[e] Série, Tome I, Gauthier-Villars, Paris 1905. 9

[38] J. Cheeger and D. Gromoll, On the structure of complete manifolds of nonnegative curvature, *Ann. of Math.* (2) 96 1972), 413–443. 205

[39] G. Choquet, *Cours d'analyse*, Tome II: Topologie, espaces topologiques et espaces métriques, 2nd ed., Masson, Paris 1969. 10

[40] S. Cohn-Vossen, Kürzeste Wege und Totalkrümmung auf Flächen, *Compositio Math.* 2 (1935), 69–133. 205

[41] S. Cohn-Vossen, Existenz kürzester Wege, *Compositio Math.* 3 (1936), 441–452. 33, 77

[42] M. Culler and J. W. Morgan, Group actions on $\mathbb{R}$-trees, *Proc. London Math. Soc.* (3) 155 (1987), 571–604. 246

[43] G. Daskalopoulos and R. Wentworth, Classification of Weil–Petersson isometries, *Amer. J. Math.* 125 (2003), 941–975. 248

[44] C. J. Earle and J. Eells, On the differential geometry of Teichmüller spaces, *J. Anal. Math.* 19 (1967), 35–52. 61

[45] P. Eberlein and B. O'Neill, Visibility manifolds, *Pacific J. Math.* 46 (1973), 45–109. 240

[46] W. Fenchel, Convexity through the ages, in *Convexity and its applications*, ed. by P. M. Gruber and J. M. Wills, Birkhäuser, Basel 1983, 120–130. 9

[47] P. Finsler, *Über Kurven und Flächen in allgemeinen Räumen*, Dissertation, Göttingen 1918. 40

[48] W. J. Floyd, Group completions and limit sets of Kleinian groups, *Invent. Math.* 57 (1980), 205–218. 125, 126

[49] M. Fréchet, Sur quelques points de calcul fonctionnel, *Rend. Circ. Mat. Palermo* 22 (1906), 1–74. 27, 33

[50] H. Freudenthal and W. Hurewicz, Dehnungen, Verkürzungen, Isometrien, *Fund. Math.* 26 (1936), 120–122. 88

[51] C. F. Gauss, *Disquisitiones generales circa superficies curvas*, Commentationes Societatis Regiae Scientiarum Gottingensis Recentiores, vol. VI, Gottingae 1823–1827; Gauss Werke, Band IV, zweiter Abdruck, herausgegeben von der Königlichen Gesellschaft der Wissenschaften zu Göttingen1880, 217–258; published with an English translation with commentaries in: Astérisque 62, Societé Math. France, Paris 1976. 3

[52] R. E. Greene and H. Wu, Integrals of subharmonic functions on manifolds of nonnegative curvature, *Invent. Math.* 27 (1974), 265–298. 205

[53] R. E. Greene and H. Wu, C^∞ convex functions and manifolds of positive curvature, *Acta Math.* 137 (1976), 209–245. 205

[54] D. Gromoll and W. Meyer, On complete open manifolds of positive curvature, *Ann. of Math.* (2) 90 (1969), 75–90. 205

[55] M. Gromov, Hyperbolic manifolds, groups and actions, in *Riemann surfaces and related topics: proceedings of the 1978 Stony Brook conference*, ed. by I. Kra and B. Maskit, Ann. of Math. Stud. 97, Princeton Univ. Press, Princeton, NJ, 1981, 183–213. 210

[56] M. Gromov and S. M. Bates, *Metric structures for Riemannian and non-Riemannian spaces*, with appendices by M. Katz, P. Pansu and S. Semmes, ed. by J. Lafontaine and P. Pansu, Progr. Math. 152, Birkhäuser, Boston 1999. 48

[57] P. M. Gruber and G. Lettl, Isometries of the space of convex bodies in Euclidean space, *Bull. London Math. Soc.* 12 (1980), 455–462. 157

[58] R. Haar, Der Massbegriff in der Theorie der kontinuierlichen Gruppen, *Ann. of Math.* (2) 34 (1933), 147–169. 203

[59] J. Hadamard, Etude sur les propriétés des fonctions entières et en particulier d'une fonction considérée par Riemann, *J. Math. Pures Appl.* 58 (1893), 171–215. 176

[60] J. Hadamard, Sur les lignes géodésiques des surfaces à courbures opposées, *C. R. Acad. Sci. Paris* 124 (1897), 149. 3

[61] J. Hadamard, Sur la courbure dans les espaces à plus de deux dimensions, Procès-verbaux des séances de la Société des sciences physiques et naturelles de Bordeaux (1897–1898), 85–87. 4

[62] J. Hadamard, Les surfaces à courbures opposées et leurs lignes geodesiques, J. Math. Pures Appl. (5) 4 (1898), 27–74. 1, 3, 228, 229, 235

[63] G. H. Hardy, J. E. Littlewood and G. Pólya, *Inequalities*, Cambridge University Press, Cambridge 1934. 74

[64] P. de la Harpe, On Hilbert's metric for simplices, in *Geometric group theory* (Sussex, 1991), vol. 1, ed. by G. A. Niblo and M. A. Roller, Lond. Math. Soc. Lecture Note Ser. 181, Cambridge University Press, Cambridge 1993, 97–119. 152

[65] P. K. Hotchkiss, The boundary of a Busemann space, *Proc. Amer. Math. Soc.* 125 (1997), 1903–1912. 229, 236, 239

[66] F. Hausdorff, Mengenlehre, de Gruyter, Leipzig 1927 and several later editions. (English translation: Set theory, Chelsea 1957 and 1962.) 112, 115, 124

[67] T. L. Heath, *The works of Archimedes*, Cambridge University Press, Cambridge 1897. 31, 156

[68] D. Hilbert, Über die stetige Abbildung einer Linie auf ein Flächenstück, *Math. Ann.* 38 (1891), 459–460. 32

[69] D. Hilbert, Über die gerade Linie als kürzeste Verbindung zweier Punkte, *Math. Ann.* 46 (1895), 91–96. 151, 153, 154, 158

[70] D. Hilbert, *Grundlagen der Geometrie*, B. G. Teubner, Stuttgart 1899, several later editions revised by the author, and several translations. (First French translation: L. Laugel, Gauthier-Villars, Paris 1900, and first English translation by Townsend, Chicago 1902. Recent French translation by P. Rossier, Dunod, Paris 1971.) 158

[71] D. Hilbert, Mathematische Probleme, *Nachr. Akad. Wiss. Göttingen, Math. Phys. Kl.* 1900, 253–297 and Archiv Math. u. Phys. 1 (1901), 43–63 and 213–237; reprinted as pp. 290–329 in *Gesammelte Abhandlungen*, Bd. III, 2nd ed., Springer-Verlag, Berlin 1970. (English translation: Bulletin of Amer. Math. Soc. 8 (1902), 437–479.) 158

[72] O. Hölder, Über einen Mittelwertsatz, *Nachr. Ges. Wiss. Goettingen* 1889, 38–47. 176

[73] H. Hopf and W. Rinow, Über den Begriff der vollständigen differentialgeometrischen Fläche, *Comm. Math. Helv.* 3 (1932), 209–225. 77

[74] J. H. Hubbard and and H. Masur, Quadratic differentials and foliations, *Acta Math.* 142 (1979), 221–274. 212

[75] V. I. Istrăţescu, *Strict convexity and complex strict convexity*, Lecture Notes in Pure and Appl. Math. 89, Marcel Dekker, Inc., New York 1984. 186

[76] J. L. W. V. Jensen, Om konvexe Funktioner og Uligheder mellem Middelvaerdier, *Nyt. Tidsskr Math. B.* 16 (1905), 49–69. 176

[77] J. L. W. V. Jensen, Sur les fonctions convexes et les inégalités entre les valeurs moyennes, *Acta Math.* 30 (1906), 175–193. 176

[78] F. Jongmans, De l'art d'être à bonne distance des ensembles dont la décomposition atteint un stade avancé, Bull. Soc. Roy. Sci. Liège 48 (1979), 237–261. 157

[79] C. Jordan, *Cours d'analyse de l'Ecole Polytechnique*, Gauthier-Villars, Paris 1883–1887. 32

[80] J. Jost, Nonpositive curvature: Geometric and analytic aspects, Lectures Math. ETH Zürich, Birkhäuser, Basel 1997. 203, 205

[81] P. J. Kelly and E. G. Straus, Curvature in Hilbert geometries, *Pacific J. Math.* 8 (1958), 119–125. 191

[82] P. J. Kelly and E. G. Straus, Curvature in Hilbert geometries II, *Pacific J. Math.* 25 (1968), 559–552. 191

[83] J. Kepler (Ioannis Keppleri), Harmonices Mundi, Lincii Austriae: Sumptibus Godofredi Tampachii Bibl. Francof., 1619. (French translation by J. Peyroux, Librairie A. Blanchard, Paris 1978.) 156

[84] S. P. Kerckhoff, The asymptotic geometry of Teichmüller space, *Topology* 19 (1980), 23–41. 61, 65, 240, 273

[85] S. P. Kerckhoff, The Nielsen realization problem, *Ann. of Math.* (2) 117 (1983), 235–265. 206

[86] S. P. Kerckhoff, Lines of minima in Teichmüller space, *Duke Math. J.* 65 (1992), 187–213. 207, 209

[87] M. A. Khamsi, On normal structure, fixed-point property and contractions of type (γ), *Proc. Amer. Math. Soc.* 106, (1989), 995–1001. 186

[88] F. Klein, Über die Transformation der elliptischen Functionen und die Auflösung der Gleichungen fünften Grades, *Math. Ann.* 14 (1879), 111–172. 241, 245

[89] F. Klein, Gutachten, betreffend den dritten Band der Theorie der Transformationsgruppen von S. Lie anlässlich der ersten Vertheilung des Lobatschewsky-Preises, *Math. Ann.* 50 (1898), 581–600. 32

[90] S. Kobayashi, Invariant distances on complex manifolds and holomorphic mappings, *J. Math. Soc. Japan* 19 (1967), 460–480. 42

[91] S. Kobayashi, *Hyperbolic complex spaces*, Grundlehren Math. Wiss. 318, Springer-Verlag, Berlin 1998. 41, 52

[92] H. von Koch, Sur une courbe continue sans tangente, obtenue par une construction géométrique élémentaire, *Arkiv Mat.* 1 (1904), 681–704. 16

[93] S. Kravetz, On the geometry of Teichmüller spaces and the structure of their modular groups, *Ann. Acad. Sci. Fenn. Ser. A I Math.* 278 (1959), 1–35. 206, 273

[94] M. R. Linch, *On metrics on Teichmüller space*, Ph.D. Thesis, Columbia University, New York 1971. 206

[95] J. D. McCarthy and A. Papadopoulos, Fundamental domains in Teichmüller space, *Ann. Acad. Sci. Fenn. Math.* 21 (1996), 151–166. 273

[96] J. D. McCarthy and A. Papadopoulos, The visual sphere of Teichmüller space and a theorem of Masur–Wolf, *Ann. Acad. Sci. Fenn. Math.* 24 (1999), 147–154. 240

[97] S. Mandelbrojt, Notice nécrologique sur Maurice Fréchet, *C. R. Acad. Sci. Paris* 277 (1973), Vie académique, 73–76. 33

[98] W. S. Massey, *A basic course in algebraic topology*, Grad. Texts in Math. 127, Springer-Verlag, New York 1991. 96

[99] H. Masur, On a class of geodesics in Teichmüller space, *Ann. of Math.* (2) 102, (1975), 205–221. 206, 273

[100] H. Masur and M. Wolf, The Weil–Petersson isometry group, *Geom. Dedicata* 93 (2002), 177–190. 248

[101] C. T. McMullen, Complex earthquakes and Teichmüller theory, *J. Amer. Math. Soc.* 11 (1998), 283–320. 43

[102] K. Menger, Grundzüge einer Theorie der Kurven, *Math. Ann.* 95 (1925), 277–306; reprinted as pp. 179–208 in vol. I of [108]. 33

[103] K. Menger, Untersuchungen über allgemeine Metrik. Untersuchungen I–III, *Math. Ann.* 100 (1928), 75–163; reprinted as pp. 237–325 in vol. 1 of [108]. 33, 69, 77

[104] K. Menger, Untersuchungen über allgemeine Metrik. IV, *Math. Ann.* 103 (1930), 466–501; reprinted as pp. 333–368 in vol. 1 of [108].

[105] K. Menger, Bericht über metrische Geometrie, *Jahresber. Deutsch. Math.-Verein.* 40 (1931), 201–219. 6

[106] K. Menger, Metrische Geometrie und Variationsrechnung, *Fund. Math.* 25 (1935), 441–458. 33

[107] K. Menger, Die metrische Methode in der Variationsrechnung, Ergebnisse eines mathematischen Kolloquiums, Heft 8, 1–32, Wien 1937. 33

[108] K. Menger, *Selecta mathematica*, vol. 1 and 2, ed. by B. Schweizer, A. Sklar, K. Sigmund, P. Gruber, E. Hlawka, L. Reich and L. Schmetterer, Springer-Verlag, Wien 2002 and 2003. 4, 280

[109] K. Menger, The formative years of Abraham Wald and his work in geometry, *Ann. Math. Stat.* 23 (1952), 14–20. 6

[110] H. Minkowski, *Geometrie der Zahlen*, B. G. Teubner, Leipzig 1896 and 1910; reprinted by Chelsea, New York 1953. 74, 156

[111] H. Minkowski, Theorie der konvexen Körper, insbesondere Bergründung ihres Oberflächenbegriffs, *Gesammelte Abhandlungen*, Bd. II, B. G. Teubner, Leipzig 1911, 131–229; reprinted in one volume by Chelsea, New York 1967. 144, 156

[112] J. W. Morgan and P. B. Shalen, Valuations, trees, and degenerations of hyperbolic structures I, *Ann. of Math.* (2) 120 (1984), 401–476. 65, 246

[113] M. Morse, A one-to-one representation of geodesics on a surface of negative curvature, *Amer. J. Math.* 43 (1921), 33–51. 2

[114] J. R. Munkres, *Topology*, 2nd ed., Prentice Hall, Upper Saddle River, NJ, 2000. 96, 238, 239

[115] S. B. Myers, Riemannian manifolds in the large, *Duke Math. J.* 1 (1935), 39–49. 77

[116] A. Papadopoulos, Sur le bord de Thurston de l'espace de Teichmüller d'une surface non compacte, *Math. Ann.* 282 (1988), 353–359. 273

[117] A. Papadopoulos, On Thurston's boundary of Teichmüller space and the extension of earthquakes, *Topology Appl.* 41 (1991), 147–177. 273

[118] B. B. Phadke, A triangular world with hexagonal circles, *Geom. Dedicata* 3 (1974/75), 511–520. 152

[119] G. Peano, *Applicazioni geometriche del calcolo infinitesimale*, Bocca, Torino 1887. 32

[120] G. Peano, Sur une courbe, qui remplit toute une aire plane, *Math. Ann.* 36 (1890), 157–160; reprinted as pp. 110–114 in *Opere Scelte*, vol. I, Edizioni Cremonese, Roma 1957. 32

[121] G. Peano, Sulla definizione dell'area d'una superficie, Rendiconti della R. Accademia dei Lincei, Serie 4a, vol. VI, 1^o Sem., A. (1890), 54–57; reprinted as pp. 102–106 in *Opere Scelte*, vol. I, Edizioni Cremonese, Roma 1957. 10, 31

[122] H. Poincaré, Sur les groupes Kleinéens, C. R. Acad. Sci. Paris 93 (1881), 44–46. 39, 79

[123] E. Rakotch, A note on α-locally contractive mappings, *Bull. Res. Council Israel Sect. F* 10 (1962), 188–191. 101

[124] J. G. Ratcliffe, Foundations of hyperbolic manifolds, Grad. Texts in Math. 149, Springer-Verlag, New York 1994. 39

[125] H. L. Royden, Automorphisms and isometries of Teichmüller space, in *Advances in the theory of Riemann surfaces* (Stony Brook, 1969), Ann. of Math. Stud. 66, Princeton Univ. Press, Princeton, NJ, 1971, 369–383. 61, 247

[126] H. L. Royden, Remarks on the Kobayashi metric, in *Several Complex Variables II* (Maryland, 1970), ed. by J. Horváth, Lecture Notes in Math. 185, Springer-Verlag, Berlin 1971, 125–137. 42, 101

[127] R. Schneider, Pairs of convex bodies with unique joining metric segment, *Bull. Soc. Roy. Sci. Liège* 50 (1981), 5–7. 157

[128] J. P. Serre, *Arbres, amalgames,* SL_2, Astérisque 46, Société Mathématique de France, Paris 1977. English translation: *Trees*, Springer-Verlag, Berlin 1980. 247

[129] O. Stolz, Grundzüge der Differential- und Integralrechnung. Erster Theil: Relle Veränderliche und Functionen, B. G. Teubner, Leipzig 1893. 176

[130] H. Tanigawa, Grafting, harmonic maps and projective structures on surfaces, J. Differential Geom. 47 (1997), 399–419. 43

[131] O. Teichmüller, Extremale quasikonforme Abbildungen und quadratische Differentiale, *Abh. Preuss. Akad. Wiss. Math.-Naturw. Kl.* 22 (1939), 1–197; reprinted as pp. 335–531 in [133]. 60

[132] O. Teichmüller, Bestimmung der extremalen quasikonformen Abbildungen bei geschlossenen orientierten riemannschen Flächen, *Abh. Preuss. Akad. Wiss., Math.-Naturw. Kl.* 4 (1943), 1–42; reprinted as pp. 635–676 in [133]. 65

[133] O. Teichmüller, *Gesammelte Abhandlungen/Collected Papers*, ed. by L. V. Ahlfors and F. W. Gehring, Springer-Verlag, Berlin, New York 1982. 65, 282

[134] W. P. Thurston, The geometry and topology of three-manifolds, Mimeographed notes, Princeton Math. Dept., 1978–1979. 39, 241, 246

[135] W. P. Thurston, Minimal Stretch maps between hyperbolic surfaces, preprint 1979. 272

[136] W. P. Thurston, On the geometry and dynamics of diffeomorphisms of surfaces, *Bull. Amer. Math. Soc. (N.S.)* 19 (1988), 417–431. 247

[137] W. P. Thurston, Three-dimensional geometry and topology, vol. 1, ed. by Silvio Levy, Princeton Math. Ser. 35, Princeton University Press, Princeton, NJ, 1997. 39, 246

[138] J. Tits, A "Theorem of Lie–Kolchin" for trees, in *Contributions to algebra, a collection of papers dedicated to Ellis Kolchin*, ed. by H. Bass, P. J. Cassidy and J. Kovacic, Academic Press, New York 1977, 377–388. 65

[139] F. A. Valentine, *Convex sets*, McGraw-Hill, New York 1964. 156, 157

[140] J. P. Vigué, La distance de Carathéodory n'est pas intérieure, *Results Math.* 6 (1983), 100–104. 52

[141] A. Wald, Sur la courbure des surfaces, *C. R. Acad. Sci. Paris* 201 (1935), 918–920. 6

[142] J. Wallis, *Arithmetica Infinitorum*, 1656, also in *Opera mathematica*, Oxoniæ : E Theatro Sheldoniano, 1699. 32

[143] A. Weil, *L'intégration dans les groupes topologiques et ses applications*, Hermann, Paris 1940. 203

[144] S. A. Wolpert, The Fenchel–Nielsen deformation, *Ann. of Math.* (2) 115 (1982), 501–528. 207

[145] S. A. Wolpert, On the symplectic geometry of deformations of a hyperbolic surface, *Ann. of Math.* (2) 117 (1983), 207–234.

[146] S. A. Wolpert, Geodesic length functions and the Nielsen problem, *J. Differential Geom.* 25 (1987), 275–296. 207

[147] S.-T. Yau, Non-existence of continuous convex functions on certain Riemannian manifolds, *Math. Ann.* 207 (1974), 269–270. 205

Index