Lecture Notes
in Control and Information Sciences 327

Editors: M. Thoma · M. Morari

J.-D. Fournier · J. Grimm · J. Leblond
J. R. Partington (Eds.)

Harmonic Analysis and Rational Approximation

Their Rôles in Signals, Control and Dynamical Systems

With 47 Figures

Springer

Editors

Dr. J.-D. Fournier
Dr. J. Grimm
Dr. J. Leblond

Département ARTEMIS
CNRS and Observatoire de la Côte d'Azur
BP 4229
06304 Nice Cedex 4
France

Prof. J. R. Partington

University of Leeds
School of Mathematics
LS2 9JT Leeds
United Kingdom

ISSN 0170-8643

ISBN-10 3-540-30922-5 **Springer Berlin Heidelberg New York**
ISBN-13 978-3-540-30922-2 **Springer Berlin Heidelberg New York**

Library of Congress Control Number: 2005937084

Springer is a part of Springer Science+Business Media

springer.com

Typesetting: Data conversion by authors.
Final processing by PTP-Berlin Protago-TeX-Production GmbH, Germany
Cover-Design: design & production GmbH, Heidelberg
Printed on acid-free paper 54/3141/Yu - 5 4 3 2 1 0

In memoriam Macieja Pindora

Preface

This book is an outgrowth of a summer school that took place on the Island of Porquerolles in September 2003. The goal of the school was mainly to teach certain pieces of mathematics to practitioners coming from three different communities: signal, control and dynamical systems theory. Our impression was indeed that, in spite of their great potential applicability, 20th century developments in approximation theory and Fourier theory, while commonplace among mathematicians, are unknown or under-appreciated within the above-mentioned communities. Specifically, we had in mind:

- some advances in analytic, meromorphic and rational approximation theory, as well as their links with identification, robust control and stabilization of infinite-dimensional systems;
- the rich correspondences between the complex and real asymptotic behavior of a function and its Fourier transform, as already described, for instance, in Wiener's books.

In this respect, it is noticeable that in the last twenty years, much effort has been devoted to the research and teaching of recent decomposition tools, like wavelets or splines, linked to real analysis. From the early stages, we shared the view that, in contrast, research in certain fields suffers from the lack of a working knowledge of modern Fourier analysis and modern complex analysis.

Finally, we felt the need to introduce at the core of the school a probabilistic counterpart to some of the questions raised above. Although familiar to specialists of signal and dynamical systems theory, probability is often ignored by members of the control and approximation theory communities. Yet we hope to convey to the reader the conviction that there is room for fascinating phenomena and useful results to be discovered at the junction of probability and complex analysis.

This book is not just a proceedings of the summer school, since the contributions made by the speakers have been totally rewritten, anonymously refereed and edited in order to reflect some of the common themes in which the authors are interested, as well as the diversity of the applications. The contributors were asked to imagine addressing a fellow-scientist with a non-negligible but modest background in mathematics.

In drawing the boundaries between the chapters of the book, we have also tried to eliminate redundancy, while allowing for repetition of a theme as seen from different points of view.

We begin in Part I with a general introduction from the late Maciej Pindor. He surveys the conceptual and practical value of complex analyticity, both in the physical and the conjugate Fourier variables, for physical theories originally built in the real domain. Obstacles to analytic extension, like polar singularities known as “resonances”, a key concept of the school, turn out to have themselves a physical meaning. It is illustrated here by means of optical dispersion relations and the scattering of particles.

Part II of this book contains basic material on the complex analysis and harmonic analysis underlying the further developments presented in the book. Candelpergher writes on complex analysis, in particular analytic continuation and the use of Borel summability and Gevrey series. Partington gives an account of basic harmonic analysis, including Fourier, Laplace and Mellin transforms, and their links with complex analysis.

Part III contains further basic material, explaining some of the aspects of approximation theory. Pindor presents the theory of Padé approximation, including convergence issues. Levin and Saff explain how potential theoretic tools such as capacity play a role in the study of efficient polynomial and rational approximation, and analyse some weighted problems. Partington discusses the use of bases of rational functions, including orthogonal polynomials, Szegö bases, and wavelets.

Finally Part IV completes the foundations by a tour in probability theory. The driving force behind the order emerging from randomness, the central limit theorem, is explained by Collet, including convergence and fractal issues. Dujardin gives an account of the properties of random real polynomials, with particular reference to the distribution of their real and complex roots. Pindor puts rational approximation into a stochastic context, the basic idea being to obtain rational interpolants to noisy data.

The major application of the themes of this book lies in signal and control theory, which is treated in Part V. Deistler gives a thorough treatment of the spectral theory of stationary processes, leading to an account of ARMA and state space systems. Cuoco's paper treats the power spectral density of physical systems and its estimation, to be used in the extraction of signals out of noisy data. Olivi continues some of the ideas of Parts II and III, and, under the general umbrella of the Laplace transform in control theory, discusses linear time-invariant systems, controllability and rational approximation. Baratchart uses Laplace–Fourier transform techniques in giving an account of recent work analysing problems originating in the identification of linear systems subject to perturbations. In a final return to the perspective of the Introduction, Parts VI and VII shows the rôle of the previously-discussed tools in extremely diverse domains of physics. In Part VI, some mathematical aspects of dynamical systems theory are discussed. Biasco and Celletti are concerned with celestial mechanics and the use of perturbation theory to analyse integrable and nearly-integrable systems. Baladi gives a brief introduction to resonances in hyperbolic and hamiltonian systems, considered via the spectra of certain transfer operators. Part VII is devoted to a modern approach to two classical physics problems. Borgnat is concerned with turbulence in fluid flow; he discusses which tools, including the Mellin transform, can be adapted to reveal the various statistical properties of intermittent signals. Finally, Bondu and Vinet give an account of the high-performance control and noise analysis required at the gravitational waves VIRGO antenna.

Last but not least, our thanks go to the authors of the 17 contributions gathered in this book, as well as to all those who have helped us produce it, with particular mention of the anonymous referees.

Nice (France), Sophia-Antipolis (France), Leeds (U.K.), July 2005.

The editors: Jean-Daniel Fournier,
José Grimm,
Juliette Leblond,
Jonathan R. Partington.

Maciej PINDOR

Our colleague Dr. Maciej Pindor of Poland, the friend, collaborator and visitor of Jean-Daniel Fournier (JDF), died on Saturday 5th July 2003 at the Nice Observatory. Apparently, he was on his way to work from the "Pavillon Magnétique", where he was staying, to his office at "CION". His death was attributed to cardiac problems. He was 62 years old. Some colleagues were present, including the Director of the "Observatoire de la Côte d'Azur" (OCA) and JDF, when help arrived.

Maciej Pindor was a senior lecturer at the Institute of Theoretical Physics at the University of Warsaw. He performed his research work with the same care that he devoted to his teaching duties. He was a specialist in complex analysis, applied to some questions of theoretical physics, and, in recent years, to the processing of data; he produced theoretical and numerical solutions, which in this regard showed an ingenuity and reliability that is hard to match. He taught effective computational methods to young physicists. From the beginning of the thesis that Benédicte Dujardin has been writing under the direction of JDF, M. Pindor participated in her supervision.

The collaboration of JDF and his colleagues with M. Pindor began in 1996. Over the years, it was supported by regular or exceptional funding from the Cassini Laboratory, the Theoretical Physics Institute of Warzaw, the Polish Academy of Sciences and from OCA (with an associated post in astronomy). Thus M. Pindor came to Nice several times, and many people knew him. His genuine modesty made him a very accessible person, and dealings with him were agreeable and fruitful in all cases.

For the summer school of Porquerolles, he had agreed to give three courses, on three different subjects. In this he was motivated by friendship, scientific interest, and his acute awareness of the teaching responsibility borne by university staff; since then he had overcome the anxiety that he felt towards the idea of presenting mathematics in front of professional mathematicians. In particular, he was due to give the opening course, showing the link between physics and mathematics, treating the ideas of analyticity and resonance. He produced his notes for the course in good time, and these are therefore included under his name in this book and listed in the table of contents. At Porquerolles his courses were given by three different people. As co-worker JDF took the topic "rational approximation and noise". We sincerely thank the two others: G. Turchetti, himself an old friend of M. Pindor, agreed to expound the rôle of analytic continuation and Padé approximants in theoretical and mathematical physics; E. B. Saff kindly offered to lecture on the mathematics behind Padé approximants.

This book is dedicated to the memory of Maciej Pindor.

This obituary and M. Pindor's photograph have been included here by agreement with his widow, Dr. Krystyna Pindor-Rakoczy.

Memories of the Porquerolles School, a word from the co-directors

As already mentioned in the Preface, we organized the editing of the present book as a separate scientific undertaking, distinct from the school itself and with a wider team including J. Grimm and J.R. Partington. Nevertheless we feel bound to stress that the book is in part the result of the intellectual and congenial atmosphere created in Porquerolles in September 2003 by the speakers and the participants. Such moments are to be cherished, and have rewarded us for our own preparatory work. This seems a natural place to thank those of our colleagues who contributed to the running of the school, either as scientists or assistants, including those whose names do not appear here. Conversely we thank especially Elena Cuoco, who agreed to write a chapter for the book, although she had not been able to attend the school for personal reasons.

List of participants

D. Avanessoff (INRIA, Sophia-Antipolis [SA]),
V. Baladi (CNRS, Univ. Jussieu, Paris),
L. Baratchart (INRIA, SA),
L. Biasco (Univ. Rome III, It.),
B. Beckermann (Univ. Lille),
F. Bondu (CNRS, Observatoire de la Côte d'Azur [OCA], Nice),

P. Borgnat (CNRS, ENS Lyon),
V. Buchin (Russian Academy of Sciences, Moscow, Russia),
B. Candelpergher (Univ. Nice, Sophia-Antipolis [UNSA]),
A. Celletti (Univ. Rome Tor Vergata, It.),
A. Chevreuil (Univ. Marne la Vallée),
C. Cichowlas (ENS Ulm, Paris),
P. Collet (CNRS, Ecole Polytechnique, Palaiseau),
D. Coulot (OCA, Grasse),
F. Deleflie (OCA, Grasse),
M. Deistler (Univ. Tech. Vienne, Aut.),
B. Dujardin (OCA, Nice),
Y. Elskens (CNRS, Univ. Provence, Marseille),
J.-D. Fournier (CNRS, OCA, Nice),
V. Fournier (Nice),
Ch. Froeschlé (CNRS, OCA, Nice),
C. Froeschlé (CNRS, OCA, Nice),
A. Gombani (CNR, LADSEB, Padoue, It.),
J. Grimm (INRIA, SA),
E. Hamann (Univ. Tech. Vienne, Aut.),
J.-M. Innocent (Univ. Provence, Marseille),
J.-P. Kahane (Acad. Sciences Paris et Univ. Orsay),
E. Karatsuba (Russian Academy of Sciences, Moscow, Russia),
J. Leblond (INRIA, SA),
M. Mahjoub (LAMSIN-ENIT, Tunis),
D. Matignon (ENST, Paris),
G. Métris (OCA, Grasse),
N.-E. Najid (Univ. Hassan II, Casablanca, Ma.),
A. Neves (Univ. Paris V),
L. Niederman (Univ. Orsay),
N. Nikolski (Univ. Bordeaux),
A. Noullez (OCA, Nice),
M. Olivi (INRIA, SA),
J.R. Partington (Univ. Leeds, GB),
J.-B. Pomet (INRIA, SA),
E.B. Saff (Univ. Vanderbilt, Nashville, USA),
F. Seyfert (INRIA, SA),
N. Sibony (Univ. Orsay),
M. Smith (Univ. York, GB),
G. Turchetti (Univ. Bologne, It.),
G. Valsecchi (Univ. Rome, It.),
J.-Y. Vinet (OCA, Nice),
P. Vitse (Univ. Laval, Québec, Ca.).

Organization:
J. Gosselin (CNRS, Nice),
F. Limouzis (INRIA, SA),
D. Sergeant (INRIA, SA).

Finally we thank the sponsors of the school: CNRS (Formation Permanente), INRIA (Formation Permanente), INRIA Sophia-Antipolis, Conseil Régional PACA, Observatoire de la Côte d'Azur (OCA), Département Cassini, Ministère délégué Recherche et Nouvelles Technologies, VIRGO-EGO. We thank them all for their support.

Nice (France), Sophia-Antipolis (France), July 2005.

The co-directors: J.-D. Fournier,
J. Leblond

Contents

Part I Introduction

Analyticity and Physics
Maciej Pindor 3
1 Introduction 3
2 The optical dispersion relations 4
3 Scattering of particles and complex energy 7
References 12

Part II Complex Analysis, Fourier Transform and Asymptotic Behaviors

From Analytic Functions to Divergent Power Series
Bernard Candelpergher 15
1 Analyticity and differentiability 15
2 Analytic continuation and singularities 19
3 Continuation of a power series 22
4 Gevrey series 27
5 Borel summability 30
6 Acknowledgments 37
References 37

Fourier Transforms and Complex Analysis
Jonathan R. Partington 39
1 Real and complex Fourier analysis 39
2 DFT, FFT, windows 45
3 The behaviour of f and $\hat{f}$ 46
4 Wiener's theorems 49
5 Laplace and Mellin transforms 51
References 55

Part III Interpolation and Approximation

Padé Approximants
Maciej Pindor 59
1 Introduction 59
2 The Padé Table 60
3 Convergence 63
4 Examples 66
5 Calculation of Padé approximants 69
References 69

Potential Theoretic Tools in Polynomial and Rational Approximation
Eli Levin and Edward B. Saff 71
1 Classical Logarithmic Potential Theory 71
2 Polynomial Approximation of Analytic Functions 78
3 Approximation with Varying Weights — a background 82
4 Logarithmic Potentials with External Fields 85
5 Generalized Weierstrass Approximation Problem 88
6 Rational Approximation 89
References 93

Good Bases
Jonathan R. Partington 95
1 Introduction 95
2 Orthogonal polynomials and Szegö bases 96
3 Wavelets 99
References 102

Part IV The Rôle of Chance

Some Aspects of the Central Limit Theorem and Related Topics
Pierre Collet 105
1 Introduction 105
2 A short elementary probability theory refresher 107
3 Another proof of the CLT 109
4 Some extensions and related results 111
5 Statistical Applications 116
6 Large deviations 116
7 Multifractal measures 119
References 125

Distribution of the Roots of Certain Random Real Polynomials
Bénédicte Dujardin . . . 129
1 Introduction . . . 129
2 Real roots . . . 130
3 Complex roots . . . 132
4 Conclusion . . . 140
References . . . 142

Rational Approximation and Noise
Maciej Pindor . . . 145
1 Introduction . . . 145
2 Rational Interpolation . . . 145
3 Convergence . . . 148
4 Rational Interpolation with Noisy Data . . . 148
5 Froissart Polynomial . . . 154
6 Conclusions . . . 156
References . . . 156

Part V Signal and Control Theory

Stationary Processes and Linear Systems
Manfred Deistler . . . 159
1 Introduction . . . 159
2 A Short View on the History . . . 160
3 The Spectral Theory of Stationary Processes . . . 162
4 The Wold Decomposition and Forecasting . . . 172
5 Rational Spectra, ARMA and State Space Systems . . . 173
6 The Relation to System Identification . . . 177
References . . . 178

Parametric Spectral Estimation and Data Whitening
Elena Cuoco . . . 181
1 Introduction . . . 181
2 Parametric modeling for Power Spectral Density: ARMA and AR models . . . 182
3 AR and whitening process . . . 184
4 AR parameters estimation . . . 185
5 The whitening filter in the time domain . . . 187
6 An example of whitening . . . 189
References . . . 190

The Laplace Transform in Control Theory
Martine Olivi 193
1 Introduction 193
2 Linear time-invariant systems and their transfer functions 194
3 Function spaces and stability 197
4 Finite order LTI systems and their rational transfer functions 200
5 Identification and approximation 206
References 209

Identification and Function Theory
Laurent Baratchart 211
1 Introduction 211
2 Hardy spaces 211
3 Motivations from System Theory 214
4 Some approximation problems 218
References 227

Part VI Dynamical Systems Theory

Perturbative Series Expansions: Theoretical Aspects and Numerical Investigations
Luca Biasco and Alessandra Celletti 233
1 Introduction 233
2 Hamiltonian formalism 235
3 Integrable and nearly–integrable systems 237
4 Perturbation theory 241
5 A discrete model: the standard map 247
6 Numerical investigation of the break–down threshold 253
References 260

Resonances in Hyperbolic and Hamiltonian Systems
Viviane Baladi 263
1 Two elementary key examples – Basic concepts 263
2 Theorems of Ruelle, Keller, Pollicott, Dolgopyat... 268
References 269

Part VII Modern Experiments in Classical Physics: Information and Control

Signal Processing Methods Related to Models of Turbulence
Pierre Borgnat 277
1 An overview of the main properties of Turbulence 277
2 Signal Processing Methods for Experiments on Turbulence 287
References 298

Control of Interferometric Gravitational Wave Detectors
François Bondu and Jean-Yves Vinet 303
1 Introduction 303
2 Interferometers 303
3 Servo systems 305
4 Conclusion and perspectives 310
References 310

List of Contributors

Viviane Baladi
CNRS UMR 7586,
Institut Mathématique de Jussieu,
75251 Paris (France)
baladi@math.jussieu.fr

Laurent Baratchart
Inria, Apics Team
2004, Route des Lucioles
06902 Sophia Antipolis (France)
baratcha@sophia.inria.fr

Luca Biasco
Dipartimento di Matematica,
Università di Roma Tre,
Largo S. L. Murialdo 1,
I-00146 Roma (Italy)
biasco@mat.uniroma3.it

François Bondu
Laboratoire Artemis
CNRS UMR 6162
Observatoire de la Côte d'Azur
BP4229 Nice (France)
Francois.Bondu@obs-nice.fr

Pierre Borgnat
Laboratoire de Physique
UMR-CNRS 5672
ÉNS Lyon 46 allée d'Italie
69364 Lyon Cedex 07 (France)
Pierre.Borgnat@ens-lyon.fr

Bernard Candelpergher
University of Nice-Sophia Antipolis
Parc Valrose
06002 Nice (France)
candel@math.unice.fr

Alessandra Celletti
Dipartimento di Matematica,
Università di Roma Tor Vergata,
Via della Ricerca Scientifica 1,
I-00133 Roma (Italy)
celletti@mat.uniroma2.it

Pierre Collet
Centre de Physique Théorique
CNRS UMR 7644
Ecole Polytechnique
F-91128 Palaiseau Cedex (France)
collet@cpht.polytechnique.fr

Elena Cuoco
INFN, Sezione di Firenze,
Via G. Sansone 1,
50019 Sesto Fiorentino (FI),
present address:
EGO, via Amaldi,
Santo Stefano a Macerata,
Cascina (PI) (Italy)
elena.cuoco@ego-gw.it

Manfred Deistler
Department of Mathematical Methods in Economics,
Econometrics and System Theory,
Vienna University of Technology
Argentinierstr. 8,
A-1040 Wien (Austria)
Deistler@tuwien.ac.at

Bénédicte Dujardin
Département Artémis,
Observatoire de la Côte d'Azur,
BP 4229, 06304 Nice (France)
dujardin@obs-nice.fr

Eli Levin
The Open University of Israel
Department of Mathematics
P.O. Box 808, Raanana (Israel)
elile@openu.ac.il

Martine Olivi
Inria, Apics Team
2004, Route des Lucioles
06902 Sophia Antipolis (France)
Martine.Olivi@sophia.inria.fr

Jonathan R. Partington
School of Mathematics
University of Leeds
Leeds LS2 9JT (U.K.)
J.R.Partington@leeds.ac.uk

Maciej Pindor
Instytut Fizyki Teoretycznej,
Uniwersytet Warszawski ul.Hoża 69,
00-681 Warszawa (Poland)
deceased

Edward B. Saff
Center for Constructive Approximation
Department of Mathematics
Vanderbilt University
Nashville, TN 37240 (USA)
esaff@math.vanderbilt.edu

Jean-Yves Vinet
ILGA, Département Fresnel
Observatoire de la Côte d'Azur
BP 4229, 06304 Nice (France)
vinet@obs-nice.fr

Analyticity and Physics

Maciej Pindor

Instytut Fizyki Teoretycznej,
Uniwersytet Warszawski ul.Hoża 69,
00-681 Warszawa, Poland.

1 Introduction

My goal is to present to you some aspects of the role that the mathematical concept as subtle and abstract as “analyticity” plays in physics.

In retrospective we could say that also the “real number” notion is in fact a very abstract one and its applicability to the description of the world external to our mind, is sort of a miracle – I do not want to dwell here on a relation between constructs of the mind and the “external world” – this is the playground for philosophers and I do not wish to compete with them. I mean here the intuitively manifest difference between the obvious nature of integer numbers (and “nearly obvious nature” of rationals) and abstractness of real numbers. This abstractness notwithstanding, I do not think that talking in terms of real numbers when describing the “real world” needed much more intellectual effort than applying rational numbers there. This fact is excellently demonstrated by the fact that in “practice” we use only rationals: e.g. floating point numbers in computer calculations – “practitioners” just ignore the subtle flavour of irrationals and treat them as rationals represented in decimal system by a “sufficient” number of digits.

The situation is completely different with complex numbers. Contrary to many other mathematical notions, they originated entirely within pure mathematics and even for mathematicians they seemed so strange that the word “imaginaire” was attributed to them! No “real world” situation seemed to demand complex numbers for its mathematical description. However already Euler (and also d’Alembert) observed that they were useful in solving problems in hydrodynamics and cartography [4]. Once domesticated by mathematicians, complex numbers slowly creeped into physical papers, though only as an auxiliary and convenient tool when dealing with periodic solutions of some mechanical systems (the spherical pendulum studied by Tissot [7]). Their particular usefulness was discovered by Riemann for describing some form of the potential field [6] and when he studied Maxwell equations [9], but again

J.-D. Fournier et al. (Eds.): Harm. Analysis and Ratio. Approx., LNCIS 327, pp. 3–12, 2006.

they played here a role of a shorthand notation for a simultaneous description of two different, though related, physical quantities. Even the advent of the quantum mechanics did not change too much – although the "wave function" was essentially complex and its real and imaginary part had no separate existence, the values of the function had no physical meaning themselves. It was its modulus that was interpretable – and so physicists could think of its "complexness" as of some mathematical trick – however with some feeling of uneasiness, this time.

As far as I know, the first individuals that truly opened the complex plane for physics were Kramers and Kronig (see [3]). They had the daredevil idea that extending the domain of a function, having a well defined physical quantity as its argument – the frequency in their case – to the complex plane, can lead to conclusions verifiable experimentally. They have shown, moreover, that properties of this function in the complex plane are connected to important physical conditions on another function. Their idea seemed a curiosity and only 25 years later it was found useful and advantages of considering energy on the complex plane were discovered. Even then physicists felt still uneasy with this, and when few years later Tulio Regge proposed extending the angular momentum to the complex plane his paper was rejected by many referees [1].

In the following I shall briefly review the original idea of Kramers and Kronig (following closely the exposition of [3]), the consequences of the extension of the energy to the complex plane in the description of particle scattering and the Regge idea.

2 The optical dispersion relations

Kramers and Kronig considered light in a material medium. The physical situation there is described by two fields: the electric field $\boldsymbol{E}(\boldsymbol{x},t)$ and the displacement field $\boldsymbol{D}(\boldsymbol{x},t)$. Let me clarify that the latter field comes from a superposition of the former one and fields produced by atoms and particles of the medium polarized by the presence of $\boldsymbol{E}$.

Their monochromatic components of frequency ω are related through

$$\tilde{\boldsymbol{D}}(\boldsymbol{x},\omega)=\varepsilon(\omega)\tilde{\boldsymbol{E}}(\boldsymbol{x},\omega) \tag{1}$$

where $\varepsilon(\omega)$ is called the dielectric constant and is frequency dependent, because the response of the medium to the presence of $\boldsymbol{E}$ depends on frequency. These frequency components are just Fourier transforms of the temporal dependence of the fields, e.g.

$$\boldsymbol{E}(\boldsymbol{x},t)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\tilde{\boldsymbol{E}}(\boldsymbol{x},\omega)\mathrm{e}^{\mathrm{i}\omega t}\mathrm{d}\omega$$

and vice versa

$$\tilde{\boldsymbol{E}}(\boldsymbol{x},\omega)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\boldsymbol{E}(\boldsymbol{x},\tau)\mathrm{e}^{-\mathrm{i}\tau t}\mathrm{d}\tau\ .$$

Using now (1) and assuming that the functions considered vanish at infinity in time and frequency fast enough as to make exchange of order of integration possible, we arrive at

$$\boldsymbol{D}(\boldsymbol{x},t)=\boldsymbol{E}(\boldsymbol{x},t)+\int_{-\infty}^{+\infty}G(\tau)\boldsymbol{E}(\boldsymbol{x},t-\tau)\mathrm{d}\tau \tag{2}$$

where $G(\tau)$ is the Fourier transform of $\varepsilon(\omega)-1$:

$$G(\tau)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}[\varepsilon(\omega)-1]\mathrm{e}^{-\mathrm{i}\omega\tau}\mathrm{d}\omega\ . \tag{3}$$

These mathematical manipulations may seem not very inspiring, but if we look carefully at (2) we can observe that it is somewhat strange – it says that the value of $\boldsymbol{D}$ at the moment t depends on the values of $\boldsymbol{E}$ at all instants of time – we say that the connection between $\boldsymbol{D}$ and $\boldsymbol{E}$ is *nonlocal* in time. Well, we can understand that polarizing of atoms and molecules takes some time and therefore the effect of changing $\boldsymbol{E}$ will be felt by the values of $\boldsymbol{D}$ after some time, but how can $\boldsymbol{D}$ at time t depend on values of $\boldsymbol{E}$ in later times – what is represented in (2) by the part of the integral from $-\infty$ to 0? Every physicist would say: IT CANNOT DEPEND! It would violate "causality".

This means that we must have $G(\tau)\equiv 0$ for $\tau<0$. Consequently, this means that there are some necessary conditions on the dependence of ε on ω. If we invert the Fourier transform in (3) we get now

$$\varepsilon(\omega)=1+\int_{0}^{\infty}G(\tau)\mathrm{e}^{\mathrm{i}\omega\tau}\mathrm{d}\tau\ . \tag{4}$$

Already at the very birth of the theoretical optics physicists used some simple "models", classical ones because quantum mechanics was not yet born, to describe the interaction between light and matter and these models lead to expressions for $\varepsilon(\omega)$ satisfying our requirement that $G(\tau)\equiv 0$ for $\tau<0$. However, truly speaking, the phenomenon of polarization of atoms and molecules is a very complicated one and even now it is not easy to describe it in all its details and it is not obvious how should one guarantee vanishing of the predicted $G(\tau)$ for negative arguments.

Kramers and Kronig observed that the most general conditions one should impose on $\varepsilon(\omega)$ to have "causality" satisfied, is just that it be of the form (4) with some real $G(\tau)$. Again, this form would not be so very interesting if not their daring concept of considering $\varepsilon(\omega)$ as a function of *complex* ω. Once they did this, many interesting conclusions followed. The most fundamental observation is that if $G(\tau)$ is finite for all τ, $\varepsilon(\omega)$ *is an analytic function of* ω *in the upper half plane.*

Although you will soon listen to a lecture on fundamentals of functions of a complex variable I am afraid I have to state here very briefly what *analyticity* is and what are its consequences for $\varepsilon(\omega)$. It sounds deceptively simple: $f(z)$ is analytic at $z = z_0$ if it has a derivative at this point. However the "point" is now a point of the plane, therefore the requirement leads to so called Cauchy-Riemann equations, which are, actually, differential equations relating the real and the imaginary parts of $f(z)$ as functions of the real and imaginary parts of z (in fact these equations were written already by d'Alembert and Euler!). The amazing consequence is that if a function possesses a first derivative at some point it possesses all derivatives there and also it has a Taylor expansion with non-zero radius of convergence at this point! Moreover if $f(z)$ is analytic inside some domain D and C is a "smooth" closed curve encircling its interior counterclockwise (*simple closed rectifiable positively oriented curve*) with ω inside the curve, then there holds the Cauchy theorem

$$f(\omega) = \frac{1}{2\pi \mathrm{i}} \int_C \frac{f(t)}{t-\omega} \mathrm{d}t. \tag{5}$$

We can now take D as the upper half plane, ω infinitesimally above the real axis and C as on the Figure 1 and write (5) for $f(\omega) = \varepsilon(\omega) - 1$. With the condition that $\varepsilon(\omega) - 1$ vanishes for large ω at least like $1/\omega^2$, which can be justified by some physical arguments, we can take $R \to \infty$ and neglect the integral over C_R. With some more maneuvering we arrive at the famous *dispersion relations* for the real and the imaginary parts of $\varepsilon(\omega)$.

$$\begin{aligned} \operatorname{Re} \varepsilon(\omega) &= 1 + \frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{\operatorname{Im} \varepsilon(t)}{t-\omega} \mathrm{d}t \\ \operatorname{Im} \varepsilon(\omega) &= -\frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{[\operatorname{Re} \varepsilon(t) - 1]}{t-\omega} \mathrm{d}t. \end{aligned} \tag{6}$$

The name comes from the fact that the dependence of ε on ω leads to the phenomenon called dispersion – the change of the shape of the light wave penetrating a material medium. The real part of ε is directly related to this phenomenon, while the imaginary part is connected with the absorption of light. Therefore they can be both measured and, not unexpectedly, experimental data confirm the validity of (6).

On the other hand the Titchmarsh theorem [8] says that if a function $F(z)$ satisfies relations of the type (6) then its Fourier transform vanishes on the real negative semiaxis. Thus, not only the physical condition of "causality" leads to definite analytical properties of some function implying a relation between its real and imaginary parts on the real axis that can be confirmed by physical experiments, but also the experimentally verifiable relation between two physical quantities, when they are considered the real and imaginary parts of an analytical function on the real axis, implies a property of the Fourier transform of this function, the one having the meaning of "causality"!

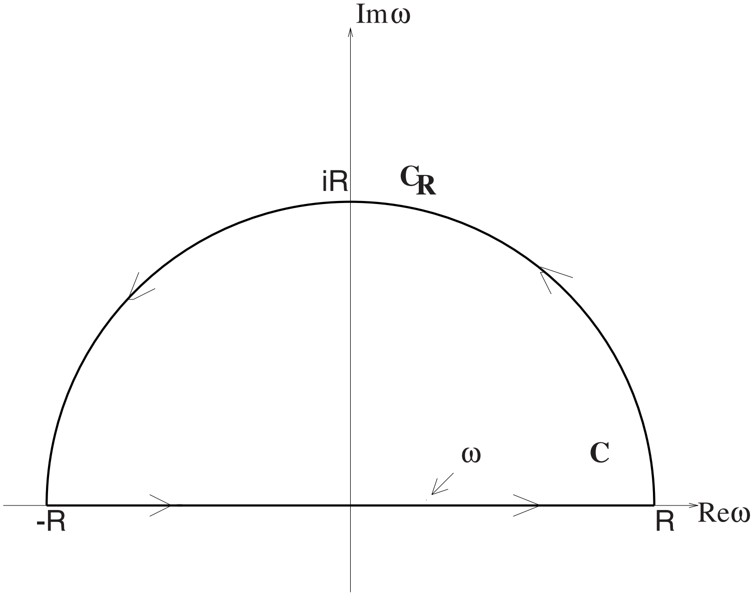


Fig. 1. Contour for the integral (5) in the complex plane of ω

3 Scattering of particles and complex energy

In the middle of fifties of the last century the physics of subatomic constituents of matter, called "elementary particles", amassed a vast amount of experimental observations which were impossible to explain on the grounds of the fundamental theory of the "microworld" – the Quantum Field Theory. Not that they were in contradiction with the QFT – simply the equations of the QFT could have been solved only in some approximation scheme, called the perturbation theory, that seemed to fail completely except in the case of electromagnetism, where it (called Quantum Electrodynamics) worked perfectly.

However the QFT had still another important deficiency – actually it had no firm mathematical foundations. In fact it was a cookbook of recipes how to deal with objects of a very obscure mathematical meaning to extract formulae containing quantities related to laboratory observations. Therefore, although the QFT came to existence as the logical extrapolation of ideas of the Quantum Mechanics – so fantastically fruitful in explaining the atomic world – to the realm of the relativistic phenomena where mass and energy are one and the same physical quantity and where physical particles are freely created and annihilated, it was slowly looked at with a growing suspicion. Its inability to deal with the growing mass of observational data concerning “elementary particles” seemed to seal its fate.

In this desperate situation it was recalled that 25 years earlier Kramers and Kronig were able to derive their “dispersion relations” using the apparatus of the functions of a complex variable with only the fundamental physical property as “causality”, as input.

The simplest process studied in elementary particle physics is the elastic scattering of two spinless particles. The word “elastic” means that the same two particles that enter the scattering process, emerge from it and no other particle is created in the process.

The states of the particles are defined by their *four-momenta* – space-time vectors with three components being the ordinary momentum, and the fourth (or rather zeroth, in the notation I shall use) component being the energy of the particle. The four-momenta of the particles before the scattering are p_1 and p_2 and after the scattering they are p_3 and p_4. Squares of this four-momenta are just masses of the particles squared – let me remind you that the space-time has the special metric

$$p_i^2 = p_{i,0}^2 - p_{i,1}^2 - p_{i,2}^2 - p_{i,3}^2 = E_i^2 - \boldsymbol{p}_i^2 = m_i^2 \quad i = 1, ..., 4 \ .$$

The total four-momentum of the system

$$P = p_1 + p_2 = p_3 + p_4; \qquad p_i = (E_i, \boldsymbol{p}_i) \quad i = 1, ..., 4$$

is conserved and so is the total energy. In the special reference system, called the center of mass system (c.m.s.), the total momentum is zero, and therefore the c.m.s. energy squared is equal to $s = P^2$. Another four-vector important in the description of the process is the *momentum transfer* q, together with its square t

$$q = p_1 - p_3 = -p_2 + p_4; \quad t = q^2 \ .$$

In the scattering of two particles of identical masses m the momentum transfer is simply related to the scattering angle θ and the energy via

$$t = -\frac{1}{2}(1 - \cos\theta)(s - 4m^2)$$

and is negative, while the (relativistic) energy is larger than $4m^2$.

The quantity relevant in this context is the *scattering amplitude* $A(s,t)$. The squared modulus of the scattering amplitude is, apart of some “kinematical” factors, the “cross-section” for the scattering – loosely speaking a probability of the registration of the scattered particle along a direction defined by the given momentum transfer when the scattering takes place at the given energy.

Using the very general formulae for this scattering amplitude following from QFT and applying as precise mathematical apparatus as was possible in this context at that time, it appeared possible to show again that relativistic causality (i.e. impossibility of any relation between events separated in such a way that they could not be connected by signals traveling with a speed inferior or equal to the velocity of light) implies some special analyticity properties of the scattering amplitude in the complex plane of energy (see e.g. [2] and references therein).

In fact, the fascinating connection between the physical requirement – causality – and the abstract mathematical property – analyticity – has been rigorously (almost) shown only for the "forward" scattering amplitude, i.e. at $t = 0$. These analytical properties allowed then one, using the theorems from complex variable functions theory, to write the *dispersion relations* for the scattering amplitude of the type

$$A(s,0) = \frac{1}{\pi}\int_{4m^2}^{\infty} \frac{\operatorname{Im} A(s',0)\mathrm{d}s'}{s'-s-\mathrm{i}\varepsilon} + \frac{1}{\pi}\int_{-\infty}^{0} \frac{\operatorname{Im} A(s',0)\mathrm{d}s'}{s'-s-\mathrm{i}\varepsilon}. \tag{7}$$

Here $-\mathrm{i}\varepsilon$ means that the integration runs just above the real axis. This *integral representation* of $A(s,0)$ as a function of complex s means that this function has the very nasty *singularities* (i.e. the points where it is not analytic) at $s = 4m^2$ and $s = 0$ (and possibly $s = \infty$) called the *branchpoints*. They are nasty, because they make the function *multivalued* – if we walk along a closed curve encircling such a point, then at the point from which we started we find a different value of the function. I cannot dwell on this horror (or, to me, the fascinating property of the complex plane) here but can only say that the multivalued function can be made univalued by removing, from the complex plane, lines joining the branchpoints – such lines are called the *cuts*. Looking at (7) you see that $A(s,0)$ is not defined on $(-\infty,0)$ and $(4m^2,\infty)$ – these are the cuts. On the other hand the function has well defined limits when s approaches these semiaxes from imaginary directions. The limit from above for $s \in (4m^2,\infty)$ is just the *physical* scattering amplitude – because these values of s correspond to physical scattering process. On the other hand the limit from below for $s \in (-\infty,0)$ corresponds to the scattering amplitude for another process related to the one we consider, through the "crossing symmetry" – a property of the scattering amplitude suggested by the QFT. Combining this property with "unitarity" – loosely speaking the requirement that the probability that anything can happen (in the context of the scattering, of course) is one, leads to conclusions that again could have been verified experimentally. This was a great triumph, because earlier no quantitative predictions concerning phenomena connected with new types of interactions (new with respect to electromagnetism) could have been given.

The great success of the simplest dispersion relations prompted many theoreticians to study the analytical structure of the scattering amplitude as suggested by the perturbation theory – though the later produced divergent expansions. This analytical structure appeared to be very rich with many branchpoints on the real axis (where the amplitude had a "physical meaning") with locations depending on masses of the scattered particles, and poles at energies of the bound states (if any) of these particles. Moreover, as mentioned above, the "crossing symmetry" implied direct connections between values of the scattering amplitude on some edges of different cuts. Causality implied that the scattering amplitude is analytic on the whole plane of complex energy properly cut along the real axis, but it was soon realized that there have to exist poles on the "unphysical sheets" – one of the fantastic properties of the analytic functions is that they can undergo the *analytic continuation*. You will learn more about it during the lectures to come, but here I shall describe it as a feature which makes the function defined on its whole domain, once it is defined on the smallest piece of it. The "domain" can mean also other "copies" (called *Riemannian sheets*) of the complex plane – if there are branchpoints – reached when one continues function analytically across the cuts. In elementary particle physics, the sheet on which energy has the "physical meaning", is called the "physical sheet". The ones reached through analytic continuation of the amplitude across the cuts, are called the "unphysical sheets". I want to make clear this fundamental fact: the assumption of analyticity of the scattering amplitude as a function of complex energy means that its values on sections of the real axis, where the values of energy correspond to the physical scattering process, define the scattering amplitude on all its Riemannian sheets. In particular for many types of scattering processes the amplitude had to have poles on the first "unphysical sheet". These poles were the manifestations of "resonances" – experimentally seen enhancements of the cross-section, related in solvable "models" of scattering (e.g. nonrelativistic scattering described by the Schrödinger equation) to short-living quasibound states of the scattered particles and therefore also in relativistic description attributed to an existence of short living non-stable particles.

Also using suggestions from the expansions of the scattering amplitude obtained in the perturbation theory, the so called *double dispersion relations* – written both in the complex s and t planes – were postulated and some verifiable – and verified! – conclusions followed from them.

Another astonishing concept was put forward by T. Regge [5]. He considered the, so called, partial waves expansion of the nonrelativistic scattering amplitude $A(q^2, t)$

$$A(q^2, t) = f(q^2, \cos(\theta)) = \sum_{l=0}^{\infty} (2l+1) A_l(q^2) P_l(\cos(\theta))$$

where $P_l(z)$ are the Legendre polynomials. $A_l(q^2)$ are called the *partial wave* amplitudes and describe the scattering at the given angular (orbital) momentum. The sum runs over integers only, because in quantum physics the angular momentum is "quantized", i.e. it can take on values only from the discreet countable set. Regge had, however, an idea to consider the angular momentum in the complex plane!

He studied the nonrelativistic scattering for a "reasonable" class of potentials (a superposition of *Yukawa potentials*) and was able to show that

$A_l(q^2)$ is meromorphic in l in the half plane $\mathrm{Re}\, l > -1/2$ where it vanishes exponentially as $|l| \to \infty$. Using this and writing the above expansion as the integral

$$f(q^2, \cos(\theta)) = \frac{\mathrm{i}}{2} \int_C \mathrm{d}l (2l+1) A(l, q^2) \frac{P_l(-\cos(\theta))}{\sin(\pi l)}$$

where the contour C encircled the positive semiaxis clockwise (so it was, in fact, the sum of small circles around all positive integers), he could deform the contour C by moving its ends at $\infty \pm \mathrm{i}\varepsilon$ to $-\frac{1}{2} \pm \mathrm{i}\infty$. As the result he got

$$\begin{aligned} f(q^2, \cos(\theta)) = \frac{\mathrm{i}}{2} \int_{-\frac{1}{2}-\mathrm{i}\infty}^{-\frac{1}{2}+\mathrm{i}\infty} \mathrm{d}l (2l+1) A(l, q^2) \frac{P_l(-\cos(\theta))}{\sin(\pi l)} \\ - \pi \sum_{n=1}^{N} \frac{(2\alpha_n(q^2)+1)\beta_n(q^2)}{\sin(\pi \alpha_n(q^2))} P_{\alpha_n(q^2)}(-\cos(\theta)) \end{aligned}$$

where the sum runs over all poles (called since then the Regge poles) of $A(l, q^2)$ in the half plane of the complex l, $\mathrm{Re}\, l > -\frac{1}{2}$.

The most exciting part came from the fact that for $q^2 < 0$, we call it *below threshold*, all these poles lie on the real axis and correspond precisely to bound states of the potential at energies $(-q^2)$ at which $\alpha_n(q^2)$ equals to an integer being the angular momentum of the given bound state! When q^2 grows above the threshold (becomes positive) $\alpha_n(q^2)$ move to the complex plane and when at some q_r the real part of it crosses an integer, the scattering amplitude has a form

$$\frac{a}{(q^2 - q_r^2)b + \mathrm{i}\, \mathrm{Im}\, \alpha_n(q_r^2)}$$

characteristic of a resonance. This way bound states and resonances were grouped into *Regge trajectories* originating from the same $\alpha_n(q^2)$.

It was then immediately conjectured that the relativistic scattering amplitude shows the same (or analogous) behaviour in the complex angular momentum plane. Though many actual resonances were grouped into Regge trajectories, other conclusions were not verified experimentally, what was attributed to a hypothetical existence of branchpoints of the scattering amplitude in the complex angular momentum plane. When such branchpoints were included the theory lost its beautiful simplicity and its predictive power was considerably limited. Because of that, its attractivity paled and though it is still considered that actually bound states and resonances form families lying on Regge trajectories, no more much importance is attributed to this fact.

This amazing fact that elements of the analytical structure of the scattering amplitude, as a function of the complex energy and momentum transfer, have direct physical meaning, induced some physicist to think that just the proper analytical properties of the scattering amplitude compatible with the fundamental physical conditions (like the “crossing symmetry” or the “unitarity”) could form the correct set of assumptions to build a complete theory of the phenomena concerning elementary particles. This point of view fell later out of fashion in the view of the spectacular success of the developments of the QFT which take now the shape of the Nonabelian Gauge Field Theory. Nevertheless the lesson that functions describing the physical observations in terms of the physically measurable parameters must be studied for complex values of these parameters because the analytic properties of such functions have direct relation to true physical phenomena underlying the observations, is now deeply rooted in the thinking of physicists.

References

1. G. Białkowski. Private information.
2. S. Gasiorowicz. *Elementary Particle Physics.* John Wiley and Sons, 1966.
3. J.D. Jackson. *Classical Electrodynamics.* John Wiley and Son, 1975.
4. A. Markushevich. Basic notions of mathematical analysis in Euler papers. *“Leonard Euler” Acad. Nauk SSSR*, 1959. (in Russian)
5. T. Regge. *Nuovo Cimento 14, p. 951.* 1959.
6. B. Riemann. *Bernhard Riemann’s gesammelte mathematische Werke.* Dover Publications, 1953. p. 431.
7. Tissot. Journal de Liouville 1857; according to P. Appel, *Traité de méchanique rationnelle* vol. 1, Paris. 1932.
8. E.C. Titchmarsh. *Introduction to the Theory of Fourier Integrals.* Oxford Univ. Press, 1948.
9. H. Weber. *Die partiellen Diffential-Gleichungen der mathematischen Physik nach Riemann’s Vorlesungen.* Friedrich Vieweg u. Sohn, 1901.

From Analytic Functions to Divergent Power Series

Bernard Candelpergher

University of Nice-Sophia Antipolis
Parc Valrose
06002 Nice (France)
candel@math.unice.fr

1 Analyticity and differentiability

1.1 Differentiability

The functions occurring commonly in classical analysis, such as x^n, e^x, $\mathrm{Log}(x)$, $\sin(x)$, $\cos(x)$, $\ldots$, are not only defined on intervals in $\mathbb{R}$, but they can also be defined when the variable x (which we shall now denote by z) lies in some subdomain of $\mathbb{C}$. These domains are the subsets of U of $\mathbb{C}$ that we call *open sets*, and are characterised by the property

$$z_0 \in U \Rightarrow \text{there exists } r > 0 \text{ such that } D(z_0, r) \subset U$$

where $D(z_0, r) = \{z \in \mathbb{C}\,,\ |z - z_0| < r\}$ is the disc with centre z_0 and radius r.

Let U be an open subset of $\mathbb{C}$ and let $f : U \to \mathbb{C}$ be a function. We say that f is *differentiable* on U if for $z_0 \in U$, the expression

$$\frac{f(z) - f(z_0)}{z - z_0}$$

tends to a finite limit when z tends to z_0 in U. We denote this limit by $f'(z_0)$ or $\partial f(z_0)$. We say also that f is *holomorphic* on U (this terminology comes from the fact that $f(z) \simeq a + b(z - z_0)$ for z in a neighbourhood of z_0, and so f is locally a similarity).

Formally, the definition of differentiability in $\mathbb{C}$ is the same as in $\mathbb{R}$, and its immediate consequences, such as the differentiability of a sum, a product and a composition of functions, will therefore continue to hold. However, the notion of differentiability in $\mathbb{C}$ is more restrictive than in $\mathbb{R}$ since the expression $\dfrac{f(z) - f(z_0)}{z - z_0}$ has to tend to the same limit no matter how z tends to z_0 in the complex plane. In particular if we write $z = x + \mathrm{i}y$, the function f, considered as a function of two real variables, x and y, will have partial derivatives

J.-D. Fournier et al. (Eds.): Harm. Analysis and Ratio. Approx., LNCIS 327, pp. 15–37, 2006.

with respect to x and y, satisfying certain equations known as the "Cauchy-Riemann equations".

Indeed, let us consider the functions

$$\begin{aligned} \Phi &: (x,y) \to \operatorname{Re} f(x+\mathrm{i}y) \\ \Psi &: (x,y) \to \operatorname{Im} f(x+\mathrm{i}y). \end{aligned}$$

It is easy to check that the differentiability of f with respect to z implies that the functions Φ and Ψ are differentiable with respect to x and y, and that

$$\begin{aligned} f'(x+\mathrm{i}y) &= \partial_x \Phi(x,y) + \mathrm{i}\partial_x \Psi(x,y) \\ &= \frac{1}{\mathrm{i}}(\partial_y \Phi(x,y) + \mathrm{i}\partial_y \Psi(x,y)) \end{aligned}$$

and hence the partial derivatives satisfy the *Cauchy-Riemann equations:*

$$\begin{aligned} \partial_x \Phi &= \partial_y \Psi, \\ \partial_y \Phi &= -\partial_x \Psi. \end{aligned}$$

The properties of holomorphic functions on an open subset U of $\mathbb{C}$ are therefore much more striking than those of functions of a real variable. In particular a function that is holomorphic on $U \setminus \{a\}$ and with a finite limit at a is holomorphic on U (this is the Riemann theorem).

1.2 Integrals

Let f be an holomorphic function on an open subset U of $\mathbb{C}$ and γ a path in U (so γ is a piecewise continuously differentiable function on an interval $[a,b]$ with values in U; if $\gamma(a) = \gamma(b)$, we say that γ is a closed path). We write

$$\int_\gamma f(z)\mathrm{d}z = \int_a^b f(\gamma(t))\gamma'(t)\mathrm{d}t.$$

A natural question is to see how this integral depends on the path γ, and in particular, what happens if we deform the path γ continuously, while remaining in U. It is the concept of homotopy that allows us to make this precise, saying that two paths γ_0 and γ_1 with the same endpoints (or two closed paths), are *homotopic* in U if there exists a family γ_s of intermediate paths (resp. of closed paths) between γ_0 and γ_1, having the same endpoints as γ_0 and γ_1, which depend continuously on the parameter $s \in [0,1]$.

The homotopy theorem

If f is holomorphic in U, and if γ_1 and γ_2 are two paths with the same endpoints, or else two closed paths, which are homotopic in U, then

$$\int_{\gamma_1} f(z)\mathrm{d}z = \int_{\gamma_2} f(z)\mathrm{d}z.$$

Since the integral along a closed path consisting of a single point z_0 (i.e., the closed path $t \to z_0$ for all t) is zero, it follows from the homotopy theorem that if f is holomorphic in U and if we can continuously contract a closed path γ down to a point z_0 in U while remaining all the time in U, then we have

$$\int_{\gamma} f(z)\mathrm{d}z = 0.$$

Connected open sets U (i.e., ones consisting of a single piece) for which every closed path in U is homotopic in U to a single point in U are called *simply connected.*

We deduce from the above that if f is holomorphic on a simply connected open set U and z_0 is a point of U, then for every closed path γ in U we have

$$\int_{\gamma} \frac{f(z) - f(z_0)}{z - z_0}\mathrm{d}z = 0.$$

Since we have

$$\int_{C(z_0,r)} \frac{1}{z - z_0}\mathrm{d}z = 2\mathrm{i}\pi,$$

with $C(z_0, r)(t) = z_0 + r\exp(\mathrm{i}t), t \in [0, 2\pi]$, the circle of center 0 and radius r, then if f is holomorphic on a simply connected open set U and $C(z_0, r) \subset U$, we have *Cauchy's formula*

$$f(z_0) = \frac{1}{2\mathrm{i}\pi} \int_{C(z_0,r)} \frac{f(z)}{z - z_0}\mathrm{d}z.$$

1.3 Power series expansions

Cauchy's formula enables us to show that a function f that is holomorphic on an open subset U of $\mathbb{C}$ is in fact infinitely differentiable, we have

$$\partial^n f(z_0) = \frac{n!}{2\mathrm{i}\pi} \int_{C(z_0,r)} \frac{f(z)}{(z - z_0)^{n+1}}\mathrm{d}z.$$

Writing the Cauchy formula at z

$$f(z) = \frac{1}{2\mathrm{i}\pi} \int_{C(z_0,r)} \frac{f(u)}{(u - z_0) - (z - z_0)}\mathrm{d}u = \frac{1}{2\mathrm{i}\pi} \int_{C(z_0,r)} \frac{f(u)}{u - z_0} \frac{1}{1 - \frac{(z - z_0)}{(u - z_0)}}\mathrm{d}u$$

and expanding

$$\frac{1}{1-\frac{(z-z_0)}{(u-z_0)}} = \sum_{n\geq 0} \frac{(z-z_0)^n}{(u-z_0)^n}$$

we see that f can be expanded in a Taylor series about every point of U. Precisely for each $z_0 \in U$ and for all $R > 0$ such that $D(z_0, R) \subset U$, we have

$$f(z) = \sum_{n=0}^{+\infty} \frac{\partial^n f(z_0)}{n!}(z-z_0)^n$$

for every $z \in D(z_0, R)$. We say that f is *analytic* on U.

We see therefore that if f is holomorphic on an open subset U of $\mathbb{C}$, then the radius of convergence of the Taylor series of f about z_0 is greater than or equal to every $R > 0$ for which $D(z_0, R) \subset U$. In other words, the disc of convergence of the Taylor series of f about z_0 is only controlled by the regions where f fails to be holomorphic.

1.4 Some properties of analytic functions

The principle of isolated zeroes

This principle may be expressed as the fact that the points where an analytic function f on U takes the value zero, i.e., the *zeroes* of f, cannot accumulate at a point in U (unless f is identically zero). In other words, no compact subset of U can contain more than finitely many zeroes of f.

Uniqueness of analytic functions

Cauchy's formula shows that a function analytic in the neighbourhood of a disc is fully determined on the interior of the disc if one knows its values on the circle bounding the disc. We see a further uniqueness property in the fact that if f is an analytic function on a connected open set U, then the values of f on a complex line segment $[z_0, z_1]$ of U, joining two different points z_0, z_1 of U, determine f uniquely on the whole of U.

To put it another way, if two analytic functions f and g on a connected open set U are equal on a segment $[z_0, z_1]$ of U, then they are equal on the whole of U.

The maximum principle

If f is a non-constant analytic function on U, then the function $|f|$ cannot have a local maximum in U, in particular if U is bounded, the maximum of $|f|$ is attained on the boundary of U.

Sequences, series and integrals of analytic functions

If (f_n) is a sequence of analytic functions in U, converging uniformly on every disc in U, then the limit function f is also analytic and we have $f'(z) = \lim_{n\to+\infty} f'_n(z)$, for every $z \in U$.

Let (f_n) be a series of analytic functions on U, and suppose that $\sum_{n\geq 0} f_n$ converge uniformly on every disc in U. Then $f = \sum_{n\geq 0} f_n$ is analytic on U and we also have $f'(z) = \sum_{n\geq 0} f'_n(z)$, for every $z \in U$.

Let $z \to f(t,z)$ be an analytic function on U depending on a real parameter $t \in]a,b[$, if there exist a function g such that

$$\int_a^b g(t)\mathrm{d}t < +\infty$$

and

$$|f(t,z)| \leq g(t)$$

for all $z \in U$, then the function

$$z \to \int_a^b f(t,z)\mathrm{d}t$$

is analytic on U.

2 Analytic continuation and singularities

2.1 The problem of analytic continuation

Let f be an analytic function on an open set U, and let V be an open set containing U. We seek a function g, analytic on V, such that $g = f$ on U.

We say that such a g is an *analytic continuation* of f to V.

If V is a connected open set containing U, then the analytic continuation g of f to V, if it exists, is unique.

On the other hand, the existence of an analytic continuation g of f to V is not guaranteed.

2.2 Isolated singularities

The obstructions to analytic continuation are the points or sets of points that we call singularities.

More precisely, if U is a non-empty open set, and a is a point on the boundary of U, then we say that a is a *singularity* of f if there is no analytic continuation of f to $U \cup D(a,r)$ for any disc $D(a,r)$ with $r > 0$.

The most simple singularities are the *isolated singularities*: a singularity a of f is an isolated singularity, if f is analytic in a punctured disc $D(a,R)\setminus\{a\}$ for some $R>0$, but there is no analytic continuation of f to $D(a,r)$.

We can distinguish two types of isolated singularity, depending on the behaviour of $f(z)$ as $z \to a$. If $|f(z)| \to +\infty$ as $z \to a$ we say that a is a *pole* of f, otherwise we say that a is an *essential singularity* of f, this is the case for example if we take $\exp(1/z)$ at 0.

2.3 Laurent expansion

If the point a is a pole of f, then there is a disc $D(a,R)$ with $R>0$, such that

$$f(z) = \frac{c_{-m}}{(z-a)^m} + \ldots + \frac{c_{-1}}{(z-a)} + \sum_{n=0}^{+\infty} c_n (z-a)^n \text{ for every } z \in D(a,r)\setminus\{a\}.$$

This is called the *Laurent expansion* of f about a, and the singular part

$$\frac{c_{-m}}{(z-a)^m} + \ldots + \frac{c_{-1}}{(z-a)}$$

is called the *principal part* of f at a.

If a is an essential singularity of f, then the expansion above becomes $\sum_{n=-\infty}^{+\infty} c_n (z-a)^n$ with an infinite number of non-zero c_n such that $n<0$.

2.4 Residue theorem

Let U be an open set, $a \in U$ and f an analytic function in $U\setminus\{a\}$. The coefficient c_{-1} of the Laurent expansion of f about a is called the *residue* of f at a, denoted $\mathrm{Res}(f,a)$. This number is all that is needed to calculate the integral of f around a small closed path winding round a.

More precisely, for every closed path γ homotopic in $U\setminus\{a\}$ to a circle centred at a we have

$$\int_\gamma f(z)\mathrm{d}z = 2\mathrm{i}\pi\,\mathrm{Res}(f,a).$$

We deduce that if U is a simply-connected open set, if $a_1, a_2, \ldots, a_n$ are points in U and f is an analytic function in $U\setminus\{a_1,a_2,\ldots,a_n\}$, then we have

$$\int_\gamma f(z)\mathrm{d}z = 2\mathrm{i}\pi \sum_{i=1}^{n} \mathrm{Res}(f,a_i),$$

where γ is a closed path in $U\setminus\{a_1,a_2,\ldots,a_n\}$ such that for every i the curve γ is homotopic in $U\setminus\{a_i\}$ to a circle centre a_i.

2.5 The logarithm

There exist examples of singularities that are not isolated but are branch points; we see an example when we try to define the function log on $\mathbb{C}$.

We can define the function log by

$$\log(z) = \int_1^z \frac{1}{u} \mathrm{d}u$$

where we integrate along the complex line segment joining 1 to z.

Since the line segment must avoid 0, we see that this function is defined and analytic on $\mathbb{C} \setminus]-\infty, 0]$; we call it the *principal value of the complex logarithm.*

We write arg for the continuous function on $\mathbb{C} \setminus]-\infty, 0]$ with values in $]-\pi, +\pi]$, such that $z = |z|\mathrm{e}^{\mathrm{i} \arg(z)}$ for each $z \in \mathbb{C} \setminus]-\infty, 0]$, and we call this function the *principal value of the argument.*

One can check that

$$\log(z) = \ln|z| + \mathrm{i} \arg(z) \qquad \text{for every } z \in \mathbb{C} \setminus]-\infty, 0].$$

It follows that $\mathrm{e}^{\log(z)} = z$ and that log has a discontinuity of $2\mathrm{i}\pi$ on the half-line $]-\infty, 0[$, that is,

$$\lim_{\varepsilon \to 0+} \log(x + \mathrm{i}\varepsilon) - \log(x - \mathrm{i}\varepsilon) = 2\mathrm{i}\pi \qquad \text{for every } x \in]-\infty, 0[.$$

Thus the point 0 is a singularity of log, but not an isolated singularity since log cannot be continued analytically to a disc centred at 0. The point 0 is a singularity of log called a *branch point.*

Let U be a connected open set; then we call any analytic function log on U satisfying $\mathrm{e}^{\log(z)} = z$ for all $z \in U$ a *branch of the logarithm* in U.

We call a continuous function θ on a connected open set U a *branch of the argument* if for each $z \in U$ we have $z = |z|\mathrm{e}^{\mathrm{i}\theta(z)}$.

Every branch of the logarithm in U can be written

$$\log(z) = \ln(|z|) + \mathrm{i}\theta(z),$$

where θ is a branch of the argument in U. Conversely, each branch of the argument allows us to define a branch of the logarithm, by the above formula. For example we define a branch of the logarithm on $\mathbb{C} \setminus [0, +\infty[$ by

$$\mathrm{Log}(z) = \ln|z| + \mathrm{i}\,\mathrm{Arg}(z) \qquad \text{for every } z \in \mathbb{C} \setminus [0, +\infty[$$

where Arg is the continuous function on $\mathbb{C} \setminus [0, +\infty[$ with values in $]0, +2\pi[$, such that $z = |z|\mathrm{e}^{\mathrm{i}\,\mathrm{Arg}(z)}$ for each $z \in \mathbb{C} \setminus [0, +\infty[$.

3 Continuation of a power series

Let $f(z) = \sum_{n\geq 0} a_n z^n$ be a power series; this will have a natural domain of convergence that is a disc $D(0, R)$ in the complex plane, where the radius of convergence R is given by

$$R = \sup\{r \geq 0,\ \text{there exists } C > 0 \text{ such that } |a_n| \leq \frac{C}{r^n} \text{ for all } n\}.$$

When $R = +\infty$, we can calculate the value of $f(z)$ at every point $z \in \mathbb{C}$ as the limit of the partial sums

$$f(z) = \lim_{N\to+\infty} \sum_{n\geq 0}^{N} a_n z^n.$$

If the radius of convergence of $\sum_{n\geq 0} a_n z^n$ is a finite number $R > 0$ (we shall look at the case $R = 0$ later), then the above formula allows us to calculate $f(z)$ for z in the disc $D(0, R)$, and the function f defined by the sum of the power series in $D(0, R)$ is analytic in this disc. There will exist at least one singularity z_0 of f on the boundary of the disc (there can be more than one, indeed even an infinite number, the whole circle $C(0, R)$ may consist of singularities).

We will say that f can be continued analytically along a half-line d starting at 0 if there exists an open set U containing d and a function g, analytic on U, such that

$$g(z) = f(z) \text{ for all } z \in U \cap D(0, R).$$

We shall suppose that f can be continued analytically along all but finitely many half-lines.

There is then an open set Star(f), the star domain of holomorphy of f. To give a formula allowing us to calculate f in this open set, we shall begin by giving, an expression for f in the interior of the disc of convergence, in terms of a Laplace integral.

An integral formula

To begin, we improve the convergence of the series $\sum_{n\geq 0} a_n z^n$ by multiplying the a_n by $1/n!$; thus we consider the series

$$\mathcal{B}(f)(\xi) = \sum_{n\geq 0} \frac{a_n}{n!} \xi^n.$$

Since $|a_n|$ is bounded by C/r^n with $0 < r < R$, it is easy to see that this series has an infinite radius of convergence and defines an analytic function $\mathcal{B}(f)$ on the whole of $\mathbb{C}$.

To recover f from $\mathcal{B}(f)$ we shall use the fact that

$$\int_0^{+\infty} \mathrm{e}^{-t}\frac{a_n}{n!}(zt)^n \mathrm{d}t = a_n z^n.$$

However, for each z in $D(0,R)$ there exists r such that $0 < |z| < r < R$, so that

$$\int_0^{+\infty} \mathrm{e}^{-t}\sum_{n\geq 0}\frac{|a_n z^n|}{n!}t^n \mathrm{d}t \leq \int_0^{+\infty} \mathrm{e}^{-t}C\mathrm{e}^{t|z|/r}\mathrm{d}t < +\infty.$$

Thus we can write

$$\int_0^{+\infty}\sum_{n\geq 0}\mathrm{e}^{-t}\frac{a_n z^n}{n!}t^n \mathrm{d}t = \sum_{n\geq 0}\int_0^{+\infty} \mathrm{e}^{-t}\frac{a_n z^n}{n!}t^n \mathrm{d}t,$$

giving, for each z in $D(0,R)$, the expression

$$\int_0^{+\infty} \mathrm{e}^{-t}\sum_{n\geq 0}\frac{a_n}{n!}(zt)^n \mathrm{d}t = \sum_{n\geq 0} a_n z^n.$$

Thus in the disc $D(0,R)$ we can write

$$f(z) = \int_0^{+\infty} \mathrm{e}^{-t}\mathcal{B}(f)(zt)\mathrm{d}t.$$

This formula will allow us to continue f analytically beyond $D(0,R)$.

Remark. For z in $[0,R[$, we can write

$$f(z) = \frac{1}{z}\int_0^{+\infty} \mathrm{e}^{-\xi/z}\mathcal{B}(f)(\xi)\mathrm{d}\xi.$$

If we define the Laplace transform of a function h by

$$\mathcal{L}(h)(z) = \int_0^{+\infty} \mathrm{e}^{-z\xi}h(\xi)\mathrm{d}\xi,$$

we then have, for every z in $[0,R[$, the expression

$$f(z) = \frac{1}{z}\mathcal{L}(\mathcal{B}(f))(\frac{1}{z}).$$

Note that the function $g : z \to \frac{1}{z}\int_0^{+\infty} \mathrm{e}^{-\xi/z}\mathcal{B}(f)(\xi)\mathrm{d}\xi$ is analytic in every domain on which the function

$$\xi \to \mathrm{e}^{-\xi\,\mathrm{Re}(1/z)}|\mathcal{B}(f)(\xi)|$$

is majorized by an integrable function on $]0,+\infty[$, independently of z. Now we know that $|a_n|$ is majorized by Const. $/(R-\varepsilon)^n$, and so we have

$$|\mathcal{B}(f)(\xi)| \leq C\mathrm{e}^{\xi/(R-\varepsilon)} \text{ for all } \varepsilon > 0;$$

the function g is therefore analytic in the open set $\{z \mid \operatorname{Re}(1/z) > 1/R\}$, i.e., the disc $D(R/2, R/2)$.

Remark. If the function $\mathcal{B}(f)$ is such that we have a better bound,

$$|\mathcal{B}(f)(\xi)| \leq C\mathrm{e}^{B\xi}$$

with $B < 1/R$, we then obtain an analytic continuation of f in the open set $\operatorname{Re}(1/z) > B$, i.e., the disc $D(1/2B, 1/2B)$.

Continuation outside the disc

We note first that if the integral $\int_0^{+\infty} \mathrm{e}^{-t}\mathcal{B}(f)(zt)\mathrm{d}t$ converges for $z = z_0$, then it converges for all z in the segment $[0, z_0]$; indeed it is enough to write, for z in the segment $[0, z_0]$,

$$\begin{aligned}\int_0^{+\infty} \mathrm{e}^{-t}\mathcal{B}(f)(zt)\mathrm{d}t &= \int_0^{+\infty} \mathrm{e}^{-t}\mathcal{B}(f)(z_0\frac{z}{z_0}t)\mathrm{d}t \\ &= (\frac{z_0}{z})\int_0^{+\infty} \mathrm{e}^{-(z_0/z)u}\mathcal{B}(f)(z_0u)\mathrm{d}u,\end{aligned}$$

and since this last integral converges for $z_0/z = 1$, it does so for $z_0/z > 1$, i.e., for in the segment $[0, z_0]$ and even for those z with $\operatorname{Re}(z_0/z) > 1$, by the following lemma:

Lemma 1 (Classical lemma). *If a is a locally integrable function on $[0,+\infty[$ such that $\int_0^{+\infty} \mathrm{e}^{-t}a(t)\mathrm{d}t$ converges, then $\int_0^{+\infty} \mathrm{e}^{-st}a(t)\mathrm{d}t$ converges for every s such that $\operatorname{Re}(s) > 1$, and the integral defines an analytic function of s in this half-plane.*

Let us consider the function

$$z \to \frac{z_0}{z}\int_0^{+\infty} \mathrm{e}^{-(z_0/z)u}\mathcal{B}(f)(z_0u)\mathrm{d}u;$$

this function is analytic in the open set consisting of all z such that

$$\operatorname{Re}(\frac{z_0}{z}) > 1,$$

that is, the disc $D(z_0/2, |z_0|/2)$.

To sum up, if the integral $\int_0^{+\infty} \mathrm{e}^{-t}\mathcal{B}(f)(zt)\mathrm{d}t$ converges for $z = z_0$, then it converges for all z in the open set $D(z_0/2, |z_0|/2)$, and defines an analytic function in this open set. This function equals f on the line segment $[0, z_0] \cap$

$D(0,R)$; by the uniqueness theorem it therefore equals f on $D(z_0/2, |z_0|/2) \cap D(0,R)$, and so we obtain an analytic continuation of f.

Consider the open set

$$E(f) = \{z_0 \in \mathrm{Star}(f),\ \text{there exists } \varepsilon > 0 \text{ such that } D(\frac{z_0}{2}, \frac{|z_0|}{2} + \varepsilon) \subset \mathrm{Star}(f)\};$$

we shall show that the function $z \to \int_0^{+\infty} \mathrm{e}^{-t}\mathcal{B}(f)(zt)\mathrm{d}t$ is defined and analytic in this open set, and therefore provides an analytic continuation of f into $E(f)$. This is a consequence of the preceding discussion together with the following lemma:

Lemma 2. *For every $z \in E(f)$ the integral $\int_0^{+\infty} \mathrm{e}^{-t}\mathcal{B}(f)(zt)\mathrm{d}t$ converges and its value is $f(z)$.*

Proof of the Lemma. Take z in $E(f)$; we deform the contour $C(z/2, |z|/2)$ to a slightly bigger contour C' surrounding 0 such that if $\xi \in C'$ then we have $\mathrm{Re}(z/\xi) < 1$.

By Cauchy's formula we have

$$f(z) = \frac{1}{2\mathrm{i}\pi} \int_{C'} \frac{f(\xi)}{\xi - z} \mathrm{d}\xi,$$

and now we see that if $\mathrm{Re}(z/\xi) < 1$ then

$$\frac{1}{1 - \frac{z}{\xi}} = \int_0^{+\infty} \mathrm{e}^{-t}\mathrm{e}^{tz/\xi}\mathrm{d}t.$$

Substituting this into Cauchy's formula we have

$$\begin{aligned} f(z) &= \frac{1}{2\mathrm{i}\pi} \int_{C'} \frac{f(\xi)}{\xi} \int_0^{+\infty} \mathrm{e}^{-t}\mathrm{e}^{tz/\xi}\mathrm{d}t \\ &= \int_0^{+\infty} \mathrm{e}^{-t}(\frac{1}{2\mathrm{i}\pi} \int_{C'} \frac{f(\xi)}{\xi} \mathrm{e}^{zt/\xi}\mathrm{d}\xi)\mathrm{d}t. \end{aligned}$$

Deforming C' into a small circle $C(0,r)$ contained in $D(0,R)$, we have

$$\begin{aligned} &\frac{1}{2i\pi} \int_{C'} \frac{f(\xi)}{\xi} \sum_{n \geq 0} \frac{1}{n!} (\frac{z}{\xi}t)^n \mathrm{d}\xi \\ &= \sum_{n \geq 0} \frac{1}{n!} (zt)^n \frac{1}{2\mathrm{i}\pi} \int_{C(0,r)} \frac{f(\xi)}{\xi^{n+1}} \mathrm{d}\xi \\ &= \sum_{n \geq 0} \frac{a_n}{n!} (zt)^n \\ &= \mathcal{B}(f)(zt). \end{aligned}$$

We deduce that the integral $\int_0^{+\infty} \mathrm{e}^{-t}\mathcal{B}(f)(zt)\mathrm{d}t$ converges to the value $f(z)$.

Continuation in the star domain

To have an analytic continuation of f in the star domain of holomorphy of f, we improve the convergence of the series $\sum_{n\geq 0} a_n z^n$ in a more delicate way. In fact it is enough to multiply the a_n by a term which behaves like $(1/n!)^\alpha$ with $0 < \alpha \leq 1$, we take the term $1/\Gamma(1+n\alpha)$ where

$$\Gamma(1+n\alpha) = \int_0^{+\infty} \mathrm{e}^{-t} t^{\alpha n} \mathrm{d}t \ .$$

So we consider the series

$$\mathcal{B}_\alpha(f)(\xi) = \sum_{n\geq 0} \frac{a_n}{\Gamma(1+n\alpha)} \xi^n.$$

This series has infinite radius of convergence, and defines an analytic function $\mathcal{B}_\alpha(f)$ on the whole complex plane $\mathbb{C}$.

To recover f from $\mathcal{B}_\alpha(f)$ we use the fact that

$$\int_0^{+\infty} \mathrm{e}^{-t} \frac{a_n z^n}{\Gamma(1+n\alpha)} t^{\alpha n} \mathrm{d}t = a_n z^n,$$

and obtain, for every z in the disc $D(0,R)$,

$$f(z) = \int_0^{+\infty} \mathrm{e}^{-t} \mathcal{B}_\alpha(f)(z t^\alpha) \mathrm{d}t.$$

This formula will allow us to continue f analytically outside $D(0,R)$.

We notice as above that if the integral $\int_0^{+\infty} \mathrm{e}^{-t} \mathcal{B}_\alpha(f)(z t^\alpha) \mathrm{d}t$ converges for $z = z_0$, then it converges for all z in the line segment $[0, z_0]$; indeed it is enough to write

$$\begin{aligned}\int_0^{+\infty} \mathrm{e}^{-t} \mathcal{B}_\alpha(f)(z t^\alpha) \mathrm{d}t &= \int_0^{+\infty} \mathrm{e}^{-t} \mathcal{B}_\alpha(f)(z_0 \frac{z}{z_0} t^\alpha) \mathrm{d}t \\ &= (\frac{z_0}{z})^{1/\alpha} \int_0^{+\infty} \mathrm{e}^{-(z_0/z)^{1/\alpha} u} \mathcal{B}_\alpha(f)(z_0 u^\alpha) \mathrm{d}u,\end{aligned}$$

and since this last integral converges for $(z_0/z)^{1/\alpha} = 1$, it does so also for $(z_0/z)^{1/\alpha} > 1$ and even for $\mathrm{Re}(z_0/z)^{1/\alpha} > 1$.

The function

$$z \to (\frac{z_0}{z})^{1/\alpha} \int_0^{+\infty} \mathrm{e}^{-(\frac{z_0}{z})^{1/\alpha} u} \mathcal{B}_\alpha(f)(z_0 u^\alpha) \mathrm{d}u$$

is analytic in the connected open set $D_\alpha(z_0)$ containing $]0, z_0]$ consisting of those z such that

$$\mathrm{Re}(\frac{z_0}{z})^{1/\alpha} > 1.$$

This is a rather thin convex open set, whose boundary $C_\alpha(z_0)$ has the following equation in polar coordinates:

$$\varrho = \varrho_0 (\cos \frac{\theta - \theta_0}{\alpha})^\alpha$$
$$-\frac{\pi}{2}\alpha < \theta - \theta_0 < \frac{\pi}{2}\alpha.$$

The smaller α is, the thinner $D_\alpha(z_0)$ is.
Summary. If the integral $\int_0^{+\infty} \mathrm{e}^{-t}\mathcal{B}_\alpha(f)(zt^\alpha)\mathrm{d}t$ converges for $z = z_0$, then it converges for all z in the open set $D_\alpha(z_0)$, and defines an analytic function in this open set, which is a continuation of f.

Consider the open set

$$E_\alpha(f) = \{z \in \mathrm{Star}(f) \text{ and } D_\alpha(z) \cup C_\alpha(z_0) \subset \mathrm{Star}(f)\}.$$

One can show that for every $z \in E_\alpha(f)$ the integral $\int_0^{+\infty} \mathrm{e}^{-t}\mathcal{B}_\alpha(f)(zt^\alpha)\mathrm{d}t$ converges.

We see therefore that on $E_\alpha(f)$ the function

$$z \to \int_0^{+\infty} \mathrm{e}^{-t}\mathcal{B}_\alpha(f)(zt^\alpha)\mathrm{d}t$$

is an analytic continuation of f.

Since

$$\mathrm{Star}(f) = \bigcup_{0<\alpha\leq 1} E_\alpha(f),$$

we therefore have a means of calculating the continuation of f for every z in $\mathrm{Star}(f)$.

4 Gevrey series

4.1 Definitions

If the radius of convergence of $\sum_{n\geq 0} a_n z^n$ is zero, then the power series $F = \sum_{n\geq 0} a_n z^n$ cannot define an analytic function f by the formula

$$f(z) = \lim_{N\to+\infty} \sum_{n\geq 0}^{N} a_n z^n,$$

since this limit does not exist for any $z \neq 0$.

We shall therefore weaken the concept of convergence, looking for an analytic function f on an open set U, with 0 in U or on the boundary of U, such that the series $\sum_{n\geq 0} a_n z^n$ is an asymptotic expansion of f in the following sense:

$$|f(z) - \sum_{n\geq 0}^{N-1} a_n z^n| \leq C_N |z|^N \text{ for all } z \in U,$$

with the above holding for every $N \geq 0$, and with C_N independent of z, although the C_N are allowed to tend to infinity.

We cannot hope that such an asymptotic condition could hold on an open disc $U = D(0,R)$ with $R > 0$, as that would imply that

$$a_n = \frac{\partial^n f(0)}{n!},$$

and since f is supposed to be analytic on $U = D(0,R)$, the series $\sum_{n\geq 0} a_n z^n$ would converge in $D(0,R)$ and hence would have a non-zero radius of convergence.

We shall therefore require that the above condition holds in a *small sector* S based at 0 of angle $\theta_1 - \theta_0$ less than 2π, i.e.,

$$S = \{z = r\mathrm{e}^{\mathrm{i}\theta} \mid 0 < r < R, \theta_0 < \theta < \theta_1\}.$$

In this case, one can show (this is the Borel–Ritt theorem) that for every power series $\sum_{n\geq 0} a_n z^n$ there exists an analytic function f on S, such that the series $\sum_{n\geq 0} a_n z^n$ is the asymptotic expansion of f about 0 in S, but the function f is not unique (for example, if the sector is contained in $\mathbb{C} \setminus]-\infty, 0]$, it is possible to add to f the function $z \to \mathrm{e}^{-1/\sqrt{z}}$).

To obtain uniqueness results, we shall strengthen slightly the asymptotic condition, as it leaves too much freedom in the terms $C_N|z|^N$ since C_N can grow arbitrarily as $N \to +\infty$.

To make this precise, we introduce the condition of *Gevrey asymptoticity* in a small sector S which consists of requiring of C_N a growth rate of at most $B^N N!$, and one then requires that

$$|f(z) - \sum_{n\geq 0}^{N-1} a_n z^n| \leq C B^N N!\, |z|^N \text{ for all } z \in S,$$

the above holding for all $N \geq 0$, with constants $C > 0$ and $B > 0$ independent of $z \in S$.

This condition implies that

$$\frac{f(z) - \sum_{n\geq 0}^{N-1} a_n z^n}{z^N} - a_N \to 0 \text{ when } z \to 0 \text{ in } S,$$

and so it cannot be satisfied unless the coefficients a_n also satisfy an inequality like

$$|a_n| \le CB^n n!\ .$$

We say in this case that the series $\sum_{n\ge 0} a_n z^n$ is *Gevrey* (or Gevrey of order 1).

We shall write this condition of Gevrey asymptoticity in S in the form

$$f(z) \backsim \sum_{n\ge 0} a_n z^n \text{ in } S.$$

4.2 Exponential smallness and uniqueness

In the condition of Gevrey asymptoticity

$$|f(z) - \sum_{n\ge 0}^{N-1} a_n z^n| \le CB^N N!\,|z|^N$$

the function $R : N \to CB^N N!\,|z|^N$ is first decreasing and then increasing, and so it has a minimum at $N_0 \backsimeq (B|z|)^{-1}$, and takes a minimal value

$$R(N_0) \backsimeq A|z|^{-1/2}\mathrm{e}^{-\frac{1}{B|z|}}$$

with $A > 0$.

We therefore have an exponentially small remainder (when $z \to 0$ in S) if we take the sum as far as N_0 (this justifies the method of summation up to the smallest term, or the “astronomers’ method”).

Note that this implies that if we have

$$f(z) \backsim \sum_{n\ge 0} 0z^n \text{ in } S,$$

then the function f is exponentially decreasing in S, i.e.,

$$|f(z)| \le C\mathrm{e}^{-D/|z|} \text{ in } S.$$

Conversely, one can show that this inequality implies that $f(z) \backsim \sum_{n\ge 0} 0z^n$ in S.

Conclusion. Given a formal series $F = \sum_{n\ge 0} a_n z^n$ that is Gevrey, *we do not have uniqueness* of the function f such that

$$f(z) \backsim \sum_{n\ge 0} a_n z^n \text{ in } S\ :$$

it is enough to add to f an analytic function decreasing exponentially in S.

4.3 Gevrey summability

Given a divergent series of Gevrey type $\sum_{n\geq 0} a_n z^n$ a small sector S, does there exist an unique analytic function f such that

$$f(z) \backsim \sum_{n\geq 0} a_n z^n \text{ in } S\,?$$

If one wants to guarantee the uniqueness of f it is enough to require that the condition of Gevrey asymptoticity holds *on a small sector S of angle $> \pi$*. Indeed, in this case one can show that the only analytic function of exponential decrease in S is the zero function.

What about the existence of f? The condition: $|a_n| \leq CB^n n!$ for all n, guarantees the convergence of the power series

$$\mathcal{B}(F)(\xi) = \sum_{n\geq 0} \frac{a_n}{n!} \xi^n$$

for all ξ in the disc $D(0, 1/B)$, and defines an analytic function in this disc.

For $0 < \varrho < 1/B$, we can define an analytic function

$$f_\varrho(z) = \frac{1}{z} \int_0^\varrho \mathrm{e}^{-(\xi/z)} \mathcal{B}(F)(\xi) \mathrm{d}\xi$$

for z in $\mathbb{C} \setminus \{0\}$. We can show that we have

$$f_\varrho(z) \backsim \sum_{n\geq 0} a_n z^n$$

in every small sector $S' = \{z = re^{i\theta} \text{ with } -\frac{\pi}{2} + \varepsilon < \theta < \frac{\pi}{2} - \varepsilon\}$ of angle $< \pi$. The disadvantage of this construction is the arbitrary choice of ϱ, since all we can say is that $f_\varrho - f_{\varrho'}$ is an analytic function decreasing exponentially in S. If one wants to guarantee existence and uniqueness of f we will need to impose stronger hypotheses on the function $\mathcal{B}(F)$.

5 Borel summability

Let $F = \sum_{n\geq 0} a_n z^n$ be a power series satisfying the Gevrey condition: $|a_n| \leq CB^n n!$ for all n. The function

$$\mathcal{B}(F)(\xi) = \sum_{n\geq 0} \frac{a_n}{n!} \xi^n$$

is defined and analytic in the disc $D(0, 1/B)$. If we want to avoid the arbitrary choice of ϱ as above, we try to define the function

$$f(z) = \frac{1}{z}\int_0^{+\infty} \mathrm{e}^{-(\xi/z)}\mathcal{B}(F)(\xi)\mathrm{d}\xi.$$

To guarantee the existence of the integral we shall suppose that the function $\mathcal{B}(F)$ is continued analytically in a sector $S = \{z = r\mathrm{e}^{\mathrm{i}\theta}$ with $-\varepsilon < \theta < +\varepsilon\}$, to give a function of at most exponential growth at infinity in this sector, i.e.,

$$|\mathcal{B}(F)(\xi)| \leq A\mathrm{e}^{B|\xi|}\ .$$

In this case we say that the series $\sum_{n\geq 0} a_n z^n$ is *Borel-summable in the direction* $\theta = 0$. The function f thereby defined is analytic in the domain $\{z \,|\, \mathrm{Re}(1/z) > B\}$, which is just the disc

$$D = D(\frac{1}{2B}, \frac{1}{2B}) = \{z = r\mathrm{e}^{\mathrm{i}\theta} \,|\, r < \frac{1}{B}\cos(\theta)\},$$

(or if $B = 0$ it is the half-plane $\mathrm{Re}(z) > 0$). Let $\varphi \in \,]-\varepsilon, \varepsilon[$; then, setting

$$f_\varphi(z) = \frac{1}{z}\int_0^{+\infty\mathrm{e}^{\mathrm{i}\varphi}} \mathrm{e}^{-(\xi/z)}\mathcal{B}(F)(\xi)\mathrm{d}\xi,$$

we obtain a function f_φ defined and analytic in the disc

$$D_\varphi = \{z = r\mathrm{e}^{\mathrm{i}\psi} \,|\, r < \frac{1}{B}\cos(\psi - \varphi)\},$$

which is just the disc $D(1/2B, 1/2B)$ rotated by the angle φ.

For $z \in D \cap D_\varphi$ we see, using the analyticity of $\xi \to \mathrm{e}^{-(\xi/z)}\mathcal{B}(f)(\xi)$ and its decay at infinity, that

$$\begin{aligned} f_\varphi(z) - f(z) &= \frac{1}{z}\left(\int_0^{+\infty\mathrm{e}^{\mathrm{i}\varphi}} \mathrm{e}^{-(\xi/z)}\mathcal{B}(F)(\xi)\mathrm{d}\xi - \int_0^{+\infty} \mathrm{e}^{-(\xi/z)}\mathcal{B}(F)(\xi)\mathrm{d}\xi\right) \\ &= \frac{1}{z}\lim_{R\to+\infty}\int_{\gamma_R} \mathrm{e}^{-(\xi/z)}\mathcal{B}(F)(\xi)\mathrm{d}\xi = 0 \end{aligned}$$

(the path γ_R consisting of the arc $R\mathrm{e}^{-\mathrm{i}t}$, $t \in [0, \varphi]$).

Letting φ vary in $]-\varepsilon, \varepsilon[$, we obtain an analytic continuation of f in an open set containing a small sector S of angle strictly greater than π.

Moreover, one can show that

$$|f(z) - \sum_{n\geq 0}^{N-1} a_n z^n| \leq CB^N N!\,|z|^N \text{ for all } z \in S.$$

The function f defined this way in S is then the only analytic function in S such that

$$f(z) \backsim \sum_{n\geq 0} a_n z^n \text{ in } S.$$

We call this the *Borel sum* of the formal series $F = \sum_{n\geq 0} a_n z^n$, and we write it $f = s(F)$.

In the same way we can define the notion of *Borel-summability in the direction* $\theta \neq 0$, and we write

$$f_\theta(z) = \frac{1}{z}\int_0^{+\infty e^{i\theta}} e^{-(\xi/z)}\mathcal{B}(F)(\xi)d\xi.$$

The function f_θ defined this way in $e^{i\theta}S$ is the the only analytic function in $e^{i\theta}S$ such that

$$f_\theta(z) \backsim \sum_{n\geq 0} a_n z^n \text{ in } e^{i\theta}S,$$

and we call it the *Borel sum in the direction* θ of the formal series $F = \sum_{n\geq 0} a_n z^n$, we note it $s_\theta(F)$.

Remark. If the series $F = \sum_{n\geq 0} a_n z^n$ has radius of convergence $R > 0$, then it is Borel-summable in every direction θ and the Borel sums $s_\theta(F)$ give the analytic continuation of the function $f : z \to \sum_{n\geq 0} a_n z^n$ to an open set containing $D(0, R)$.

Properties of s_θ

a) s_θ is linear:

$$s_\theta(F + G) = s_\theta(F) + s_\theta(G),$$
$$s_\theta(c \cdot F) = c \cdot s_\theta(F) \text{ if } c \in \mathbb{C},$$

since J, $\mathcal{L}_\theta$ and $\mathcal{B}$ are linear.

b) s_θ commutes with differentiation $\partial = d/dz$:

$$s_\theta(\partial F) = \partial s_\theta(F).$$

c) s_θ is a morphism:

$$s_\theta(F \cdot G) = s_\theta(F)s_\theta(G)$$

(where the product $F \cdot G$ denotes the usual product of formal series).

5.1 Connection with the usual Laplace transform

The integral formula

$$s_\theta(F)(z) = \frac{1}{z}\int_0^{+\infty e^{i\theta}} e^{-(\xi/z)}\mathcal{B}(F)(\xi)d\xi,$$

which we use to construct the Borel sum of $F = \sum_{n\geq 0} a_n z^n$, can be expressed in terms of an ordinary Laplace integral as

$$\frac{1}{z}\mathcal{L}_\theta(\mathcal{B}(F))(\frac{1}{z})$$

where

$$\mathcal{L}_\theta(g)(z) = \int_0^{+\infty e^{i\theta}} e^{-z\xi} g(\xi) d\xi.$$

Let J be the mapping

$$h \to J(h),$$
$$J(h)(z) = \frac{1}{z}h(\frac{1}{z}).$$

This satisfies $J \circ J = Id$ and it interchanges behaviour at 0 and behaviour at ∞, as

$$J(\sum_{n\geq 0} a_n z^n) = \sum_{n\geq 0} a_n \frac{1}{z^{n+1}}.$$

We then have

$$s_\theta = J \circ \mathcal{L}_\theta \circ \mathcal{B}.$$

The behaviour at 0 of $s_\theta(F)$ is then linked to the behaviour at ∞ of $\mathcal{L}_\theta(\mathcal{B}(F))$.

The asymptoticity condition at 0:

$$f_\theta(z) \backsim \sum_{n\geq 0} a_n z^n \text{ in } S_\theta,$$
$$S_\theta = \{z = re^{i\psi} \,|\, r < R \text{ and } \theta - \frac{\pi}{2} - \varepsilon < \psi < \theta + \frac{\pi}{2} + \varepsilon\},$$

translates into the asymptoticity condition at ∞:

$$\mathcal{L}_\theta(\mathcal{B}(F))(z) \backsim \sum_{n\geq 0} a_n \frac{1}{z^{n+1}} \text{ in } S_{\infty,\theta},$$
$$S_{\infty,\theta} = \{z = re^{i\varphi} \,|\, r > 1/R \text{ and } -\theta - \frac{\pi}{2} - \varepsilon < \varphi < -\theta + \frac{\pi}{2} + \varepsilon\}.$$

Given a formal series $F = \sum_{n\geq 0} a_n z^n$ that is Borel-summable in the direction θ, then the function

$$\mathcal{L}_\theta(\mathcal{B}(F))(z) = \int_0^{+\infty e^{i\theta}} e^{-z\xi} \mathcal{B}(F)(\xi) d\xi$$

can be continued analytically in the sector

$$S_{\infty,\theta} = \{z = re^{i\varphi} \,|\, r > 1/R \text{ and } -\theta - \frac{\pi}{2} - \varepsilon < \varphi < -\theta + \frac{\pi}{2} + \varepsilon\},$$

and satisfies

$$\mathcal{L}_\theta(\mathcal{B}(F))(z) \backsim \sum_{n\geq 0} a_n \frac{1}{z^{n+1}} \text{ in } S_{\infty,\theta}.$$

5.2 Alien derivations

Let F be a formal power series; then some ambiguities in summation can appear in directions θ in which the function $\mathcal{B}(F)$ has singularities.

Suppose for example that $\mathcal{B}(F)$ possesses a singularity $\omega = r\mathrm{e}^{\mathrm{i}\theta}$, $r \neq 0$, and that in a sector S containing the half-line in the direction θ one has

$$\mathcal{B}(F)(\xi) = \frac{1}{2\mathrm{i}\pi}\varphi(\xi-\omega)\,\mathrm{Log}(\xi-\omega) + \psi(\xi-\omega),$$

where φ and ψ are analytic in an open neighbourhood of $\omega + S$ with subexponential growth.

If we take two half-lines in S in the directions $\theta_- < \theta$ and $\theta_+ > \theta$, we have

$$\begin{aligned}\mathcal{L}_{\theta-}(\mathcal{B}(F))(z) - \mathcal{L}_{\theta+}(\mathcal{B}(F))(z) &= \int_{\omega}^{+\infty\mathrm{e}^{\mathrm{i}\theta}} \mathrm{e}^{-z\xi}\varphi(\xi-\omega)\mathrm{d}\xi \\ &= \mathrm{e}^{-\omega z}\mathcal{L}_{\theta}(\varphi)(z).\end{aligned}$$

Suppose that $\varphi = \mathcal{B}(\Phi)$ where Φ is a formal power series, we have

$$\mathcal{L}_{\theta-}(\mathcal{B}(F)) - \mathcal{L}_{\theta+}(\mathcal{B}(F)) = \mathrm{e}^{-\omega z}\mathcal{L}_{\theta}(\mathcal{B}(\Phi)).$$

We can write this as

$$\mathcal{L}_{\theta-}(\mathcal{B}(F)) = \mathcal{L}_{\theta+}(\mathcal{B}(F)) + \mathrm{e}^{-\omega z}\mathcal{L}_{\theta+}(\mathcal{B}(\Phi)),$$

thus

$$s_{\theta-}(F) = s_{\theta+}(F) + \mathrm{e}^{-\omega/z}s_{\theta+}(\Phi).$$

The ambiguity in the summation shows itself in the appearance of the exponential $\mathrm{e}^{-\omega/z}$ multiplied by the function $s_{\theta+}(\Phi)$. To allow for this we extend the summation operators s_θ to the formal products $\mathrm{e}^{-\omega/z}(\Phi)$ by

$$s_\theta(\mathrm{e}^{-\omega/z}(\Phi)) = \mathrm{e}^{-\omega/z}s_\theta(\Phi)).$$

We can then write

$$s_{\theta-}(F) = s_{\theta+}(F + \mathrm{e}^{-\omega/z}(\Phi)),$$

where the formal series Φ only depends on F and ω, since the singular part of $\mathcal{B}(F)$ at ω is

$$\frac{1}{2\mathrm{i}\pi}\mathcal{B}(\Phi)(\xi-\omega)\,\mathrm{Log}(\xi-\omega).$$

We shall write $S_\omega F = \Phi$; then $S_\omega F$ describes the singularity of $\mathcal{B}(F)$ at the point ω, and we then have

$$s_{\theta-}(F) = s_{\theta+}(F + \mathrm{e}^{-\omega/z} S_\omega F).$$

This formula can be generalized to other singularities than logarithmic ones; it is the basis of the definition of alien derivations due to J. Ecalle. Let us show that S_ω is a derivation, i.e., that it satisfies

$$S_\omega(FG) = (S_\omega F)G + F(S_\omega G).$$

If F and G are two formal series as above, such that we have

$$s_{\theta-}(F) = s_{\theta+}(F + \mathrm{e}^{-\omega/z} S_\omega F),$$

$$s_{\theta-}(G) = s_{\theta+}(G + \mathrm{e}^{-\omega/z} S_\omega G).$$

Using the fact that $s_{\theta-}$ and $s_{\theta+}$ are morphisms, we deduce that

$$\begin{aligned} s_{\theta-}(FG) &= s_{\theta+}((F + \mathrm{e}^{-\omega/z} S_\omega F)(G + \mathrm{e}^{-\omega/z} S_\omega G)) \\ &= s_{\theta+}(FG + \mathrm{e}^{-\omega/z}(S_\omega F)G + \mathrm{e}^{-\omega/z} F(S_\omega G) + \mathrm{e}^{-2\omega/z}(S_\omega F)(S_\omega G)). \end{aligned}$$

We see that the product of two formal power series F and G such that $\mathcal{B}(F)$ and $\mathcal{B}(G)$ have singularities at ω, can have one at ω, but the exponential $\mathrm{e}^{-2\omega/z}$ show us that we can also have a singularity at 2ω.

On the other hand, we have as above

$$s_{\theta-}(FG) = s_{\theta+}(FG + \mathrm{e}^{-\omega/z} S_\omega(FG) + \mathrm{e}^{-2\omega/z} S_{2\omega}(FG))$$

where $S_{2\omega}(FG)$ represents the singularity of $\mathcal{B}(FG)$ at 2ω.

Equating the coefficients of the exponential, we obtain

$$S_\omega(FG) = (S_\omega F)G + F(S_\omega G),$$

or, in other words, the mapping S_ω is a derivation; it is also written Δ_ω.

More generally, in order to take arbitrary products of power series, it is therefore necessary to allow $\mathcal{B}(F)$ the possibility of singularities at the points $n\omega$, $n = 1, 2, \ldots$ The ambiguity in summation is then described by all the $S_{n\omega}$, since

$$s_{\theta-}(F) = s_{\theta+}(F + \mathrm{e}^{-\omega/z} S_\omega F + \mathrm{e}^{-2\omega/z} S_{2\omega} F + \ldots).$$

The mappings $S_{n\omega}$ are defined as above, but for $n \geq 2$ they are not derivations; for example, we have

$$S_{2\omega}(FG) = (S_{2\omega}(F))G + F(S_{2\omega}(G)) + S_\omega(F) S_\omega(G).$$

We can construct derivations $\Delta_{n\omega}$ by suitable combination of the $S_{k\omega}$. To find this combination, we use the mapping

$$S(\theta) : F \to \mathrm{e}^{-\omega/z} S_\omega F + \mathrm{e}^{-2\omega/z} S_{2\omega} F + \dots .$$

Since

$$(I + S(\theta))(F) = s_{\theta+}^{-1} s_{\theta-}(F),$$

we see that

$$(I + S(\theta))(F \cdot G) = (I + S(\theta))(F) \cdot (I + S(\theta))(G).$$

We call the mapping $I + S(\theta)$ the *passage morphism in the direction θ.*

The mapping $\varDelta(\theta)$ given by

$$\varDelta(\theta) = \sum_{n \geq 1} \frac{(-1)^{n-1}}{n} (S(\theta))^n .$$

is a derivation because it satisfies

$$I + S(\theta) = \exp(\varDelta(\theta)),$$

This is the *global alien derivation in the direction θ.*

If we expand $(S(\theta))^n$ we see that we can write

$$\varDelta(\theta) = \mathrm{e}^{-\omega/z} \Delta_\omega F + \mathrm{e}^{-2\omega/z} \Delta_{2\omega} F + \dots$$

where

$$\begin{aligned}
\Delta_\omega &= S_\omega, \\
\Delta_{2\omega} &= S_{2\omega} - \frac{1}{2} S_\omega S_\omega, \\
\Delta_{3\omega} &= S_{3\omega} - \frac{1}{2}(S_\omega S_{2\omega} + S_{2\omega} S_\omega) + \frac{1}{3} S_\omega S_\omega S_\omega, \\
&\dots
\end{aligned}$$

By construction, the $\Delta_{n\omega}$ are derivations, they are not of the form $a(z)\frac{d}{dz}$, they are the “alien derivations” of J. Ecalle.

5.3 Real summation

If F is a series $\sum_{n \geq 0} a_n z^n$ where the a_n are *real*, it is natural to calculate the Borel sum of F in the real direction $\theta = 0$ in order to obtain a real sum when $z \in \mathbb{R}$. If there exist singularities of $\mathcal{B}(F)$ on $\mathbb{R}_+$, then we will have two lateral sums $s_{0+} = s_+$ and $s_{\theta-} = s_-$, and the ambiguity in summation is described by the global derivation $\Delta(0) = \Delta$.

If we take as the sum

$$s(F) = \frac{1}{2}(s_+(F) + s_-(F)),$$

we do obtain a real sum for real z, although it does not necessarily have the property

$$s(FG) = s(F)s(G).$$

In order to obtain a real sum with this property, we introduce the operator C defined by

$$C(F)(z) = \overline{F(\overline{z})}.$$

We have $C(F) = F$ if F is a series $\sum_{n\geq 0} a_n z^n$ where the a_n are real; in this case we wish to determine a sum f of F such that

$$C(f) = f.$$

We may see from the explicit formula for Borel summation that $s_+C = Cs_-$. Since

$$C^2 = I \text{ and } s_+^{-1}s_- = \mathrm{e}^{\Delta},$$

we deduce that

$$Cs_+\mathrm{e}^{\Delta/2} = s_+\mathrm{e}^{\Delta/2}C.$$

This implies that

$$Cs_+\mathrm{e}^{\Delta/2}(F) = s_+\mathrm{e}^{\Delta/2}(F).$$

In other words, the function

$$s(F) = s_+\mathrm{e}^{\Delta/2}(F)$$

has the property that $s(F)(x)$ is real if x is real, and

$$s(FG) = s(F)s(G).$$

6 Acknowledgments

I am indebted for the quality of the English version of this paper to Jonathan Partington from Leeds University who kindly agreed to translate it. My warmest thanks to him.

References

1. E. Borel, Leçons sur les séries divergentes, Gabay, 1988.
2. B. Candelpergher, Une introduction à la résurgence, La Gazette des Mathématiciens, SMF, 42, 1989.
3. J. Ecalle, Les fonctions résurgentes, Publ. Math., Orsay, 1985.
4. B. Malgrange, Sommation des séries divergentes. Expo. Math. 13:163-222, 1995.
5. G. Sansone, J. Gerretsen, Lectures on the theory of functions of a complex variable, Noordhoff, Groningen, 1960.

Fourier Transforms and Complex Analysis

Jonathan R. Partington

School of Mathematics, University of Leeds,
Leeds LS2 9JT, U.K.
J.R.Partington@leeds.ac.uk

1 Real and complex Fourier analysis

1.1 Fourier series

Let f be a real or complex-valued function defined on the real line $\mathbb{R}$, having period $T > 0$, say; by this we mean that $f(t+T) = f(t)$ for all real t. Then, assuming that f is sufficiently well-behaved that the following definitions make sense (in practice this means that f is locally Lebesgue integrable), we can form its Fourier series

$$f(t) \sim \frac{a_0}{2} + \sum_{k=1}^{\infty} \left(a_k \cos \frac{2\pi k t}{T} + b_k \sin \frac{2\pi k t}{T} \right),$$

where

$$a_k = \frac{2}{T} \int_0^T f(t) \cos \frac{2\pi k t}{T} \, \mathrm{d}t \qquad \text{and} \qquad b_k = \frac{2}{T} \int_0^T f(t) \sin \frac{2\pi k t}{T} \, \mathrm{d}t,$$

are the *real Fourier coefficients* of f.

For example, consider the *sawtooth function*, $f(t) = t$ on $(-\pi, \pi]$, extended with period 2π to $\mathbb{R}$. Then

$$f(t) \sim \sum_{n=1}^{\infty} \frac{2}{n} (-1)^{n+1} \sin nt$$

(the cosine terms vanish). The Fourier series converges to the function except at odd multiples of π, where it is discontinuous.

We have used the symbol "$\sim$" rather than "$=$" above, since, even for continuous functions, the Fourier series need not converge pointwise. However, if f is C^1 (has a continuous derivative), then in fact there is no problem and the series converges absolutely. For all continuous functions the partial sums

J.-D. Fournier et al. (Eds.): Harm. Analysis and Ratio. Approx., LNCIS 327, pp. 39–55, 2006.

$$s_n(f)(t) = \frac{a_0}{2} + \sum_{k=1}^{n} \left(a_k \cos \frac{2\pi kt}{T} + b_k \sin \frac{2\pi kt}{T} \right)$$

converge in an L^2 (mean-square) sense, by which we mean that

$$\int_0^T |f(t) - s_n(f)(t)|^2 \, \mathrm{d}t \to 0 \qquad \text{as} \quad n \to \infty,$$

and there are other famous results in the literature, such as Fejér's theorem, which asserts that the Cesàro averages

$$\sigma_m(f) = \frac{1}{m+1}(s_0(f) + \ldots + s_m(f))$$

converge uniformly to f whenever f is continuous. Thus a continuous periodic function can always be approximated by trigonometric polynomials (finite sums of sines and cosines).

It is often more convenient to re-express the Fourier series using the complex exponential function $\mathrm{e}^{\mathrm{i}x} = \cos x + \mathrm{i} \sin x$, and this produces a somewhat simpler expression, namely

$$f(t) \sim \sum_{k=-\infty}^{\infty} c_k \mathrm{e}^{2\pi \mathrm{i} kt/T},$$

where

$$c_k = \frac{1}{T} \int_0^T f(t) \mathrm{e}^{-2\pi \mathrm{i} kt/T} \, \mathrm{d}t$$

are the *complex Fourier coefficients* of f, and often written $c_k = \hat{f}(k)$. Indeed, the real and complex coefficients are related by the identities

$$a_k = \hat{f}(k) + \hat{f}(-k) \qquad \text{and} \qquad b_k = \mathrm{i}(\hat{f}(k) - \hat{f}(-k)).$$

There is no essential difference between these two approaches: the partial sums are now given by

$$s_n(f) = \sum_{k=-n}^{n} c_k \mathrm{e}^{2\pi \mathrm{i} kt/T},$$

as is easily verified.

Underlying all this theory is an inner-product structure, and the basic orthogonality relation

$$\frac{1}{T} \int_0^T \mathrm{e}^{2\pi \mathrm{i} jt/T} \overline{\mathrm{e}^{2\pi \mathrm{i} kt/T}} \, \mathrm{d}t = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{otherwise,} \end{cases}$$

which can be used to deduce *Parseval's identity*, namely

$$\frac{1}{T}\int_0^T |f(t)|^2\,\mathrm{d}t = \sum_{k=-\infty}^{\infty} |\hat{f}(k)|^2.$$

This expresses the idea that the energy in a signal is the sum of the energies in each mode.

Fourier series can be used to study the vibrating string (wave equation), as well as the heat equation, which was Fourier's original motivation. We illustrate this by an example.

The temperature in a rod of length π with ends held at zero temperature is governed by the heat equation

$$\frac{\partial^2 y}{\partial x^2} = \frac{1}{K^2}\frac{\partial y}{\partial t},$$

with boundary conditions $y(0,t) = y(\pi,t) = 0$. Suppose an initial temperature distribution $y(x,0) = F(x)$.

We look for solutions $y(x,t) = f(x)g(t)$, so that

$$f''(x)g(t) = f(x)g'(t)/K^2,$$

or

$$\frac{f''(x)}{f(x)} = C = \frac{1}{K^2}\frac{g'(t)}{g(t)}.$$

It turns out we should take $f(x) = \sin nx$ (times a constant), and $C = -n^2$, in which case

$$g'(t) + K^2n^2g(t) = 0.$$

Thus one solution is

$$y(x,t) = f(x)g(t),$$

with

$$f(x) = \sin nx$$

and

$$g(t) = a_n\mathrm{e}^{-K^2n^2t}.$$

We can now superimpose solutions for different n, so we build in the initial conditions and write

$$y(x,0) = F(x) = \sum_{n=1}^{\infty} a_n \sin nx.$$

We then arrive at the formal solution

$$y(x,t) = \sum_{n=1}^{\infty} a_n \sin nx\,\mathrm{e}^{-K^2n^2t}.$$

1.2 Fourier transforms

We now move to what is sometimes regarded as a limiting case of Fourier series when T tends to infinity and infinite sums turn into integrals. Here we work with real or complex functions f defined on $\mathbb{R}$. In fact we assume that f is in $L^1(\mathbb{R})$, i.e., Lebesgue integrable on the real line; in this case we can define its Fourier transform by

$$\hat{f}(w) = \int_{-\infty}^{\infty} f(t)\mathrm{e}^{-\mathrm{i}wt}\,\mathrm{d}t.$$

This is a function of w, which is sometimes interpreted as denoting "frequency", while the variable t denotes "time". *WARNING:* one can find various alternative expressions in the literature, for example

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f(t)\mathrm{e}^{-\mathrm{i}wt}\,\mathrm{d}t \qquad \text{or} \qquad \int_{-\infty}^{\infty} f(t)\mathrm{e}^{-2\pi\mathrm{i}wt}\,\mathrm{d}t.$$

Each has its advantages and disadvantages, so we have had to make a choice. On another day we might prefer a different one.

Here is an important example. If

$$f(x) = \mathrm{e}^{-x^2/2},$$

then

$$\hat{f}(w) = \sqrt{2\pi}\mathrm{e}^{-w^2/2}\,;$$

that is, the Gaussian function is (up to a constant) the same as its Fourier transform.

In the same way that one can reconstruct a function from its Fourier series, it is possible to get back from the Fourier transform to the original function. Accordingly, define the *inverse Fourier transform* by

$$\check{g}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} g(w)\mathrm{e}^{\mathrm{i}wt}\,\mathrm{d}w = \frac{1}{2\pi}\hat{g}(-t). \tag{1}$$

We now have *Fourier's inversion theorem*, which asserts that if $f : \mathbb{R} \to \mathbb{C}$ is continuous and satisfies

$$\int_{\mathbb{R}} |f(t)|\,\mathrm{d}t < \infty \qquad \text{and} \qquad \int_{\mathbb{R}} |\hat{f}(w)|\,\mathrm{d}w < \infty, \tag{2}$$

then $(\hat{f})\check{} = f$; that is,

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(w)\mathrm{e}^{\mathrm{i}wt}\mathrm{d}w.$$

We thus deduce a uniqueness theorem for Fourier transforms, namely, that two continuous and integrable functions with the same Fourier transform must be identical.

In the interests of beauty as well as truth, we mention *Plancherel's theorem*, which is a continuous analogue of Parseval's identity. If (2) holds, and in addition

$$\int_{\mathbb{R}} |f(t)|^2 \, \mathrm{d}t < \infty$$

(more concisely: if f and $\hat{f}$ lie in $L^1(\mathbb{R})$ and f also lies in $L^2(\mathbb{R})$), then

$$\int_{-\infty}^{\infty} |f(t)|^2 \, \mathrm{d}t = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\hat{f}(w)|^2 \, \mathrm{d}w.$$

Thus, up to a possible constant, f and $\hat{f}$ have the same energy.

We shall now say a few words about Fourier transforms in $\mathbb{R}^n$, that is, for functions $f(\boldsymbol{x}) = f(x_1, \dots, x_n)$. The appropriate definition is

$$\hat{f}(\boldsymbol{w}) = \int_{\mathbb{R}^n} f(\boldsymbol{x}) \mathrm{e}^{-\mathrm{i}\boldsymbol{w}.\boldsymbol{x}} \, \mathrm{d}\boldsymbol{x},$$

giving another function defined on $\mathbb{R}^n$. The corresponding inversion theorem asserts that

$$f(\boldsymbol{x}) = \left(\frac{1}{2\pi}\right)^n \int_{\mathbb{R}^n} \hat{f}(\boldsymbol{w}) \mathrm{e}^{\mathrm{i}\boldsymbol{w}.\boldsymbol{x}} \, \mathrm{d}\boldsymbol{w},$$

at least if f is continuous and $\int_{\mathbb{R}^n} |f|$ and $\int_{\mathbb{R}^n} |\hat{f}|$ are both finite.

One application of the multi-dimensional Fourier transform is in the theory of partial differential equations. The partial derivative $\dfrac{\partial f}{\partial x_k}$ has transform $\mathrm{i}w_k \hat{f}(\boldsymbol{w})$, and so the Laplacian

$$\nabla^2 f = \sum_{k=1}^{n} \frac{\partial^2 f}{\partial x_k^2}$$

has transform equal to $-\|\boldsymbol{w}\|^2 \hat{f}(\boldsymbol{w})$; we shall not go into further details here.

1.3 Harmonic and analytic functions

For simplicity, let us consider 2π-periodic functions f. These correspond to functions g defined on the unit circle

$$\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$$

in the complex plane, by setting $g(\mathrm{e}^{\mathrm{i}t}) = f(t)$. Conversely, any function $g : \mathbb{T} \to \mathbb{C}$ gives a 2π-periodic function f by the same formula.

Note that the formula for the Fourier coefficients can be written

$$\hat{g}(k) = \frac{1}{2\pi}\int_0^{2\pi} g(\mathrm{e}^{\mathrm{i}t})\mathrm{e}^{-\mathrm{i}kt}\,\mathrm{d}t = \frac{1}{2\pi\mathrm{i}}\int_{\mathbb{T}} \frac{g(z)}{z^{k+1}}\,\mathrm{d}z, \tag{3}$$

where the last integral is a contour integral round the unit circle.

Suppose (for simplicity) that g is continuous. Then it has a harmonic extension to the unit disc

$$\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\},$$

namely

$$g(r\mathrm{e}^{\mathrm{i}\theta}) = \sum_{k=-\infty}^{\infty} \hat{g}(k) r^{|k|} \mathrm{e}^{\mathrm{i}k\theta},$$

for $0 \le r < 1$ and $0 \le \theta \le 2\pi$.

Write $z = x + \mathrm{i}y = r\mathrm{e}^{\mathrm{i}\theta}$ as usual. Then the extension of g is a solution to the *Dirichlet problem*, i.e., it satisfies *Laplace's equation*

$$\frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2} = 0,$$

with the boundary values of g specified on the unit circle.

One important special case arises if $\hat{g}(k) = 0$ for all $k < 0$; then the harmonic extension is

$$g(r\mathrm{e}^{\mathrm{i}\theta}) = \sum_{k=0}^{\infty} \hat{g}(k) r^k \mathrm{e}^{\mathrm{i}k\theta},$$

or

$$g(z) = \sum_{k=0}^{\infty} \hat{g}(k) z^k,$$

where again $z = r\mathrm{e}^{\mathrm{i}\theta}$. This is an analytic function (not just harmonic).

There is a one–one correspondence between power series with square-summable Taylor coefficients (the Hardy class H^2), and square-integrable functions g on the unit circle with $\hat{g}(k) = 0$ for all $k < 0$.

Suppose now that g has an analytic extension to an annulus containing the unit circle, say, $\mathcal{A} = \{A < |z| < B\}$ with $0 < A < 1 < B$. Then the formula (3) can be replaced by integrals round circles of radius a or b for any $A < a < 1 < b < B$, and we obtain useful estimates for the rate of decrease of the Fourier coefficients, namely,

$$\begin{aligned} |\hat{g}(k)| &\le M_b b^{-k}, \qquad \text{and} \\ |\hat{g}(-k)| &\le M_a a^k, \end{aligned} \tag{4}$$

for $k \geq 0$, where M_r denotes the maximum value of $|g|$ on the circle of radius r.

If now g has an isolated simple pole at a point z_0 with $A < |z_0| < 1$, with residue c, but is otherwise analytic in the annulus $\mathcal{A}$, then the identity

$$\frac{c}{z - z_0} = c \sum_{k=1}^{\infty} \frac{z_0^{k-1}}{z^k},$$

valid on $|z| = 1$, shows that $\hat{g}(-k)$ is asymptotic to cz_0^{k-1} as $k \to \infty$. Likewise, if the location of the pole satisfies $1 < |z_0| < B$ instead, then the identity

$$\frac{c}{z - z_0} = -c \sum_{k=0}^{\infty} \frac{z^k}{z_0^{k+1}}$$

shows that $\hat{g}(k)$ is asymptotic to $-c/z_0^{k+1}$ as $k \to \infty$. The extension to finitely many poles, and to poles of multiplicity greater than 1, is similar. Thus the singularities of g are reflected in the behaviour of its Fourier coefficients, a phenomenon that we shall see again in Section 3.

2 DFT, FFT, windows

We consider again the following formula for Fourier coefficients:

$$\hat{g}(k) = \frac{1}{2\pi} \int_0^{2\pi} g(\mathrm{e}^{\mathrm{i}t})\mathrm{e}^{-\mathrm{i}kt}\,\mathrm{d}t = \frac{1}{2\pi \mathrm{i}} \int_{\mathbb{T}} \frac{g(z)}{z^{k+1}}\,\mathrm{d}z.$$

In order to compute Fourier transforms numerically from data, a natural approximation to the above integral is obtained by discretising. Let us take N equally-spaced points: to do this set $\omega = \mathrm{e}^{2\pi\mathrm{i}/N}$ and consider the expression

$$\tilde{g}_N(k) = \frac{1}{N} \sum_{j=0}^{N-1} g(\omega^j)\omega^{-jk}.$$

This is a *discrete Fourier transform* of g. Since $\omega^N = 1$, the values of $\tilde{g}_N$ repeat themselves, and we need only work with $\tilde{g}_N(-\frac{N}{2}), \ldots, \tilde{g}_N(\frac{N}{2} - 1)$. It is not difficult to convince oneself that, if g is continuous, then for each fixed k the number $\tilde{g}_N(k)$ should be close to $\hat{g}(k)$ when N is sufficiently large (basically, we have replaced a Riemann integral by a Riemann sum). An approximation to the Fourier series for g is now given by taking the function

$$g_N(\mathrm{e}^{\mathrm{i}t}) = \sum_{k=-N/2}^{N/2-1} \tilde{g}_N(k)\mathrm{e}^{\mathrm{i}kt},$$

The *Fast Fourier Transform (FFT)* was introduced by Cooley and Tukey as a numerical algorithm for computing the discrete Fourier coefficients of g

for values of N which are powers of 2, say $N = 2^n$. At first sight it seems that, starting with N values of g, we require approximately $2N^2$ operations (additions and multiplications) to calculate the N values of $\tilde{g}_N$. In fact, if we have an even number of points, say $2r$, and divide them into two halves (the even ones and the odd ones), then we can exploit the algebraic relations existing between $\tilde{g}_r$ and $\tilde{g}_{2r}$. These imply that, if we can find the coefficients g_r in M operations, then we can obtain the coefficients g_{2r} in not more than $2M + 8r$ operations.

The upshot is that, for $N = 2^n$, computers can calculate the coefficients $\tilde{g}_N$ in at most $n2^{n+2} = 4N \log_2 N$ operations. This is a significant saving if N is of the order of several thousand.

In many applications, it is convenient to work with a *windowed discrete Fourier transform* of g, which is a function of the form

$$g_w(\mathrm{e}^{\mathrm{i}t}) = \sum_{k=-\infty}^{\infty} \tilde{g}_N(k) w_k \mathrm{e}^{\mathrm{i}kt},$$

where (w_k) is a sequence of weights, of which usually only finitely many are non-zero. For example, for $0 \le m < N$ we may take the sequence

$$w_k = \begin{cases} \dfrac{m+1-|k|}{m+1} & \text{for} |k| \le m, \\ 0 & \text{otherwise,} \end{cases}$$

in which case the corresponding functions g_w form a sequence of trigonometric polynomials known as the *Jackson polynomials*, $J_{m,N}(g)$. These have many attractive properties, in particular they converge uniformly to the original function g as $N \to \infty$, for any sequence of $m = m(N)$ remaining less than N but also tending to infinity. They are also *robust*, in the sense that small measurement errors or perturbations lead to small errors in the polynomials. For rather more rapid convergence, one may use the discrete *de la Vallée Poussin polynomials*, $V_{m,N}(g)$, defined for $N \ge 3m$ using the following window:

$$w_k = \begin{cases} 1 & \text{for } |k| \le m, \\ \dfrac{2m-|k|}{m} & \text{for } m \le |k| \le 2m, \\ 0 & \text{otherwise.} \end{cases}$$

These have been used in various interpolation and approximation schemes, for example in the identification of linear systems from noisy frequency-domain data.

3 The behaviour of f and $\hat{f}$

We return to Fourier transforms for functions defined on $L^1(\mathbb{R})$, and consider how the properties of f and $\hat{f}$ are linked. For example, it is easily seen that, if f is a real function, then $\hat{f}(-w) = \overline{f(w)}$; if f is a real even function, then $\hat{f}$ is purely real, and if f is a real odd function, then $\hat{f}$ is purely imaginary.

The properties of f and $\hat{f}$ behave well under translations and dilations: let

$$(T_a f)(t) = f(t-a) \qquad \text{and} \qquad (D_b f)(t) = f(t/b)$$

for $a \in \mathbb{R}$ and $b > 0$. Then

$$(T_a f)\hat{}(w) = \mathrm{e}^{-\mathrm{i}aw}\hat{f}(w) \qquad \text{and} \qquad (D_b f)\hat{}(w) = b\hat{f}(bw).$$

Similarly, derivatives transform in a simple fashion: if f is an $L^1(\mathbb{R})$ function with a continuous derivative, such that $\int_{\mathbb{R}} |f'| < \infty$, then

$$(f')\hat{}(w) = \mathrm{i}w\hat{f}(w).$$

In particular, there is a constant $C > 0$ such that $|\hat{f}(w)| \le C/|w|$. This argument can be repeated with higher derivatives, and we obtain the slogan: *the smoother the function, the faster its Fourier transform decays.* A similar phenomenon holds for Fourier series of periodic functions: for smooth functions the Fourier coefficients tend rapidly to zero.

By means of the inversion theorem, we can argue in the other direction too: if $\hat{f}$ is smooth, then this corresponds to rapid decay of f at ∞.

In many applications, it is convenient to work with smooth functions of rapid decay. Thus we define the *Schwartz class*, $\mathcal{S}$, to be the class of all infinitely differentiable functions $f : \mathbb{R} \to \mathbb{C}$ such that every derivative is rapidly decreasing: thus, for all n, k, there is $C_{n,k} > 0$ such that

$$|f^{(n)}(t)| \le \frac{C_{n,k}}{(1+|t|)^k}$$

for all $t \in \mathbb{R}$. A simple example is $\exp(-at^2)$ with $a > 0$, but one can even find such functions with compact support (so-called "bump functions").

Now if $\int_{\mathbb{R}} |t^k f(t)|\,\mathrm{d}t < \infty$, it follows that $\hat{f}$ is differentiable k times, and

$$(\hat{f})^{(k)}(w) = \int_{-\infty}^{\infty} (-\mathrm{i}t)^k f(t)\mathrm{e}^{-\mathrm{i}tw}\,\mathrm{d}t.$$

This can be used to show that the Fourier transform is a linear bijection from $\mathcal{S}$ onto itself.

Suppose now that f is smooth apart from jump discontinuities of the function and its derivatives at the origin, so that we may define

$$\delta_k = \lim_{t\to 0+} f^{(k)}(t) - \lim_{t\to 0-} f^{(k)}(t)$$

for $k = 0, 1, 2, \ldots$. Then it can be shown that $\hat{f}$ possesses an asymptotic expansion of the form

$$\hat{f}(w) \sim \sum_{k=0}^{\infty} \frac{\delta_k}{(\mathrm{i}w)^{k+1}}$$

as $|w| \to \infty$.

Moreover, since the Fourier transform and inverse Fourier transform are related by (1), we may similarly conclude that jumps ε_k in $\hat{f}$ and its derivatives at the origin are reflected in an asymptotic expansion

$$f(t) \sim \frac{1}{2\pi} \sum_{k=0}^{\infty} \frac{\varepsilon_k}{(-\mathrm{i}t)^{k+1}}$$

as $|t| \to \infty$.

Finally, the expansions corresponding to jumps occurring at other points on the real line may be derived by a straightforward change of variables.

We now consider the case when f has an analytic extension to a horizontal band $\mathcal{B} = \{A < \operatorname{Im} z < B\}$, where $A < 0$ and $B > 0$. Then certain estimates hold for the Fourier transform, which are analogous to those obtained for Fourier coefficients in (4). If we take $0 < b < B$ and suppose that f is absolutely integrable on the line $\{\operatorname{Im} z = b\}$, tending to zero uniformly in $\mathcal{B}$ as $\operatorname{Re} z \to \pm\infty$, then we can move the contour of integration, and obtain the estimate

$$|\hat{f}(-w)| \le \int_{-\infty}^{\infty} |f(x + \mathrm{i}b)|\, |\mathrm{e}^{\mathrm{i}w(x+\mathrm{i}b)}|\, \mathrm{d}x = O(\mathrm{e}^{-bw})$$

as $w \to \infty$. Similarly, analyticity in the lower half-plane leads to estimates of the form $|\hat{f}(w)| = O(\mathrm{e}^{aw})$ as $w \to \infty$, provided that we may integrate along the line $\{\operatorname{Im} z = a\}$ with $A < a < 0$.

Once more we may see the existence of singularities of f reflected in the asymptotic behaviour of $\hat{f}$. The Fourier transform of the function $f(t) = c/(t - z_0)$, with $\operatorname{Re} z_0 > 0$, is easily calculated by contour integration, and is given by

$$\hat{f}(w) = \begin{cases} 2\pi\mathrm{i}c\mathrm{e}^{-\mathrm{i}wz_0} & \text{if } w < 0, \\ 0 & \text{if } w > 0. \end{cases}$$

Similarly, if $\operatorname{Re} z_0 < 0$, the Fourier transform is

$$\hat{f}(w) = \begin{cases} 0 & \text{if } w < 0, \\ -2\pi\mathrm{i}c\mathrm{e}^{-\mathrm{i}wz_0} & \text{if } w > 0. \end{cases}$$

Thus, if f is sufficiently regular in $\mathcal{B}$ except for an isolated pole with residue c occurring at $p + \mathrm{i}q$ with $q > 0$, then there is an asymptotic formula valid for $w \to -\infty$, namely

$$\hat{f}(w) \sim 2\pi\mathrm{i}c\mathrm{e}^{-\mathrm{i}pw}\mathrm{e}^{qw}.$$

The extensions to poles in the lower half-plane, to a finite number of poles, and to poles of multiple order, are very similar and we omit them. As before, we may exchange the roles of f and $\hat{f}$, using the identity (1), so that singularities in $\hat{f}$ are reflected in the asymptotic behaviour of f.

4 Wiener's theorems

Suppose that a 2π-periodic function f has an absolutely convergent Fourier series, that is

$$f(t) = \sum_{k=-\infty}^{\infty} \hat{f}(k)\mathrm{e}^{\mathrm{i}kt},$$

with $\sum_{k=-\infty}^{\infty} |\hat{f}(k)| < \infty$. So in fact f is necessarily continuous (although this is not a sufficient condition), but need not be differentiable. These functions form a linear space, and indeed an algebra, closed under multiplication, since if

$$f(t) = \sum_{k=-\infty}^{\infty} \hat{f}(k)\mathrm{e}^{\mathrm{i}kt} \qquad \text{and} \quad g(t) = \sum_{k=-\infty}^{\infty} \hat{g}(k)\mathrm{e}^{\mathrm{i}kt},$$

then

$$f(t)g(t) = \sum_{k=-\infty}^{\infty} c_k \mathrm{e}^{\mathrm{i}kt},$$

where

$$c_k = \sum_{j=-\infty}^{\infty} \hat{f}(j)\hat{g}(k-j),$$

and so

$$\begin{aligned}\sum_{k=-\infty}^{\infty} |c_k| &\le \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} |\hat{f}(j)|\,|\hat{g}(k-j)| \\ &= \sum_{j=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} |\hat{f}(j)|\,|\hat{g}(l)| < \infty,\end{aligned}$$

i.e., $f.g$ has an absolutely convergent Fourier series.

It is a much deeper result, due to Wiener, that, if f never takes the value 0, then $1/f$ has an absolutely convergent Fourier series. Originally proved by "hard" analysis, it can now be deduced more easily using the Gelfand theory of commutative Banach algebras.

There is an analogous result for Taylor series (linked by the change of variable $z = \mathrm{e}^{it}$): suppose that

$$f(z) = \sum_{k=0}^{\infty} a_k z^k$$

has an absolutely convergent Taylor series, so that $\sum_{k=0}^{\infty} |a_k| < \infty$. Such functions are analytic in the open unit disc $\mathbb{D}$ and continuous on the closed disc $\overline{\mathbb{D}}$. If $f(z) \neq 0$ for $z \in \overline{\mathbb{D}}$, then the function $1/f$ also has an absolutely convergent Taylor series.

A more general result is the *Wiener–Lévy theorem*: if G is a function holomorphic in a neighbourhood of the range of f, then the composite function $G \circ f$ also has an absolutely convergent Fourier series. The special case $G(x) = 1/x$ is the classical Wiener theorem, but one can consider other functions such as $\sqrt{f}$ if $|f(z) - 1| < 1$ for all z, and these too have absolutely convergent Fourier series.

There are analogous results for Fourier transforms. We remark first that if f and g lie in $L^1(\mathbb{R})$, then their *convolution* $f * g$, given by

$$(f * g)(x) = \int_{-\infty}^{\infty} f(x - y) g(y) \, \mathrm{d}y,$$

also lies in $L^1(\mathbb{R})$. Indeed $f * g = g * f$, and we also have

$$\|f * g\|_1 \leq \|f\|_1 \|g\|_1 \qquad \text{and} \quad (f * g)\hat{}\,(w) = \hat{f}(w)\hat{g}(w).$$

The main consequence of the non-vanishing of the Fourier transform of f is *Wiener's Tauberian theorem.* This may be presented in three forms.

(i) If $f \in L^1(\mathbb{R})$, then the translates of f, namely $f_\lambda(x) = f(x - \lambda)$ for $\lambda \in \mathbb{R}$, span a dense subspace of $L^1(\mathbb{R})$ if and only if $\hat{f}$ is non-zero everywhere.

(ii) If $f \in L^1(\mathbb{R})$, then the convolutions $f * g$ for functions g in $L^1(\mathbb{R})$ form a dense subspace of $L^1(\mathbb{R})$ if and only if $\hat{f}$ is non-zero everywhere.

(iii) If $f \in L^1(\mathbb{R})$ and $\hat{f}$ is non-zero everywhere, and if in addition the identity

$$\lim_{x \to \infty} (f * K)(x) = A \int_{-\infty}^{\infty} f(x) \, \mathrm{d}x,$$

holds for a given function $K \in L^\infty(\mathbb{R})$ and $A \in \mathbb{C}$, then in fact

$$\lim_{x \to \infty} (g * K)(x) = A \int_{-\infty}^{\infty} g(x) \, \mathrm{d}x.$$

holds for every $g \in L^1(\mathbb{R})$.

The first two forms of the theorem may be seen as results in approximation theory; the last one, Wiener's original version of the theorem, is a "Tauberian" theorem (a name given to a certain kind of theorem that deduces the convergence of a series or integral from other hypotheses).

There is also an L^2 version of (i), which is useful in some applications.

(iv) If $f \in L^2(\mathbb{R})$, then the translates of f span a dense subspace of $L^2(\mathbb{R})$ if and only if $\hat{f}$ is non-zero almost everywhere.

5 Laplace and Mellin transforms

5.1 Laplace

The Laplace transform is an important tool in the theory of differential equations, and we give its basic properties. Let f be a measurable function defined on $(0, \infty)$. Then we define its *Laplace transform* $F = \mathcal{L}f$ by

$$F(s) = \int_0^\infty f(t)\mathrm{e}^{-st}\,\mathrm{d}t,$$

which will, in general be a holomorphic function of a complex variable lying in some half-plane $\operatorname{Re} s > a$. For example, if f is an exponential function $f(t) = \mathrm{e}^{\lambda t}$, then $F(s) = 1/(s - \lambda)$, and the integral converges for $\operatorname{Re} s > \operatorname{Re} \lambda$.

There is an inversion formula available. Namely, if $b > a$, we have

$$f(t) = \lim_{y \to \infty} \frac{1}{2\pi\mathrm{i}} \int_{b-\mathrm{i}y}^{b+\mathrm{i}y} F(s)\mathrm{e}^{st}\,\mathrm{d}s,$$

which is an integral along a vertical contour in the complex plane.

Suppose f is differentiable, and we take the Laplace transform of f'. We may integrate by parts to obtain:

$$\begin{aligned}
(\mathcal{L}f')(s) &= \int_0^\infty \mathrm{e}^{-st} f'(t)\,\mathrm{d}t \\
&= [\mathrm{e}^{-st} f(t)]_{t=0}^\infty + s\int_0^\infty \mathrm{e}^{-st} f(t)\,\mathrm{d}t \\
&= -f(0) + s(\mathcal{L}f)(s).
\end{aligned}$$

Thus a differential equation can be turned into an algebraic equation, using Laplace transforms.

For example, suppose that we have a "linear system"

$$y''(t) + ay'(t) + by(t) = cu'(t) + du(t),$$

where a, b, c and d are real.

Here u is the input, and y the output. We suppose also (for simplicity) that $u(0) = y(0) = 0$. Then, writing $U = \mathcal{L}u$ and $Y = \mathcal{L}y$, we arrive at

$$(s^2 + as + b)Y(s) = (cs + d)U(s),$$

and we have an algebraic relation between U and Y.

Similarly, we may shift/translate/delay a function f by an amount $T > 0$, to get g defined by

$$g(t) = \begin{cases} f(t-T) & \text{if } t \geq T, \\ 0 & \text{if } t < T. \end{cases}$$

Then

$$\begin{aligned} (\mathcal{L}g)(s) &= \int_0^\infty \mathrm{e}^{-st} g(t)\,\mathrm{d}t \\ &= \int_T^\infty \mathrm{e}^{-st} f(t-T)\,\mathrm{d}t \\ &= \int_0^\infty \mathrm{e}^{-s(x+T)} f(x)\,\mathrm{d}x = \mathrm{e}^{-sT}(\mathcal{L}f)(s). \end{aligned}$$

Thus a differential–delay equation also looks simpler in the "frequency domain".

For example, suppose we have (again with zero initial conditions for simplicity) the equation

$$y'(t) + ay(t-1) = u(t).$$

Taking Laplace transforms $Y = \mathcal{L}y$ and $U = \mathcal{L}u$ gives

$$(s + a\mathrm{e}^{-s})Y(s) = U(s).$$

Thus, if we know u, we can find y by taking Laplace transforms and inverse Laplace transforms.

A key theorem due to Paley and Wiener says that the Laplace transform provides a linear mapping from the Lebesgue space $L^2(0,\infty)$ onto the Hardy class $H^2(\mathbb{C}_+)$ of the right half-plane $\mathbb{C}_+$. This consists of all analytic functions $F : \mathbb{C}_+ \to \mathbb{C}$ such that

$$\|F\|_2 := \left(\sup_{x>0} \int_{-\infty}^\infty |F(x+\mathrm{i}y)|^2 \mathrm{d}y \right)^{1/2} < \infty,$$

(roughly speaking, functions analytic in the right half-plane, with L^2 boundary values), and moreover the Laplace transform is an isomorphism in the sense that

$$\|F\|_2 = \sqrt{2\pi}\|f\|_2.$$

This is the basis of various approaches to control theory and approximation theory. One consequence is that if we have an input/output relation

$$Y(s) = G(s)U(s)$$

as in our examples, then we can decide whether L^2 inputs (finite energy) guarantee L^2 outputs. The answer is that it is necessary and sufficient that

$G(s)$ be analytic and bounded in $\mathbb{C}_+$ (i.e., lie in the Hardy class $H^\infty(\mathbb{C}_+)$). For example, if

$$y'(t) + ay(t-1) = u(t),$$

then we have this form of stability precisely when

$$\frac{1}{s + a\mathrm{e}^{-s}} \in H^\infty(\mathbb{C}_+), \qquad \text{i.e., for } 0 < a < \pi/2.$$

5.2 Mellin

Note that the Laplace transform is closely related to the Fourier transform (put $s = \mathrm{i}w$), if we consider only functions which are 0 on the negative real axis. A still closer analogue is the *bilateral Laplace transform*, where we define

$$G(s) = \int_{-\infty}^{\infty} f(t)\mathrm{e}^{-st}\,\mathrm{d}t = \hat{f}(-\mathrm{i}s),$$

as this is simply the Fourier transform with a (sometimes useful) change of variable.

A more complicated change of variable gets us to the *Mellin transform.* For a function f defined on $(0,\infty)$, we set

$$F(s) = \int_0^\infty x^{s-1} f(x)\,\mathrm{d}x,$$

so that F is the Mellin transform of f. The variable s will in general be complex, and then the function F is holomorphic in some strip.

For example, if we take $f(x) = \mathrm{e}^{-x}$, then

$$F(s) = \int_0^\infty x^{s-1}\mathrm{e}^{-x}\,\mathrm{d}x = \Gamma(s),$$

which is in fact analytic in $\mathbb{C}_+$.

If we set $x = \mathrm{e}^{-t}$, so that $t \in \mathbb{R}$, we obtain, at least formally,

$$F(s) = \int_{-\infty}^{\infty} f(\mathrm{e}^{-t})\mathrm{e}^{-st}\,\mathrm{d}t,$$

which expresses the Mellin transform as a Fourier transform (or a bilateral Laplace transform). The Mellin inversion formula asserts that for suitable functions f such that $x^{a-1}f(x)$ is integrable, we have

$$f(t) = \frac{1}{2\pi\mathrm{i}} \lim_{y\to\infty} \int_{a-\mathrm{i}y}^{a+\mathrm{i}y} F(s)x^{-s}\,\mathrm{d}s,$$

which is again an integral along a vertical contour in the complex plane.

We conclude with a further application. Let us consider Laplace's equation in a sector $\{(r,\theta) : r > 0 \text{ and } a < \theta < b\}$. In polar coordinates we have

$$r^2 u_{rr} + r u_r + u_{\theta\theta} = 0,$$

with some appropriate boundary conditions. Let us take a Mellin transform in r, i.e.,

$$U(s,\theta) = \int_0^\infty r^{s-1} u(r,\theta)\,\mathrm{d}r.$$

Then it is easily checked that we now have

$$U_{\theta\theta} + s^2 U = 0,$$

for $0 < \operatorname{Re} s < \mu$, if $u(r,\theta) = O(r^{-\mu})$ at 0.

Suppose, to make life simple, we take $0 < \theta < 1$ and

$$u(r,0) = 0, \qquad u(r,1) = \begin{cases} 1 & \text{for } 0 \le r \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

(This might represent the heat flow in a piece of cake, heated on one side only.) Then it is easily verified that

$$U(s,\theta) = \frac{1}{s}\frac{\sin s\theta}{\sin s},$$

and we can find u by inverting the Mellin transform.

$$u(r,\theta) = \frac{1}{2\pi \mathrm{i}} \int_{a-\mathrm{i}\infty}^{a+\mathrm{i}\infty} \frac{r^{-s}}{s} \frac{\sin s\theta}{\sin s}\,\mathrm{d}s,$$

where $0 < a < \pi$.

Curiously, the integral can be done using Cauchy's residue theorem. In that case our story comes full circle, as we obtain a Fourier series solution

$$u(r,\theta) = \frac{1}{\pi} \sum_{n=1}^{\infty} \frac{(-1)^n r^{-n\pi}}{n} \sin n\pi\theta.$$

References

1. Y. Katznelson, *An introduction to harmonic analysis.* Dover Publications, Inc., New York, 1976.
2. T.W. Körner, *Fourier analysis.* Cambridge University Press, Cambridge, 1988.
3. M. Levitin, Fourier Tauberian theorems. Appendix in Yu. Safarov and D. Vassiliev, *The asymptotic distribution of eigenvalues of partial differential operators.* AMS Series Translations of Mathematical Monographs 155, AMS, Providence, R. I., 1997.
4. J.R. Partington, *Interpolation, identification, and sampling.* The Clarendon Press, Oxford University Press, 1997.
5. N. Wiener, *The Fourier integral and certain of its applications.* Reprint of the 1933 edition. Cambridge University Press, Cambridge, 1988.
6. W.E. Williams, *Partial differential equations.* The Clarendon Press, Oxford University Press, 1980.
7. A. Zygmund, *Trigonometric series.* Vol. I, II. Third edition. Cambridge University Press, Cambridge, 2002.

Padé Approximants

Maciej Pindor

Instytut Fizyki Teoretycznej,
Uniwersytet Warszawski ul.Hoża 69,
00-681 Warszawa, Poland.

1 Introduction

The frequent situation one encounters in applied science is the following: the information we need is contained in values, or some features of the analytical structure, of some function of which we have a knowledge only in the form of its power expansion in a vicinity of some point. Favourably, it is the Taylor expansion with some finite radius of convergence, but it may also be an asymptotic expansion. Let us concentrate on the first case, some remarks concerning the second one will be given later, if time allows.

If the information we need concerns points within the circle of convergence of the Taylor series, then the problem is (almost) trivial. If it concerns points outside the circle, then the problem becomes that of analytic continuation. Unfortunately, the method of direct rearrangements of the series, used in theoretical considerations on the analytic continuation, is practically useless here. The method of the "practical analytic continuation" which I shall discuss is called the method of "Padé Approximation". There exist ample monographs on Padé Approximants [6], [2], [3] and my purpose here is to present you a subjective glimpse of the subject.

Actually, the method is based on the very direct idea of using rational functions instead of polynomials to approximate the function of interest. They are practically as easy to calculate as polynomials, but when we recall that the truncated Laurent expansions is just a rational function, we can expect that they could provide reasonable approximations of functions also in a vicinity of the poles of the latter, not only in circles of analyticity. Therefore the concept, born already in XIXth century, was to substitute partial sums of the Taylor series, by rational functions having the corresponding partial sums of their own Taylor series identical to that former one. To formulate it precisely, let us assume we have a function $f(z)$ with its Taylor expansion

$$f(z) = \sum_{i=0}^{\infty} f_i z^i \text{ for } |z| < R \; . \tag{1}$$

J.-D. Fournier et al. (Eds.): Harm. Analysis and Ratio. Approx., LNCIS 327, pp. 59–69, 2006.

Having the partial sum of the above series up to the power M, we seek a rational function $r_{m,n}(z)$ which will have first $M+1$ terms of its Taylor expansion identical to that of $f(z)$ what I shall represent by

$$r_{m,n}(z) - f(z) = O(z^{M+1}) \; . \tag{2}$$

Unfortunately this problem seems to be badly defined – there are probably many rational functions that can satisfy this condition: possibly all such that $m+n=M$. In other words, assuming for the moment that all such rational functions can be found, to the infinite *sequence* of partial sums of the Taylor series (1) there correspond an infinite *table* (or a double sequence) of *rational approximants* defined by (2). As we are after the analytic continuation of $f(z)$, we expect that some sequence of rational approximants defined this way would converge, in some sense, to $f(z)$ outside the convergence circle of (1). But which one? Is it a case of advantageous flexibility, or that of "embarras du choix"? I shall argue in a moment that it is this first one!

2 The Padé Table

Let us, however, discuss first the problem of existence of rational functions defined by (2). If we denote the numerator of $r_{m,n}(z)$ by $P_m(z)$ and its denominator by $Q_n(z)$ and $r_{m,n}(z)$ by $[m/n]_f(z)$ then (2) becomes

$$[m/n]_f(z) - f(z) = \frac{P_m(z)}{Q_n(z)} - f(z) = O(z^{m+n+1}) \; . \tag{3}$$

Let me make here an obvious remark that P_m depends also on n and Q_n depends on m and they should be denoted, e.g., $P_m^{[m/n]}$, but for hygienic reasons I shall almost everywhere skip this additional index. Finding coefficients of P_m and Q_n by the expansion of $[m/n]_f(z)$ and then comparing the two series, would be a horror, but the problem can immediately be reduced to the linear one:

$$P_m(z) - Q_n(z) f(z) = O(z^{m+n+1}) \; . \tag{4}$$

This is how Frobenius [5] defined "Näruhngsbrüchen" already in 1881 and therefore (4) is called the Frobenius definition. One can immediately see that it leads to a system of linear equations for coefficients of $Q_n(z)$ and formulae expressing coefficients of P_m by those of Q_n. Denoting the former by $\{p_i\}_0^m$ and the later by $\{q_i\}_0^n$ we have (assuming that $f_i \equiv 0$ for $i<0$)

$$\begin{pmatrix} f_{m+1} & f_m & \cdots & f_{m-n+2} & f_{m-n+1} \\ f_{m+2} & f_{m+1} & \cdots & f_{m-n+3} & f_{m-n+2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ f_{m+n} & f_{m+n-1} & \cdots & f_{m+1} & f_m \end{pmatrix} \begin{pmatrix} q_0 \\ q_1 \\ \cdots \\ q_n \end{pmatrix} = 0 \tag{5}$$

and

$$
\begin{aligned}
p_0 &= f_0 q_0 \\
p_1 &= f_1 q_0 + f_0 q_1 \\
p_2 &= f_2 q_0 + f_1 q_1 + f_0 q_2 \\
\cdots &\cdots \\
p_m &= \sum_{i=0}^{\min(m,n)} f_{m-i} q_i \,.
\end{aligned}
\tag{6}
$$

The system (5) is an homogeneous one but it has n equations for $n+1$ unknowns. The reason is that $[m/n]_f$ has $m+n+1$ free coefficients, but we have written equations for $m+n+2$ ones. The result is that the system (5) has always at least one nontrivial solution. One could think that we can take then an arbitrary value for one of the coefficients q_i and next solve (5) for the remaining coefficients of Q_n. However, it may happen that the determinant of this linear system vanishes and either we have again an infinite number of solutions, or no solution at all. The problem can also be stated in this way: although the rational approximant defined by (4) always exists, it may happen that it does not satisfy (3). The first study of the table of all approximants $[m/n]_f$ has been done by Henry Padé [8] in his PhD dissertation and it is why now they are called *Padé Approximants* and the table is called *Padé Table*. The result was that there were square areas of the Padé Table where all entries were identical rational functions of degrees equal to those of its upper left corner and they all fulfill (4). However, only approximants on the antidiagonal of the square and to the left (up) to it fulfill also (3). According to one of the contemporary definitions introduced by Baker [1], we take $Q_n(0) \neq 0$ (e.g. 1, i.e. $q_0 = 1$) which is possible only when (3) is satisfied, and say that only in this case "Padé Approximants exist". See Fig. 1.

I shall present here only the very brief discussion of situations leading to an appearance of blocks in the Padé table. Of course their existence is due to special relations between coefficients of the Taylor series – e.g. vanishing of some coefficients or possibility of representing higher coefficients by algebraic functions of lower ones.

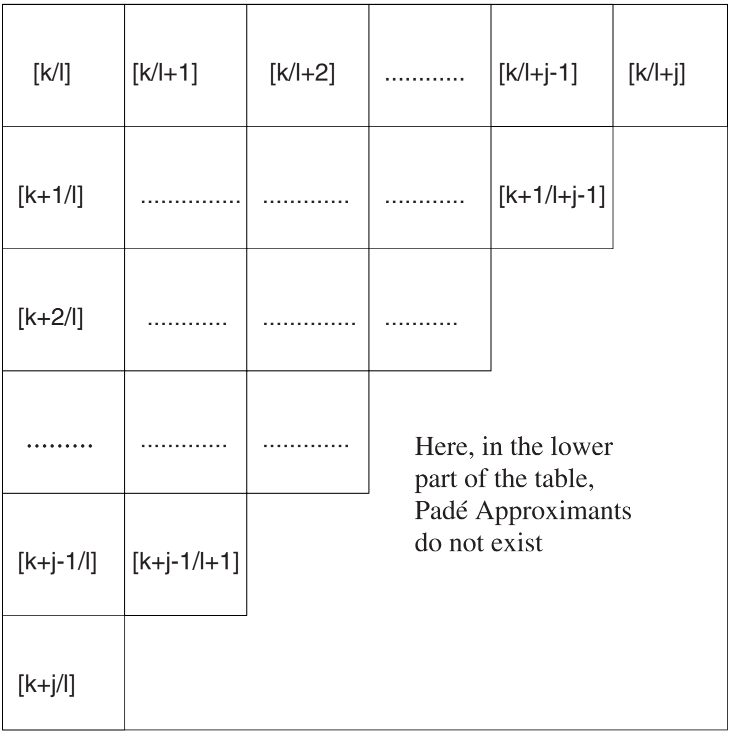
The first situation is exemplified by the series containing only even powers of the variable

$$f(z) = \frac{\log(1+z^2)}{z^2} = 1 - \frac{z^2}{2} + \frac{z^4}{3} - \frac{z^6}{4} + \frac{z^8}{5} + \cdots$$

For this series we have

$$[2/2]_f = \frac{1+\frac{z^2}{6}}{1+\frac{2z^2}{3}} = 1 - \frac{z^2}{2} + \frac{z^4}{3} - \frac{2z^6}{9} + \cdots$$

Obviously, $[2/2]_f$ is simultaneously $[3/2]_f$ and $[2/3]_f$ because its Taylor series matches that of $f(z)$ up to z^5. On the other hand, there is no rational function



[k/l]	[k/l+1]	[k/l+2]	....	[k/l+j-1]	[k/l+j]
[k+1/l]	....	....	....	[k+1/l+j-1]	
[k+2/l]	....	....	....		
....	....	....			
[k+j-1/l]	[k+j-1/l+1]				
[k+j/l]					

Here, in the lower part of the table, Padé Approximants do not exist

Fig. 1. A block of the size $j+1$ in the Padé Table. All Padé Approximants on the positions indicated by their symbols, or by dots, exist and are identical to $[k/l]$, therefore they are rational functions of degrees k and l in the numerator and the denominator, however they fulfill equation (3) with m and n corresponding to their positions in the Padé Table.

of degrees of the numerator and of the denominator both ≤ 3 that would match the series for $f(z)$ up to z^6. $[4/2]_f$ satisfies this condition, but then it is identical with $[5/3]_f$ and $[4/3]_f$

$$[4/2] = \frac{1+\frac{z^2}{4}-\frac{z^4}{24}}{1+\frac{3z^2}{4}} = 1-\frac{z^2}{2}+\frac{z^4}{3}-\frac{z^6}{4}+\frac{3z^8}{16}+\cdots$$

In this case the whole Padé Table consists of blocks of the size 2.

The second situation appears typically when $f(z)$ is a rational function itself. In this case there is one infinite block with the left upper corner at the entry corresponding to the exact degrees of the numerator and the denominator of this function. Obviously all the Padé Approximants with degrees of numerators and denominators larger or equal to these of the function, are equal to this function, because it matches it own Taylor expansion to any order!

3 Convergence

Rational functions are meromorphic, and therefore the first speculation that comes to the mind (at least mine) is that Padé approximants should be well suited to approximate just the former ones.

This speculation appears to absolutely correct, because there holds the de Montessus theorem ([3] p. 246):

Theorem 1 (de Montessus, 1902). *Let $f(z)$ be a function meromorphic in the disk $|z| < R$ with m poles at distinct points $z_1, z_2, ..., z_m$ with*

$$0 < |z_1| \leq |z_2| \leq \cdots \leq |z_m| < R \,.$$

Let the pole at z_k have multiplicity μ_k and let the total multiplicity $\sum_{k=1}^{m} \mu_k = M$ precisely. Then

$$f(z) = \lim_{L \to \infty} [L/M]$$

uniformly on any compact subset of

$$\mathcal{D} = \{z,\, |z| \leq R,\, z \neq z_k,\ k = 1, 2, \ldots, m\} \,.$$

One could be very enthusiastic about this theorem, considering that it "solves" completely the problem of analytic continuation inside a disc of meromorphy. There is however a practical obstacle in applying the theorem: generally, we cannot say what M we should use. We cannot expect anything particularly interesting if M is too small (e.g. smaller than the multiplicity of the nearest singularity), but when it is too large, the uniform convergence can be expected only for subsequences on rows in the Padé Table. This is well illustrated by [4]:

Theorem 2 (Beardon, 1968). *Let $f(z)$ be analytic in $|z| \leq R$. Then an infinite subsequence of $[L/1]$ Padé approximants converges to $f(z)$ uniformly in $|z| \leq R$.*

which casts into doubt whether the sequence $[L/1]$ must converge even in a disc of analyticity of the function! Although the theorem does not exclude that the subsequence could be the complete sequence, many counterexamples were constructed to show that the above theorem is the optimal result. Maybe the best known is the one due to Perron [9] – he has constructed the series representing an entire function, but such that poles of $[L/1]$ were dense in the plane.

On the other hand such ugly phenomena do not appear in "practice" – e.g. for $f(z) = \mathrm{e}^z$ poles of $[L/a]$ lie at $L+1$, while these of $[L/2]$ at $L+1 \pm \mathrm{i}\sqrt{L+1}$ and both rows (and also all the other ones) of the Padé table converge to $f(z)$ on any compact subset of the complex plane containing the origin.

Happily, problems caused by the stray poles are not as acute as one could think, as explained by the following theorem ([3] p. 264)

Theorem 3. *Let $f(z)$ be analytic at the origin and also in a given disk $|z| \leq R$ except for m poles counting multiplicity. Consider a row of Padé table $[L/M]$ of $f(z)$ with M fixed, $M \geq m$, and $L \to \infty$. Suppose that arbitrarily small, positive ε and δ are given. Then L_0 exists such that $|f(z) - [L/M]| < \varepsilon$ for any $L > L_0$ and for all $|z| \leq R$ except for $z \in \mathcal{E}_L$ where $\mathcal{E}_L$ is a set of points in the z-plane of measure less than δ.*

This type of convergence is known as the convergence in measure and seems to be used in this context first by Nuttal [7]. It means that we cannot guarantee convergence at any given point in the z-plane, but it assures us that the area where our Padé approximants do not approximate $f(z)$ arbitrarily well can be made as small as we wish.

It is important to understand that the theorem says nothing about where this set $\mathcal{E}_L$ is, and the practice shows that undesired poles are accompanied by undesired zeros and form so called *defects* which spoil convergence in smaller and smaller neighborhoods, but shift unpredictably from order to order.

But what about functions with more rich analytical structure – essential singularities and branch points?

The amazing (at least for me) fact is that if we are content with convergence in measure (or even stronger convergence in capacity) also such functions can be approximated by Padé approximants, if we consider sequences with growing degrees of the numerator and of the denominator. The fundamental theorem on convergence of Padé approximants for functions with essential singularities is due to Pommerenke [10]

Theorem 4 (Pommerenke, 1973). *Let $f(z)$ be a function which is analytic at the origin and analytic in the entire z-plane except for a countable number of isolated poles and essential singularities. Suppose $\varepsilon > 0$ and $\delta > 0$ are given. Then M_0 exists such that any $[L/M]$ Padé approximant of the ray sequence $(L/M = \lambda;\ \lambda \neq 0,\ \lambda \neq \infty)$ satisfies*

$$|f(z) - [L/M]_f(z)| < \varepsilon$$

for any $M \geq M_0$, on any compact set of the z-plane except for a set $\mathcal{E}_L$ of capacity less than δ.

As you see, the essential notion here is that of capacity. It is also known as Chebishev constant, or transfinite diameter. I do not have time here to define it, as it is a difficult concept concerning geometry of the complex plane. Anyway to understand practical implications of the theorem above and the ones to follow, it is sufficient to know that the capacity is a function on sets in the complex plane such that it vanishes for countable sets of points, but is different from zero on line segments, e.g. for a section of a straight line it equals to one fourth of its length. For a circle it is the same as for the disk inside the circle and equals to their radius. Actually it is proportional to the electrostatic capacity in the plane electrostatics.

If we want to approximate functions having branchpoints the first question that comes to mind is how can rational functions approximate a function in a vicinity of its branchpoint? The astonishing answer is – they can do it very well, simulating a cut as a line of coalescence of infinite number of zeros and poles! This answer may seem puzzling for you – which cut? There seem to be the enormous arbitrariness in joining branchpoints by cuts, and why should Padé approximants choose just this set of cuts and not another, or why should all Padé approximants choose the same cuts? The answer to these questions lies in the interesting fact that although "all cuts are equal", but some of them "are more equal than others".

This fact is established by the following theorem ($\hat{\mathbb{C}}$ denotes here the extended complex plane) [11]

Theorem 5 (Stahl, 1985). *Let f be given by an analytic function element in a neighborhood of infinity. There uniquely exists a compact set $\mathcal{K}_0 \subseteq \mathbb{C}$ such that*

(i) $\mathcal{D}_0 := \hat{\mathbb{C}} \backslash \mathcal{K}_0$ *is a domain in which $f(z)$ has a single-valued analytic continuation,*
(ii) $\operatorname{cap}(\mathcal{K}_0) = \inf \operatorname{cap}(\mathcal{K})$*, where the infimum extends over all compact sets $\mathcal{K} \subseteq \mathbb{C}$ satisfying (i),*
(iii) $\mathcal{K}_0 \subseteq \mathcal{K}$ *for all compact sets $\mathcal{K} \subseteq \mathbb{C}$ satisfying (i) and (ii).*

The set $\mathcal{K}_0$ is called *minimal set* (for single-valued analytical continuation of $f(z)$) and the domain $\mathcal{D}_0 \subseteq \hat{\mathbb{C}}$ – *extremal domain.*

The following theorem, due to H. Stahl [11], refers to, so called, close-to-diagonal sequences of Padé approximants. By the latter one means the sequence $[m/n]$ such that $\lim_{m+n\to\infty} m/n = 1$.

Theorem 6 (Stahl, 1985). *Let the function $f(z)$ be defined by*

$$f(z) = \sum_{j=0}^{\infty} f_j z^{-j}$$

and have all its singularities in a compact set $E \subseteq \hat{\mathbb{C}}$ of capacity zero. Then any close to diagonal sequence of Padé approximants $[m/n](z)$ to the function $f(z)$ converges in capacity to $f(z)$ in the extremal domain $\mathcal{D}_0$.

In simple words, the theorem says that close-to-diagonal sequences of Padé approximants converge "practically", for a very wide class of functions, everywhere, except on a set of "optimal" cuts. However, we must keep in the mind that it is not the uniform convergence, therefore when applying Padé approximants, we must be careful and compare few different approximants from a close-to-diagonal sequence.

4 Examples

Let us see some examples how Padé approximants work for different types of functions. In illustrations below, I shall devote more attention to demonstrating that Padé approximants "discover" correctly singularities and zeros than to approximating values of functions, though I shall not forget about the latter.

Let $f(z) = \tanh(z)/z + 1/[2(1+z)]$. This function has an infinite number of poles uniformly distributed on the imaginary axis at $z = (2k+1)\pi/2$ $k = 0, \pm 1, \pm 2, \ldots$ and the pole at $z = -1$. It has also infinite number of zeros, the ones closest to origin are: $z = -2.06727$, $-.491559 \pm 2.93395$i, $-.535753 \pm 6.17741$i, $-.545977 \pm 12.5134$i and so on. I have added the geometric series mainly to have a function with a series containing all powers of z, not the one with even powers only. A small curiosity is that there is a block in the Padé table of this function – the one consisting of [0/1], [0/2], [1/1], [1/2].

As in any circle centred at the origin there is an odd number of zeros and poles, we consider the sequence $[M/3]$. In the tables below I shall compare positions of zeros and poles of the approximants in this sequence.

P.A.	zeros	poles
[3/3]	$-\,2.06806$, -1.02990 ± 3.17939i	-1.00065, $-.002435 \pm 1.58229$i
[4/3]	-1.98353, $-.963506 \pm 2.89352$i 25.5413	$-.999348$, $-.006303 \pm 1.57462$i
[5/3]	-2.06711, $-.645195 \pm 2.98225$i $-6.10166, 8.37705$	$-.999974$, $-.000237 \pm 1.57193$
[6/3]	-2.08494, $-.632818 \pm 2.91477$i $-4.85867, 10.7332 \pm 4.29698$i	-1.00003, $-.000584 \pm 1.57118$i
[7/3]	-2.06730, $-.547353 \pm 2.93852$i -4.73415 ± 2.76689i, 5.76997 ± 3.47249i	-1.00000, $-.000025 \pm 1.57091$i
[8/3]	-2.06422, $-.540386 \pm 2.91715$i -4.13620 ± 2.49272i, 11.6809 5.90481 ± 4.63395i	$-.999999$, $-.000062 \pm 1.57084$i

We clearly see that first three poles and first three zeros of $[M/3]$ converge to corresponding zeros and poles of $f(z)$ as expected from the de Montessus theorem. We could have also checked that values of $[M/3]$ converge to values of $f(z)$ in the circle of the radius smaller than $3\pi/2$ – the distance of the next pair of poles. There appeared also "stray zeros" but they were outside this circle.

Our function has infinite number of poles, so let us see how "diagonal" Padé approximants work here.

P.A.	zeros	poles
[4/4]	$-\,2.02230,\ -.906076 \pm 3.08279\mathrm{i}$ 2.07416	$-.999772,\ -.001036 \pm 1.57569\mathrm{i}$ 2.08395
[5/5]	$-2.06727,\ -.499934 \pm 2.93370\mathrm{i}$ $-2.74928 \pm 8.40343\mathrm{i}$	$-1.00000,\ -2 \cdot 10^{-6} \pm 1.57081\mathrm{i}$ $-.003750 \pm 5.06207\mathrm{i}$
[6/6]	$-2.06716,\ -.497714 \pm 2.93312\mathrm{i}$ $2.03302,\ -2.66063 \pm 8.25624\mathrm{i}$	$-1.00000,\ -1 \cdot 10^{-6} \pm 1.57080\mathrm{i}$ $2.03304,\ -.003394 \pm 5.02527\mathrm{i}$

If we remember that [7/3] and [5/5] both use the same number of the coefficients (11), we can conclude that the diagonal Padé approximants approximate our function better than approximants with a prescribed degree of the denominator. We could say, there is a price to pay: [4/4] – using 9 coefficients like [5/3] – has an unwanted pole at 2.08395. We see however that it is accompanied by a zero at 2.07416 and can (correctly) guess that values of [4/4] deviate considerably from those of $f(x)$ only close to the pair, which is called *the defect.* The analogous *defect* appears in [6/6], but the pair is much more "tight" here and we can (correctly) guess that it spoils the approximative quality of [6/6] in even smaller area close to the defect. This is just how convergence in measure (and in capacity) manifests itself.

We can also see on Fig. 2 how the behavior of some Padé approximants, mentioned above, compares with the behavior of $f(x)$ on the interval $[-6, -1.5]$ i.e. "behind" the singularity at $x = -1$.

If you are curious what happens when $f(z)$ has a multiple pole – let me tell you that in that case Padé approximants have as many single poles as is a multiplicity of that pole and they all converge to this one when order the of the approximation increases.

Finally, let me say that I would be glad if you have read the message: diagonal Padé approximants are beautiful – do not be discouraged by their defects – others can also have defects, but none are as useful.

You should not, however, think that diagonal Padé approximants are always the best ones – there are some situations when *paradiagonal* sequences of Padé approximants, i.e. sequences $[m+k/m]$ with k constant, are optimal. It can happen if we have some information on the behavior of the function at infinity. Obviously $[m+k/m](x)$ behaves like x^k for $x \to \infty$. If our function behaves at infinity in a similar way, such sequences of Padé approximants can converge faster. This is well exemplified by a study of Padé approximants for $f(x) = \sqrt{x+1}\sqrt{2x+1} + 2/(1-x)$. It has zeros at 1.60415 and −1.39193, a pole at $x = 1$ and two branch points at $x = -1/2$ and $x = -1$. Look at zeros and poles of [4/3] and [4/4], remembering that [4/4] uses one coefficient the series more.

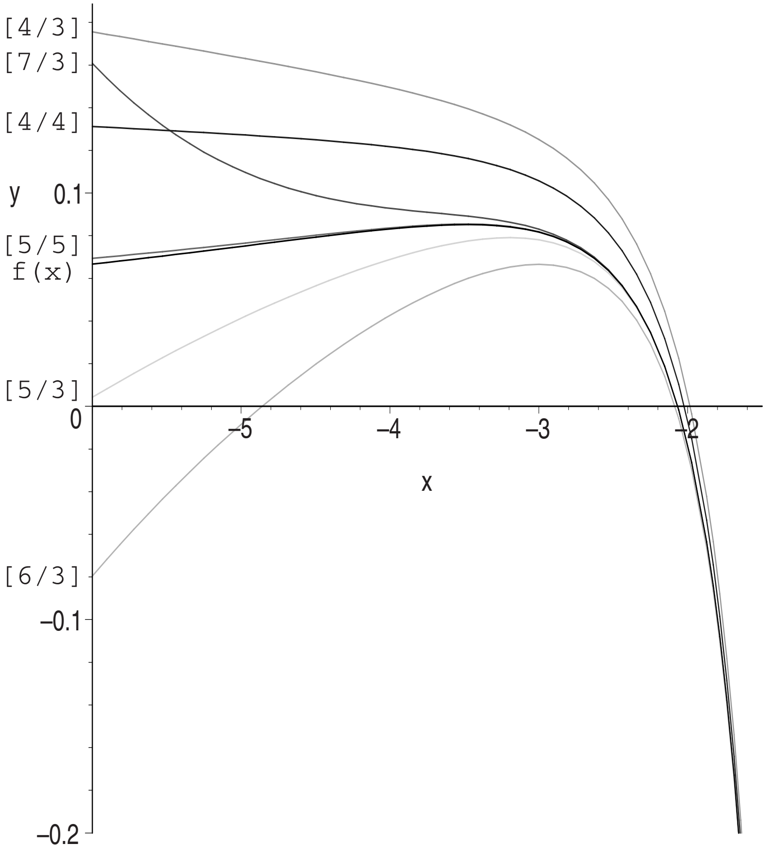


Fig. 2. Values of different PA to $f(x) = \tanh(x) + 1/(1+x)/2$

P.A.	zeros	poles
[4/3]	$-\,1.38833,\ -.754876,$ $-.556928,\ 1.60403$	$-.782628,\ -.564096$ $.999985$
[4/4]	$-1.38548,\ -.739811,$ $-.552442,\ 1.60441$	$-2499.85,\ -.767038$ $-.558807,\ 1.00002$

Positions of zeros and of the pole are clearly better reproduced by [4/3] than [4/4]. Moreover, when $x \to \infty$ [4/3](x) behaves like $1.4146x$ ($\sqrt{2} \approx 1.4142$). Additionally we see that the cut $(-1, -1/2)$ is simulated by a line of interlacing zeros and poles – the line of minimal capacity connecting branch-points.

5 Calculation of Padé approximants

In practical applications there appears a problem of how to calculate the given Padé approximants. In principle one should avoid solving a system of linear equations, because it is the process very sensitive both to errors of data and to precision of calculations. Forty and thirty years ago much activity was devoted to finding different algorithms of recursive calculation of Padé approximants. It is well documented in [3] ch. 2.4. However you can see that the system of equations for coefficients of the denominator is the one with the Toeplitz matrix and for such systems there exist relatively fast and reliable routines in all numerical programs libraries. With the speed of computers now in use, quadruple precision as a standard option in all modern Fortran compilers and also multiprecision libraries spreading around, I think that finding Padé approximants this way is in practice the most convenient solution. This is, e.g., the method used for calculation of Padé approximants in the symbolic algebra system Maple.

References

1. G.A. Baker, Jr. Existence and Convergence of Subsequences of Padé Approximants. *J. Math. Anal. Appl.*, 43:498–528, 1973.
2. G.A. Baker, Jr. *Essentials of Padé Approximants.* Academic Press, 1975.
3. G.A. Baker, Jr., P. Graves-Morris. *Padé Approximants*, volume 13 and 14 of *Encyclopedia of Mathematics and Applications.* Addison-Wesley, 1981.
4. A.F. Beardon. The convergence of Padé Approximants. *J. Math. Anal. Appl.*, 21:344–346, 1968.
5. G. Frobenius. Ueber Relationen zwischen den Näherungsbrüchen von potenzreihen. *J für Reine und Angewandte Math.*, 90:1–17, 1881.
6. J. Gilewicz. *Approximants de Padé.* Number 667 in Springer Lecture Notes in Mathematics. Springer Verlag, 1978.
7. J. Nuttal. Convergence of Padé approximants of meromorphic functions. *J. Math. Anal. Appl.*, 31:147–153, 1970.
8. H. Padé. Sur la représentation approchée d'une fonction par des fractions rationelles. *Ann. de l'Ecole Normale*, 9(3ieme série, Suppl. 3-93).
9. O. Perron. *Die Lehre von den Kettenbrüchen.* B.G. Tuebner, 1957. Chapter 4.
10. Ch. Pommerenke. Padé approximants and convergence in capacity. *J. Math. Anal. Appl.*, 31:775–780, 1973.
11. H. Stahl. Three different approaches to a proof of convergence for Padé approximants. In *Rational Approximation and its Applications in Mathematics and Physics*, number 1237 in Lecture Notes in Mathematics. Springer Verlag, 1987.

Potential Theoretic Tools in Polynomial and Rational Approximation

Eli Levin and Edward B. Saff

A.L. Levin
The Open University of Israel
Department of Mathematics
P.O. Box 808, Raanana
Israel
elile@openu.ac.il

E.B. Saff
Center for Constructive Approximation
Department of Mathematics
Vanderbilt University
Nashville, TN 37240, USA
esaff@math.vanderbilt.edu

Logarithmic potential theory is an elegant blend of real and complex analysis that has had a profound effect on many recent developments in approximation theory. Since logarithmic potentials have a direct connection with polynomial and rational functions, the tools provided by classical potential theory and its extensions to cases when an external field (or weight) is present, have resolved some long-standing problems concerning orthogonal polynomials, rates of polynomial and rational approximation, convergence behavior of Padé approximants (both classical and multi-point), to name but a few.

In this article we provide an introduction to the tools of classical and "weighted" potential theory, along with a taste of various applications. We begin by introducing three "different" quantities associated with a compact (closed and bounded) set in the plane.

1 Classical Logarithmic Potential Theory

Potential theory has its origin in the following

Problem 1 (Electrostatics Problem). Let E be a compact set in the complex plane $\mathbb{C}$. Place a unit positive charge on E so that equilibrium is reached in the sense that the energy is minimized.

To create a mathematical framework for this problem, we let $\mathcal{M}(E)$ denote the collection of all positive unit measures μ supported on E (so that $\mathcal{M}(E)$ contains all possible distributions of charges placed on E). The *logarithmic potential* associated with μ is

$$U^{\mu}(z) := \int \log \frac{1}{|z-t|} \mathrm{d}\mu(t),$$

J.-D. Fournier et al. (Eds.): Harm. Analysis and Ratio. Approx., LNCIS 327, pp. 71–94, 2006.

which is harmonic outside the support $S(\mu)$ of μ and is *superharmonic* in $\mathbb{C}$. The latter means that the value of the potential at any point z is not less than its average over any circle centered at z. Notice that, since μ is a unit measure,

$$\lim_{z\to\infty}(U^{\mu}(z)+\log|z|)=0. \tag{1}$$

The *energy* of such a potential is defined by

$$I(\mu):=\int U^{\mu}\mathrm{d}\mu=\int\int\log\frac{1}{|z-t|}\mathrm{d}\mu(t)\mathrm{d}\mu(z).$$

Thus, the electrostatics problem involves the determination of

$$V_E:=\inf\{I(\mu):\ \mu\in\mathcal{M}(E)\},$$

which is called the *Robin constant* for E. Note that since E is bounded, we have

$$\operatorname{diam}E:=\sup_{z,t\in E}|z-t|<\infty,$$

which implies that

$$-\infty<V_E\leq+\infty.$$

The *logarithmic capacity* of E, denoted by $\operatorname{cap}(E)$, is defined by

$$\operatorname{cap}(E):=\mathrm{e}^{-V_E}.$$

If $V_E=+\infty$, we set $\operatorname{cap}(E)=0$. Such sets are called *polar* and they are very "thin". In particular, the "area" (= planar Lebesgue measure) and the "length" (= one-dimensional Hausdorff measure) of any polar set, are both equal to zero. For example, any countable set has capacity zero. (However, the classical Cantor set has positive capacity.)

A fundamental theorem of Frostman asserts that if $\operatorname{cap}(E)>0$, there exists a unique measure $\mu_E\in\mathcal{M}(E)$ such that $I(\mu_E)=V_E$. This extremal measure is called the *equilibrium measure* (or *Robin measure*) for E.

We do not dwell on the proof of the Frostman result, but only mention that it utilizes three important properties:

(i) $\mathcal{M}(E)$ is compact with respect to weak-star convergence of measures;
(ii) $I(\mu)$ is a lower semi-continuous function on $\mathcal{M}(E)$;
(iii) $I(\mu)$ is a strictly convex function on $\mathcal{M}(E)$.

The existence of μ_E follows from (i), (ii), while (iii) guarantees the uniqueness. The weak-star convergence (denoted weak*) is defined as follows: we say that a sequence $\{\mu_n\}$ converges weak* to μ (write $\mu_n\overset{*}{\to}\mu$), if

$$\int f \mathrm{d}\mu_n \to \int f \mathrm{d}\mu \quad \text{as} \quad n \to \infty$$

for any function f continuous in $\mathbb{C}$.

The potential U^{μ_E} associated with μ_E is called the *equilibrium potential* (or *conductor potential*) for E. Some basic facts about $\operatorname{cap}(E)$ and U^{μ_E} are:

(a) Let $\partial_\infty E$ denote the *outer boundary* of E (that is, the boundary of the unbounded component of $\mathbb{C} \setminus E$; see Fig. 1). Then μ_E is supported on $\partial_\infty E$:

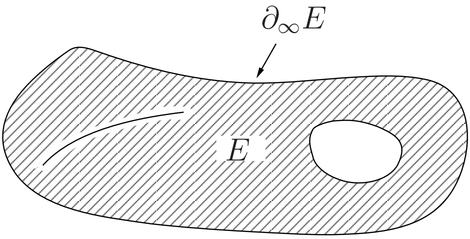


Fig. 1. Outer boundary of E

$$S(\mu_E) \subseteq \partial_\infty E.$$

Moreover, if strict inclusion takes place, then the set $\partial_\infty E \setminus S(\mu_E)$ has capacity zero. It follows from the above inclusion that, being unique, the equilibrium measures for E and for $\partial_\infty E$ coincide. Therefore

$$\operatorname{cap}(E) = \operatorname{cap}(\partial_\infty E).$$

(b) For all $z \in \mathbb{C}$,

$$U^{\mu_E}(z) \le V_E$$

with equality holding *quasi-everywhere* on E; that is, except possibly for a set of capacity zero. We write this as

$$U^{\mu_E}(z) = V_E = \log \frac{1}{\operatorname{cap}(E)} \qquad \text{q.e. on } E. \tag{2}$$

Moreover, such equality *characterizes* μ_E:
If the potential of some $\mu \in \mathcal{M}(E)$ is constant q.e. on E and $I(\mu) < \infty$, then $\mu = \mu_E$.

(c) A point $z \in E$ is called *regular* if (2) holds at z. If the interior[1] Int$E$ of E is not empty, it follows from (a) that the conductor potential is harmonic there.

[1] A point $z_0 \in \operatorname{Int} E$ if and only if there is some open disk with center at z_0 that lies entirely in E.

Then (b) guarantees that (2) holds at every point of Int$E$. The following fact is deeper: if $\partial_\infty E$ is *connected*, then every point of $\partial_\infty E$ is regular. Furthermore, at every regular point the conductor potential is *continuous*.

It is helpful to keep in mind the following two simple examples.

Example 1. Let E be the closed disk of radius R, centered at 0. Then $d\mu_E = ds/2\pi R$, where ds is the arclength on the circle $|z| = R$. One way to derive this is to observe that E is invariant under rotations. The equilibrium measure, being unique and supported on $|z| = R$, must enjoy the same property, and therefore must be of the above form. Calculating the potential, we obtain

$$U^{\mu_E}(z) = \log \frac{1}{|z|}, \quad |z| > R \qquad \text{and} \qquad U^{\mu_E}(z) = \log \frac{1}{R}, \quad |z| \leq R.$$

Therefore (see (2)), cap$(E) = R$.

Example 2. Let $E = [a, b]$ be a segment on the real line. Then cap$(E) = (b - a)/4$ and $d\mu_E$ is the arcsine measure; i.e.

$$\mathrm{d}\mu_E = \frac{1}{\pi} \frac{\mathrm{d}x}{\sqrt{(x-a)(b-x)}}, \qquad x \in [a, b].$$

If $a = -1$, $b = 1$, the conductor potential is given by

$$U^{\mu_E}(z) = \log 2 - \log |z + \sqrt{z^2 - 1}|$$

(for arbitrary a, b the expression is a bit more complicated). These results can be obtained from Example 1 by applying the Joukowski conformal map of $\mathbb{C} \setminus [-1, 1]$ onto $|w| > 1$.

There is an important relation between the equilibrium potential and the notion of *Green function*. Assume, for simplicity, that $\partial_\infty E$ is connected and let Ω denote the unbounded component of $\mathbb{C} \setminus E$ (so that $\partial\Omega = \partial_\infty E$ and $\Omega \cup \{\infty\}$ is a simply connected domain in the extended complex plane).

Let $w = \Phi(z)$ denote the conformal map of Ω onto $|w| > 1$, normalized by $\Phi(\infty) = \infty$, $\Phi'(\infty) > 0$. That is, for some constant $c > 0$,

$$\Phi(z) = \frac{1}{c} z + \text{ lower order terms}, \qquad \text{as} \qquad z \to \infty.$$

By the Riemann Mapping Theorem, such a Φ exists and is unique. Moreover, its absolute value $|\Phi|$ becomes a continuous function in the whole plane if we set

$$|\Phi(z)| = 1, \qquad z \in E.$$

Let us examine some properties of the function $g = \log |\Phi|$.

First, g is the real part of $\log \Phi(z)$ which is analytic in Ω. Therefore

(i) *g is harmonic in Ω;*
Second, our normalization implies that
(ii) $\lim_{z\to\infty}(g(z)-\log|z|)$ *exists and is finite*;
Finally,
(iii) *g is continuous in the closed domain $\overline{\Omega}$ and equals zero on its boundary.*

There is a unique function that enjoys these three properties. It is called the *Green function for Ω with pole at infinity* and is denoted by $g_\Omega(\cdot,\infty)$. So we have just shown that

$$\log|\Phi(z)| = g_\Omega(z,\infty)$$

(and that the limit in (ii) is equal to $\log(1/c)$). It is now easy to see that

$$U^{\mu_E}(z) = \log\frac{1}{\operatorname{cap}(E)} - g_\Omega(z,\infty). \tag{3}$$

Indeed, let h denote the difference of the two sides of (3). Then h is harmonic in the domain Ω, and is equal to zero on its boundary. Moreover, h has a finite limit at infinity, namely $\log(1/c) - \log(1/\operatorname{cap}(E))$, recall (1). By the maximum principle, h is identically zero and we are done. We also obtain that the constant c is just $\operatorname{cap}(E)$.

In the case when $\partial_\infty E$ is a smooth closed Jordan curve, there is a simple representation for μ_E. The equilibrium measure of any arc γ on $\partial_\infty E$ is given by

$$\mu_E(\gamma) = \frac{1}{2\pi}\int_\gamma \frac{\partial g_\Omega}{\partial n}\mathrm{d}s = \frac{1}{2\pi}\int_\gamma |\Phi'|\mathrm{d}s,$$

where the derivative in the first integral is taken in the direction of the *outer* normal on $\partial_\infty E$. Alternatively, $\mu_E(\gamma)$ is given by the normalized angular measure of the image $\Phi(\gamma)$:

$$\mu_E(\gamma) = \frac{1}{2\pi}\int_{\Phi(\gamma)}\mathrm{d}\theta \tag{4}$$

(for this representation, the smoothness of $\partial_\infty E$ is not needed).

The reader is invited to carry out the above calculations, for the special case of a disk, considered in Example 1.

We now introduce another quantity associated with E. It arises in the following

Problem 2 (Geometric Problem). Place n points on E so that they are "as far apart" as possible in the sense of the geometric mean of the distances between the points. Since the number of different pairs of n points is $n(n-1)/2$, we consider the quantity

$$\delta_n(E) := \max_{z_1,\dots,z_n\in E}\left(\prod_{1\le i<j\le n}|z_i-z_j|\right)^{2/n(n-1)}.$$

Any system of points $\mathcal{F}_n = \left\{ z_1^{(n)}, \dots, z_n^{(n)} \right\}$ for which the maximum is attained, is called an *n-point Fekete set* for E; the points $z_i^{(n)}$ in $\mathcal{F}_n$ are called *Fekete points.*

For example, if $n = 2$, then $\mathcal{F}_2 = \left\{ z_1^{(2)}, z_2^{(2)} \right\}$, where $\left| z_1^{(2)} - z_2^{(2)} \right| = \operatorname{diam} E$. Obviously, these 2 points lie on the outer boundary of E. In general, it follows from the maximum modulus principle for analytic functions, that for all n, the Fekete sets lie on the outer boundary of E.

It turns out (cf. [11], [12]), that the sequence δ_n decreases, so we may define

$$\tau(E) := \lim_{n \to \infty} \delta_n(E).$$

The quantity $\tau(E)$ is called the *transfinite diameter* of E.

Example 3. Let E be the closed unit disk. Then one can show that the set of n-th roots of unity is an n-point Fekete set for E (and so is any of its rotations). Furthermore, $\tau(E) = 1$.

Example 4. Let $E = [-1, 1]$. Then (cf. [15]) the set $\mathcal{F}_n$ turns out to be unique and it coincides with the zeros of $(1 - x^2) P_{n-2}^{(1,1)}(x)$, where $P_{n-2}^{(1,1)}$ is the Jacobi polynomial with parameters $(1, 1)$ of degree $n - 2$. Also, $\tau(E) = 1/2$.

Finally, we introduce a third quantity — the *Chebyshev constant*, $\operatorname{cheb}(E)$ — which arises in a mini-max problem.

Problem 3 (Polynomial Extremal Problem). Determine the minimal sup-norm on E for monic polynomials of degree n. That is, determine

$$t_n(E) := \min_{p \in \mathcal{P}_{n-1}} \|z^n + p(z)\|_E,$$

where $\mathcal{P}_{n-1}$ denotes the collection of all polynomials of degree $\leq n - 1$ and $\|\cdot\|_E$ is defined by

$$\|f\|_E := \max_{z \in E} |f(z)|.$$

We assume that E contains infinitely many points (which is always the case if $\operatorname{cap}(E) > 0$). Then for every n there is a unique monic polynomial $T_n(z) = z^n + \cdots$ such that $\|T_n\|_E = t_n(E)$. It is called the n-th *Chebyshev polynomial* for E.

In view of the simple inequality

$$t_{m+n}(E) = \|T_{m+n}\|_E \leq \|T_m T_n\|_E \leq \|T_m\|_E \|T_n\|_E = t_m(E) t_n(E),$$

one can show (cf. [11], [12]) that the sequence $t_n(E)^{1/n}$ converges, so we may define

$$\operatorname{cheb}(E) := \lim_{n \to \infty} t_n(E)^{1/n}.$$

Example 5. Let E be the closed disk of radius R, centered at 0. For any $p \in \mathcal{P}_{n-1}$, the ratio $(z^n + p(z))/z^n$ represents an analytic function in $|z| \geq 1$ that takes the value 1 at ∞. By the maximum principle,

$$\|z^n + p(z)\|_E = \max_{|z|=R} |z^n + p(z)| = R^n \max_{|z|=R} \left| \frac{z^n + p(z)}{z^n} \right| \geq R^n,$$

and strict inequality takes place if $p(z)$ is not identically zero. It follows that $T_n(z) = z^n$. Therefore $t_n(E) = R^n$ and $\mathrm{cheb}(E) = R$.

Example 6. Let $E = [-1, 1]$. Then T_n is the classical monic Chebyshev polynomial

$$T_n(x) = 2^{1-n} \cos(n \arccos x), \quad x \in [-1, 1], \quad n \geq 1.$$

Also, $t_n(E) = 2^{1-n}$ from which it follows that $\mathrm{cheb}(E) = 1/2$.

Closely related to Chebyshev polynomials are *Fekete polynomials*. An n-th Fekete polynomial $F_n(z)$ is a monic polynomial having all its zeros at the n points of a Fekete set $\mathcal{F}_n$.

Example 7. If E is the closed unit disk centered at 0, then one can take $F_n(z) = z^n - 1$, so that $\|F_n\|_E = 2$. Comparing this with Example 5 we see that the F_n's are asymptotically optimal for the Chebyshev problem:

$$\lim_{n\to\infty} \|F_n\|_E^{1/n} = \lim_{n\to\infty} \|T_n\|_E^{1/n} = 1 = \mathrm{cheb}(E).$$

Moreover, uniformly on compact subsets of $|z| > 1$, we have

$$\lim_{n\to\infty} |F_n(z)|^{1/n} = \lim_{n\to\infty} |T_n(z)|^{1/n} = |z| = \exp\{-U^{\mu_E}(z)\},$$

(the last equality follows from Example 1). Finally, it is easy to see that the zeros of F_n are asymptotically uniformly distributed on $|z| = 1$. By that we mean that for any arc γ on this circle,

$$\frac{1}{n} \times \{\text{number of zeros of } F_n \text{ in } \gamma\} \to \frac{1}{2\pi} \times \{\text{length of } \gamma\}, \quad n \to \infty.$$

Note that the second ratio coincides with $\mu_E(\gamma)$ (cf. Example 1).

The examples of this section illustrate the following fundamental theorem, various parts of which are due to Fekete, Frostman, and Szegő.

Theorem 1 (Fundamental Theorem of Classical Potential Theory). *For any compact set $E \subset \mathbb{C}$,*

(a) cap$(E) = \tau(E) =$ cheb(E);
(b) Fekete polynomials are asymptotically optimal for the Chebyshev problem:

$$\lim_{n\to\infty} \|F_n\|_E^{1/n} = cheb(E) = cap(E).$$

If cap$(E) > 0$ (so that μ_E is defined), then we also have:

(c) Uniformly on compact subsets of the unbounded component of $\mathbb{C} \setminus E$,

$$\lim_{n\to\infty} |F_n(z)|^{1/n} = \exp\{-U^{\mu_E}(z)\};$$

(d) Fekete points (the zeros of F_n*) have asymptotic distribution* μ_E.

The last statement is illustrated in Example 7, but let us make it more precise. Let

$$P_n(z) = \prod_{k=1}^{n} (z - z_k)$$

and let δ_{z_k} denote the unit mass placed at z_k. Then $U^{\delta_{z_k}}(z) = \log \dfrac{1}{|z - z_k|}$, and we see that

$$|P_n(z)|^{1/n} = e^{-U^{\nu}(z)},$$

where ν is the unit measure (*normalized zero counting measure for* P_n) given by

$$\nu = \nu_{P_n} := \frac{1}{n} \sum_{k=1}^{n} \delta_{z_k}.$$

Notice that for any set K,

$$\nu(K) = \frac{1}{n} \times \{\text{number of zeros of } P_n \text{ in } K\}.$$

We can now rigorously formulate part (d) of the Fundamental Theorem:
The normalized zero counting measures for Fekete polynomials converge weak to* μ_E.

In applications, the following result is also useful:
Let $\{P_n\}$ *be any sequence of monic polynomials having all their zeros in* E *and such that* $\nu_{P_n} \overset{*}{\to} \mu_E$. *If* $\partial_\infty E$ *is regular (e.g., if it is connected), then the assertions (b) and (c) of the Fundamental Theorem hold for the* P_n*'s.*

Such sequences can be constructed by various "discretization" techniques. One of the simplest discretizations was employed by J.L. Walsh in his work on polynomial and rational approximation; see Remark (a) at the end of the next section.

2 Polynomial Approximation of Analytic Functions

Let f be a continuous function on a compact set E (symbolically, $f \in C(E)$) and let

$$e_n(f;E) = e_n(f) := \min_{p \in \mathcal{P}_n} \|f - p\|_E \tag{5}$$

be the error in best uniform approximation of f by polynomials of degree at most n. We denote by p_n^* the polynomial of best approximation: $\|f - p_n^*\|_E = e_n(f)$.

If $e_n(f) \to 0$ as $n \to \infty$, the series

$$p_1^* + \sum_{n=1}^{\infty} (p_{n+1}^* - p_n^*)$$

converges to f uniformly on E, so that the continuous function f must be analytic at every interior point of E. (The collection of all functions that are continuous on E and analytic in Int$E$ is denoted by $\mathcal{A}(E)$.) Furthermore, it follows from the maximum principle, that the above series automatically converges on every bounded component of $\mathbb{C} \setminus E$, so that its sum represents an analytic continuation of f to these components (e.g., if E is the unit circle $|z| = 1$, then the convergence holds in the unit disk $|z| \leq 1$). Such a continuation, however, may be impossible. Therefore, in order to ensure that $e_n(f) \to 0$ for *every function* f in $\mathcal{A}(E)$, it is necessary to assume that the only component of $\mathbb{C} \setminus E$ is the unbounded one; that is, $\mathbb{C} \setminus E$ is connected (so that E does not separate the plane).

A celebrated theorem of S.N. Mergelyan (cf. [3]) asserts that this assumption is also sufficient. Here we prove this result in a special case when E is connected and f is analytic in some neighborhood of E. The proof will also give the *rate of approximation*.

So let $\mathbb{C} \setminus E$ and E both be connected. Then the complement of E with respect to the extended complex plane is a simply-connected domain. Let Φ be the conformal map considered in Section1 and recall that

$$\log|\Phi(z)| = \log\frac{1}{\text{cap}(E)} - U^{\mu_E}(z) = g_{\mathbb{C}\setminus E}(z, \infty), \quad z \in \mathbb{C} \setminus E. \tag{6}$$

For any $R > 1$, let Γ_R denote the level curve $\{z : |\Phi(z)| = R\}$, see Fig. 2 (we call such a curve a *level curve with index* R).

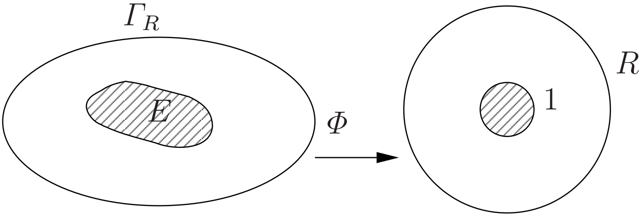


Fig. 2. Level curve of Φ

Note that Γ_R is also a level curve for the potential:

$$U^{\mu_E}(z) = \log \frac{1}{R \operatorname{cap}(E)}, \quad z \in \Gamma_R. \tag{7}$$

Let F_{n+1} be the $(n+1)$-st Fekete polynomial for E and let P_n be the polynomial of degree $\leq n$ that interpolates f at the zeros of F_{n+1}. We are given that f is analytic in a neighborhood of E; hence there exists $R > 1$ such that f is analytic on and inside Γ_R. For any such R, the *Hermite interpolation formula* yields

$$f(z) - P_n(z) = \frac{1}{2\pi \mathrm{i}} \int_{\Gamma_R} \frac{F_{n+1}(z)}{F_{n+1}(t)} \frac{f(t)\mathrm{d}t}{t-z}, \quad z \text{ inside } \Gamma_R. \tag{8}$$

(The validity of the Hermite formula follows by first observing that the right-hand side vanishes at the zeros of $F_{n+1}(z)$, and then by replacing $f(z)$ by its Cauchy integral representation to deduce that the difference between f and the right-hand side is indeed a polynomial of degree at most n).

Formula (8) leads to a simple estimate:

$$e_n(f) \leq \|f - P_n\|_E \leq K \frac{\|F_{n+1}\|_E}{\min_{\Gamma_R} |F_{n+1}(t)|},$$

where K is some constant independent of n. Applying parts (b), (c) of the Fundamental Theorem we obtain, with the aid of (7), that

$$\limsup_{n\to\infty} e_n(f)^{1/n} \leq \frac{\operatorname{cap}(E)}{R \operatorname{cap}(E)} = \frac{1}{R} < 1. \tag{9}$$

We have proved that indeed $e_n(f) \to 0$ and that the convergence is geometrically fast. Since $R > 1$ was arbitrary (but such that f is analytic on and inside Γ_R), we have actually proved that (9) holds with R replaced by $R(f)$, where

$$R(f) := \sup\{R : \ f \text{ admits analytic continuation to the interior of } \Gamma_R\}.$$

Can we improve on this? The answer is — no! In order to show this, we need the following very useful result.

Theorem 2 (Bernstein-Walsh Lemma). *Assume that both E and $\mathbb{C} \setminus E$ are connected. If a polynomial p of degree n satisfies $|p(z)| \leq M$ for $z \in E$, then $|p(z)| \leq Mr^n$ for $z \in \Gamma_r$, $r > 1$.*

The proof uses essentially the same argument as in Example 5. The function $p(z)/\Phi^n(z)$ is analytic outside E, even at ∞. Since $|\Phi| = 1$ on ∂E, we know that $|p(z)/\Phi^n(z)| \leq M$ for $z \in \partial E$. Hence the maximum principle yields

$$\left| \frac{p(z)}{\Phi^n(z)} \right| \leq M, \quad z \in \mathbb{C} \setminus E$$

and the result follows by the definition of Γ_r.

Assume now that (9) holds for some $R > R(f)$ and let $R(f) < \varrho < R$. Then for some constant $c > 1$,

$$e_n(f) \leq \frac{c}{\varrho^n}, \quad n \geq 1.$$

Since, from the triangle inequality,

$$\|p_{n+1}^* - p_n^*\|_E = \|p_{n+1}^* - f + f - p_n^*\|_E \leq e_{n+1}(f) + e_n(f) \leq 2c\varrho^{-n},$$

we obtain from the Bernstein-Walsh Lemma that for any $r > 1$,

$$\|p_{n+1}^* - p_n^*\|_{\Gamma_R} \leq 2c \left(\frac{r}{\varrho}\right)^n, \quad n \geq 1.$$

If we choose $R(f) < r < \varrho$, we obtain that the series $p_1^* + \sum_{n=1}^{\infty}(p_{n+1}^* - p_n^*)$ converges uniformly inside Γ_r. Hence it gives an analytic continuation of f to the interior of Γ_r, which contradicts the definition of $R(f)$.

Let us summarize what we have proved.

Theorem 3 (Walsh [17, Ch. VII]). *Let the compact set E be connected and have a connected complement. Then for any $f \in \mathcal{A}(E)$,*

$$\limsup_{n\to\infty} e_n(f)^{1/n} = \frac{1}{R(f)}.$$

Remarks.

(a) The proof of this theorem shows that on interpolating f at Fekete points we obtain a sequence of polynomials that gives, asymptotically, the best possible rate of approximation. It may be not easy, however, to find these points and it is desirable to have other methods at hand. Assume, for example, that E is bounded by a smooth Jordan curve Γ. With Φ as above, let the points $w_1, \dots, w_n$ be equally-spaced on $|w| = 1$ and let $z_i = \Phi^{-1}(w_i) \subset \partial E$ be their preimages. These points (called the *Fejér points*) divide Γ into n subarcs, each having μ_E-measure $1/n$ (the latter can be derived from the formula (4)). Therefore, the Fejér points have asymptotic distribution μ_E. Let P_n be the monic polynomial with zeros at $z_1, \dots, z_n$. According to the statement in the end of Section 1, the sequence $\{P_n\}$ enjoys the same properties (b), (c) as $\{F_n\}$ does, and the proof of Theorem 3 shows that

$$\limsup_{n\to\infty} \|f - P_n\|_E^{1/n} = \frac{1}{R(f)}.$$

(b) $R(f)$ is the first value of R for which the level curve Γ_R contains a singularity of f. It may well be possible that f is analytic at some other points of $\Gamma_{R(f)}$, but the geometric rate of best polynomial approximation “does not feel this” — whether every point of $\Gamma_{R(f)}$ is a singularity or merely one point is a singularity, the rate of approximation remains the same as if f was analytic only inside of $\Gamma_{R(f)}$! To take advantage of any extra analyticity, different approximation tools are needed; e.g., rational functions. We demonstrate this in Section 6.

(c) It follows from (6) that

$$\Gamma_R = \{z \in \mathbb{C} \setminus E : \ g_{\mathbb{C}\setminus E}(z, \infty) = \log R\}. \tag{10}$$

Assume now that $\mathbb{C} \setminus E$ is connected but is E *not.* Then one can still define the Green function $g_{\mathbb{C}\setminus E}$ via the formula

$$g_{\mathbb{C}\setminus E} = \log \frac{1}{\operatorname{cap}(E)} - U^{\mu_E},$$

from which it follows that properties (i)–(iii) described in Section 1 will hold, provided E is regular. Then, with Γ_R defined by (10), it is easy to modify the above proof to show that Walsh's Theorem 3 holds in this case as well.

Example 8. Let $E = [-1, -\alpha] \cup [\alpha, 1]$, $0 < \alpha < 1$, and let $f = 0$ on $[-1, -\alpha]$ and $f = 1$ on $[\alpha, 1]$. Some level curves Γ_R of $g_{\mathbb{C}\setminus E}$ are depicted on Fig. 3. For R small, Γ_R consists of two pieces, while for R large, Γ_R is a single curve. There is a "critical value" $R_0 = g_{\mathbb{C}\setminus E}(0, \infty)$ for which Γ_{R_0} represents a self-intersecting lemniscate-like curve (the bold curve in Fig. 3). Clearly, f can be extended as an analytic function to the interior of Γ_{R_0} (define $f = 0$ inside the left lobe and $f = 1$ inside the right lobe). For $R > R_0$, the interior of Γ_R is a (connected) domain; hence there is no function analytic inside of Γ_R that is equal to 0 on $[-1, -\alpha]$ and to 1 on $[\alpha, 1]$. Therefore

$$R(f) = R_0 = \exp\left\{g_{\mathbb{C}\setminus E}(0, \infty)\right\},$$

and by the (extension of) Walsh's theorem:

$$\limsup_{n\to\infty} e_n(f)^{1/n} = \exp\left\{-g_{\mathbb{C}\setminus E}(0, \infty)\right\}.$$

3 Approximation with Varying Weights — a background

We start with two problems that have triggered much of the recent potential theoretic research on polynomial and rational approximation and on orthogonal polynomials.

Let $0 < \theta < 1$. A polynomial $P(x) = \sum_{k=0}^{n} a_k x^k$ is said to be *incomplete of type* θ ($P \in I_\theta$), if $a_k = 0$ for $k < n\theta$. The study of such polynomials was introduced in [6] by Lorentz who proved the following.

Theorem 4 (G.G. Lorentz, 1976). *If* $P_n \in I_\theta$, $\deg P_n \to \infty$ *as* $n \to \infty$ *and*

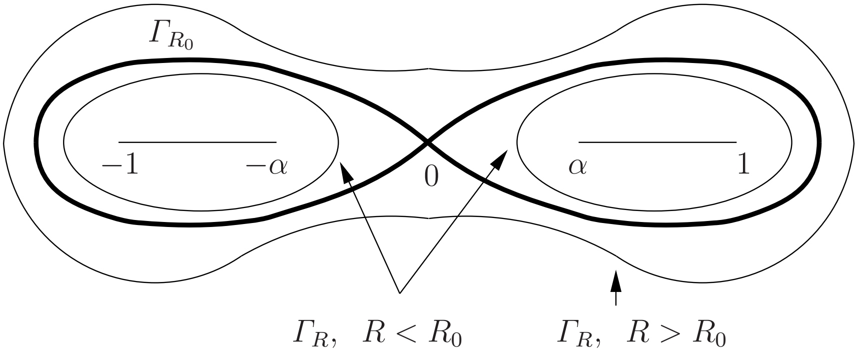


Fig. 3. Level curves of $g_{\mathbb{C}\setminus E}$

$$\|P_n\|_{[0,1]} = \max_{[0,1]} |P_n(x)| \leq M, \qquad \textit{all } n,$$

then

$$P_n(x) \to 0 \qquad \textit{for} \qquad x \in [0, \theta^2).$$

Concerning the sharpness of this result, we state

Problem 4. Is $[0, \theta^2)$ the largest interval where the convergence to zero is guaranteed?

Another problem, dealing with the asymptotic behavior of recurrence coefficients for orthogonal polynomials, was posed by G. Freud, also in 1976 [2]. Let

$$w_\alpha(x) := \mathrm{e}^{-|x|^\alpha}, \quad \alpha > 0 \tag{11}$$

be a weight on the real line and let $\{p_n\}$ be *orthonormal polynomials* with respect to this weight:

$$\int_{-\infty}^{\infty} p_m(x) p_n(x) e^{-|x|^\alpha} \mathrm{d}x = \delta_{mn}$$

(for $\alpha = 2$ these are the classical Hermite polynomials). Since the weight is even, the polynomials p_n satisfy the following 3-term recurrence relation

$$x p_n(x) = a_{n+1} p_{n+1}(x) + a_n p_{n-1}(x),$$

where $\{a_n\}$ is some sequence of real numbers (cf. [15]). For the weights (11), G. Freud conjectured that

$$\lim_{n\to\infty} n^{1/\alpha} a_n \quad \text{exists}.$$

Problem 5. Resolve this conjecture.

Seemingly very different, these two problems are connected by a common thread — both can be formulated in terms of *weighted polynomials* of the form

$$w^n(x)P_n(x), \quad \deg P_n \leq n.$$

For the Lorentz Problem, one simply observes that any $P \in I_\theta$ of degree $n/(1-\theta)$ (which for simplicity we assume to be an integer) can be written in the form

$$P(x) = x^{n\theta/(1-\theta)}P_n(x),$$

where P_n is a polynomial of degree $\leq n$. Therefore, this problem deals with sequences of weighted polynomials that satisfy

$$\|w^n P_n\|_{[0,1]} \leq M, \quad w(x) = x^{\theta/(1-\theta)}, \quad \deg P_n \leq n.$$

Regarding Problem 5, we observe that from the normalization

$$\int_{-\infty}^{\infty} p_n^2(x)\mathrm{e}^{-|x|^\alpha}\mathrm{d}x = 1$$

the substitution

$$x \to n^{1/\alpha}x, \quad p_n(x) \to P_n(x) := n^{1/2\alpha}p_n(n^{1/\alpha}x)$$

leads again to a sequence of weighted polynomials for which

$$\|w^n P_n\|_{L_2(\mathbb{R})} = 1, \quad w(x) = \mathrm{e}^{-|x|^\alpha/2}, \quad \mathrm{deg}P_n \leq n,$$

where $\|\cdot\|_{L_2(\mathbb{R})}$ is defined by

$$\|f\|_{L_2(\mathbb{R})} := \left(\int_{\mathbb{R}} |f(x)|^2\mathrm{d}x\right)^{1/2}.$$

In this framework, the following question is of fundamental importance:

Problem 6 (Generalized Weierstrass Approximation Problem). For $E \subset \mathbb{R}$ closed, $w : E \to [0,\infty)$, characterize those functions f continuous on E that are uniform limits on E of some sequence of weighted polynomials $\{w^n P_n\}$, $\mathrm{deg}P_n \leq n$.

It turns out that Problems 4, 5, and 6 can be resolved with the aid of potential theory, when an *external field* is introduced.

4 Logarithmic Potentials with External Fields

Let E be a closed (not necessarily compact) subset of $\mathbb{C}$ and let $w(z)$ be a nonnegative weight on E. We define a new "distance function" on E, replacing $|z-t|$ by $|z-t|w(z)w(t)$. This gives rise to weighted versions of logarithmic capacity, transfinite diameter and Chebyshev constant.

Weighted capacity: $\operatorname{cap}(w, E)$.

As before, let $\mathcal{M}(E)$ denote the collection of all unit measures supported on E. We set

$$Q := \log \frac{1}{w}$$

and call it the *external field.* Consider the modified energy integral for $\mu \in \mathcal{M}(E)$:

$$\begin{aligned} I_w(\mu) &:= \int\int \log \frac{1}{|z-t|w(z)w(t)} \mathrm{d}\mu(z)\mathrm{d}\mu(t) \\ &= \int\int \log \frac{1}{|z-t|} \mathrm{d}\mu(z)\mathrm{d}\mu(t) + 2\int Q(z)\mathrm{d}\mu(z) \end{aligned} \tag{12}$$

and let

$$V_w := \inf_{\mu\in\mathcal{M}(E)} I_w(\mu).$$

The *weighted capacity* is defined by

$$\operatorname{cap}(w, E) := \mathrm{e}^{-V_w}.$$

In the sequel, we assume that w satisfies the following conditions:

(i) $w > 0$ on a subset of positive logarithmic capacity;
(ii) w is continuous (or, more generally, upper semi-continuous);
(iii) If E is unbounded, then $|z|w(z) \to 0$ as $|z| \to \infty$, $z \in E$.

Under these restrictions on w, there exists a unique measure $\mu_w \in \mathcal{M}(E)$, called the *weighted equilibrium measure*, such that

$$I(\mu_w) = V_w.$$

The above integral (12) can be interpreted as the total energy of the unit charge μ, in the presence of the external field Q (in this electrostatics interpretation, the field is actually $2Q$). Since this field has a strong repelling effect near points where $w = 0$ (i.e. $Q = \infty$), assumption (iii) physically means that, for the equilibrium distribution, no charge occurs near ∞. In other words, the support $S(\mu_w)$ of μ_w is necessarily *compact.* However, unlike the unweighted case, the support need not lie entirely on $\partial_\infty E$ and, in fact, it can be quite an arbitrary closed subset of E. Determining this set is one of the most important aspects of weighted potential theory.

Weighted transfinite diameter: $\tau(w, E)$.

Let

$$\delta_n(w) := \max_{z_1, \ldots, z_n \in E} \left(\prod_{1 \le i < j \le n} |z_i - z_j| w(z_i) w(z_j) \right)^{2/n(n-1)} .$$

Points $z_1^{(n)}, \ldots, z_n^{(n)}$ at which the maximum is attained are called *weighted Fekete points*. The corresponding *Fekete polynomial* is the monic polynomial with all its zeros at these points.

As in the unweighted case, the sequence $\delta_n(w)$ is decreasing, so one can define

$$\tau(w, E) := \lim_{n \to \infty} \delta_n(w),$$

which we call the *weighted transfinite diameter* of E.

Weighted Chebyshev constant: $\mathrm{cheb}(w, E)$.

Let

$$t_n(w) := \min_{p \in \mathcal{P}_{n-1}} \| w^n(z)(z^n - p(z)) \|_E .$$

Then the *weighted Chebyshev constant* is defined by

$$\mathrm{cheb}(w, E) := \lim_{n \to \infty} t_n(w)^{1/n} .$$

The following theorem (due to Mhaskar and Saff) generalizes the classical results of Section 1.

Theorem 5 (Generalized Fundamental Theorem). *Let E be a closed set of positive capacity. Assume that w satisfies the conditions* (i)–(iii) *and let $Q = \log(1/w)$. Then*

$$cap(w, E) = \tau(w, E) = cheb(w, E) \exp\left\{ - \int Q \mathrm{d}\mu_w \right\} .$$

Moreover, weighted Fekete points have asymptotic distribution μ_w as $n \to \infty$, and weighted Fekete polynomials are asymptotically optimal for the weighted Chebyshev problem.

How can one find μ_w?

In most applications, the weight w is continuous and the set E is regular. Recall that the latter means that the classical (unweighted) equilibrium potential for E is equal to V_E *everywhere* on E, not just quasi-everywhere.

Under these assumptions, the equilibrium measure $\mu = \mu_w$ is characterized by the conditions that $\mu \in \mathcal{M}(E)$, $I(\mu) < \infty$ and, for some constant c_w, the following *variational conditions* hold:

$$\begin{cases} U^{\mu} + Q = c_w & \text{on } S(\mu) \\ U^{\mu} + Q \geq c_w & \text{on } E. \end{cases} \tag{13}$$

On integrating (against $\mu = \mu_w$) the first condition, we obtain that the constant is given by

$$c_w = I_w(\mu_w) + \int Q \mathrm{d}\mu_w = V_w - \int Q \mathrm{d}\mu_w.$$

When trying to find μ_w, an essential step (and a nontrivial problem in its own right!) is to determine the support $S(\mu_w)$. There are several methods by which $S(\mu_w)$ can be numerically approximated, but they are complicated from the computational point of view. Therefore, knowing properties of the support can be useful and we list some of them.

Properties of the support $S(\mu_w)$

(a) The sup-norm of weighted polynomials "lives" on $S(\mu_w)$. That is, for any n and for any polynomial P_n of degree at most n, there holds

$$\|w^n P_n\|_E = \|w^n P_n\|_{S(\mu_w)}.$$

(b) Let K be a compact subset of E of positive capacity, and define

$$F(K) := \log \operatorname{cap}(K) - \int_K Q \mathrm{d}\mu_K,$$

where μ_K is the classical (unweighted) equilibrium measure for K. This so-called *F-functional* of Mhaskar and Saff is often a helpful tool in finding $S(\mu_w)$. Since $\operatorname{cap}(K)$ and μ_K remain the same if we replace K by $\partial_\infty K$, we obtain that $F(K) = F(\partial_\infty K)$. It turns out that the outer boundary of $S(\mu_w)$ maximizes the F-functional:

$$\max_K F(K) = F(\partial_\infty S(\mu_K)).$$

This result is especially useful when E is a real interval and Q is *convex.* It is then easy to derive from (13) that $S(\mu_w)$ is an *interval.* Thus, to find the support, one merely needs to maximize $F(K)$ only over intervals $K \subset E$, which amounts to a standard calculus problem for the determination of the endpoints of $S(\mu_w)$.

Example 9 (Incomplete polynomials). Here $E = [0, 1]$ and

$$Q(x) = \log(1/w(x)) = -\frac{\theta}{1-\theta}\log x$$

is convex. Maximizing the F-functional one gets $S(\mu_w) = [\theta^2, 1]$. (For details, see [12, Sec. IV.1]).

Example 10 (Freud Weights). Here $E = \mathbb{R}$ and $w(x) = \exp(-|x|^\alpha)$. Hence $Q(x) = |x|^\alpha$ is convex provided that $\alpha > 1$, and we obtain $S_w = [-a_\alpha, a_\alpha]$, where a_α can be given explicitly in terms of the Gamma function. (Actually, this result also holds for all $\alpha > 0$; see [12, Sec. IV.1].) For example, when $\alpha = 2$, we get $S_w = [-1, 1]$.

5 Generalized Weierstrass Approximation Problem

We address here Problems 4, 5, and 6. Let E be a regular closed subset of $\mathbb{R}$ and $w(x)$ be continuous on E. Then we have the following weighted analogue of the Bernstein-Walsh lemma:

$$|w^n(x)P_n(x)| \le \|w^n P_n\|_{S(\mu_w)} \exp\{-n(U^{\mu_n}(x) + Q(x) - c_w)\}, \quad x \in E \setminus S(\mu_w).$$

With the aid of (13) and a variant of the Stone-Weierstrass theorem (cf. [12]), one can show that if a sequence $\{w^n(x)P_n(x)\}$, $\deg P_n \le n$, converges uniformly on E, then it tends to 0 for every $x \in E \setminus S(\mu_w)$.

Thus, if some $f \in C(E)$ is a uniform limit on E of such a sequence, it must vanish on $E \setminus S(\mu_w)$. The converse is not true, in general, but it is true in many important cases.

Incomplete polynomials

For the weight $w = x^{\theta/(1-\theta)}$, we have mentioned that $S(\mu_w) = [\theta^2, 1]$. It was proved by Saff and Varga and, independently, by M. v. Golitschek (cf. [13], [5]), that any $f \in C[0,1]$ that vanishes on $[0, \theta^2]$ is a uniform limit on $[0,1]$ of incomplete polynomials of type θ.

In particular, choosing $f(x) = 0$ for $x \in [0, \theta^2]$, and $f(x) = x - \theta^2$ for $x > \theta^2$, the sequence of type θ polynomials converging uniformly to f on $[0,1]$ is uniformly bounded on $[0,1]$, but does not tend to zero for $x > \theta^2$. Thus the answer to Problem 4 is — yes, Lorentz's Theorem 4 is indeed sharp!

Freud Conjecture

For $\alpha > 1$, let $[-a_\alpha, a_\alpha]$ be the support of the equilibrium measure for the weight $e^{-|x|^\alpha}$. Lubinsky and Saff showed in [7], that any $f \in C(\mathbb{R})$ that vanishes outside this support is a uniform limit of a sequence of the form $\exp\{-n|x|^\alpha\}P_n(x)$, $n \ge 1$. This result was the major ingredient in the argument given by Mhaskar, Lubinsky, and Saff [8], that resolved the Freud Conjecture in the affirmative.

Concerning more general weights, Saff made the following conjecture:

Conjecture 1. Let E be a real interval, and assume that $Q = \log(1/w)$ is convex on E. Then any function $f \in C(E)$ that vanishes on $E \setminus S(\mu_E)$ is the uniform limit on E of some sequence of weighted polynomials $\{w^n P_n\}$, $\deg P_n \leq n$.

This conjecture was proved by V. Totik [16] utilizing a careful analysis of the smoothness of the density of the weighted equilibrium measure. We remark that for more general Q and E, the conjecture is false, and additional requirements on f are needed.

6 Rational Approximation

For a rational function $R(z) = P_1(z)/P_2(z)$, where P_1 and P_2 are monic polynomials of degree n, one can write

$$-\frac{1}{n}\log|R(z)| = U^{\nu_1}(z) - U^{\nu_2}(z),$$

where ν_1, ν_2 are the normalized zero counting measures for P_1, P_2, respectively. The right-hand side represents the logarithmic potential of the *signed measure* $\mu = \nu_1 - \nu_2$:

$$U^{\nu_1}(z) - U^{\nu_2}(z) = U^{\mu}(z) = \int \log\frac{1}{|z-t|}\mathrm{d}\mu(t).$$

The theory of such potentials can be developed along the same lines as in Section 1. We present below only the very basic notions of this theory that are needed to formulate the approximation results. A more in-depth treatment can be found in the works of Bagby [1], Gonchar [4], as well as [12].

The analogy with electrostatics problems suggests considering the following energy problem. Let $E_1, E_2 \subset \mathbb{C}$ be two closed sets that are a positive distance apart. The pair (E_1, E_2) is called a *condenser* and the sets E_1, E_2 are called the *plates*. Let μ_1 and μ_2 be positive unit measures supported on E_1 and E_2, respectively. Consider the energy integral of the signed measure $\mu = \mu_1 - \mu_2$:

$$I(\mu) = \int\int \log\frac{1}{|z-t|}\mathrm{d}\mu(z)\mathrm{d}\mu(t).$$

Since $\mu(\mathbb{C}) = 0$, the integral is well-defined, even if one of the sets is unbounded. While not obvious, it turns out that such $I(\mu)$ is always *positive*. We assume that E_1 and E_2 have positive logarithmic capacity. Then the minimal energy (over all signed measures of the above form)

$$V(E_1, E_2) := \inf_{\mu} I(\mu)$$

is finite and positive. We then define the *condenser capacity* $\mathrm{cap}(E_1, E_2)$ by

$$\operatorname{cap}(E_1, E_2) := 1/V(E_1, E_2).$$

One can show, as with the Frostman theorem, that there exists a unique signed measure $\mu^* = \mu_1^* - \mu_2^*$ (the *equilibrium measure* for the condenser) for which $I(\mu^*) = V(E_1, E_2)$. Furthermore, the corresponding potential (called the *condenser potential*) is constant on each plate:

$$U^{\mu^*} = c_1 \text{ on } E_1, \quad U^{\mu^*} = -c_2 \text{ on } E_2, \tag{14}$$

(we assume throughout that E_1, E_2 are regular — otherwise the above equalities hold only quasi-everywhere). On integrating against μ^*, we deduce from (14) that

$$c_1 + c_2 = V(E_1, E_2) = 1/\operatorname{cap}(E_1, E_2). \tag{15}$$

We mention that (similar to the case of the conductor potential) the relations of type (14) *characterize* μ^*. Moreover, one can deduce from (14) that the measure μ_i^* is supported on the boundary (not necessarily the outer one) of E_i, $i = 1, 2$. Therefore, on replacing each E_i by its boundary, we do not change the condenser capacity or the condenser potential.

Example 11. Let E_1, E_2 be, respectively, the circles $|z| = r_1$, $|z| = r_2$, $r_1 < r_2$

These sets are invariant under rotations. Being unique, the measure μ^* is therefore also invariant under rotations and we obtain that

$$\mu_1^* = \frac{1}{2\pi r_1} ds, \quad d\mu_2^* = \frac{1}{2\pi r_2} ds,$$

where ds denotes the arclength over the respective circles E_1, E_2. Applying the result of Example 1, we find that

$$U^{\mu^*}(z) = \begin{cases} 0, & |z| > r_2 \\ \log(r_2/|z|), & r_1 \le |z| \le r_2 \\ \log(r_2/r_1), & |z| < r_1. \end{cases}$$

Therefore (recall (15))

$$\operatorname{cap}(E_1, E_2) = 1/\log \frac{r_2}{r_1}. \tag{16}$$

Assume now that each plate of a condenser is a single Jordan arc or curve (without self-intersections), and let G be the doubly-connected domain that is bounded by E_1 and E_2, see Fig. 4. We call such a G a *ring domain.*

For ring domains one can give an alternative definition of condenser capacity. Let

$$u(z) := \int \log(z - t) \mathrm{d}\mu^*(t) + c_1.$$

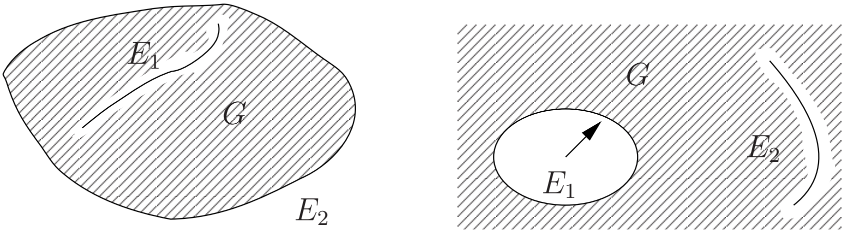


Fig. 4. Ring domains

This function is locally analytic but not single-valued in G (notice that there is no modulus sign in the integral). Moreover, if we fix t and let z move along a simple closed counterclockwise oriented curve in G that encircles E_1, say, then the imaginary part of $\log(z-t)$ increases by 2π, for $t \in E_1$, while for $t \in E_2$ it returns to the original value. Since μ_1^* and μ_2^* are unit measures, it follows that the function $\phi : z \to w = \exp(u(z))$ is analytic and single-valued. Moreover, it can be shown to be one-to-one in G. By its definition, ϕ satisfies

$$\log|\phi| = -U^{\mu^*} + c_1 = 0 \text{ on } E_1; \quad \log|\phi| = -U^{\mu^*} + c_1 = c_1 + c_2 \text{ on } E_2.$$

Therefore ϕ maps G conformally onto the annulus $1 < |w| < e^{c_1+c_2}$.

It is known from the theory of conformal mapping, that, for a ring domain G, there exists unique $R > 1$, called the *modulus of* G (we denote it by $\operatorname{mod}(G)$), such that G can be mapped conformally onto the annulus $1 < |w| < R$. We have thus shown that

$$\operatorname{cap}(E_1, E_2) = 1/\log(\operatorname{mod}(G)). \tag{17}$$

We remark that if $G_1 \supset G_2$ are two ring domains, then $\operatorname{mod}(G_1) \geq \operatorname{mod}(G_2)$.

Example 12. Let E_1, E_2 be as above, and assume that E_2 is the R-th level curve for E_1. That is, $|\Phi(z)| = R$ for $z \in E_2$, where Φ maps conformally the unbounded component of $\mathbb{C} \setminus E_1$ onto $|w| > 1$. In particular, Φ maps the corresponding ring domain G onto the annulus $1 < |w| < R$, and we conclude that $\operatorname{mod}(G) = R$ (so that $\operatorname{cap}(E_1, E_2) = 1/\log R$). Applying this to the configuration of Example 11, we see that $\Phi(z) = z/r_1$, so that $R = r_2/r_1$, and we obtain again (16).

We now turn to rational approximation. Let $E \subset \mathbb{C}$ be compact. We denote by $\mathcal{R}_n$ the collection of all rational functions of the form $R = P/Q$, where P, Q are polynomials of degree at most n, and Q has no zeros in E. For $f \in \mathcal{A}(E)$, let

$$r_n(f; E) = r_n(f) := \inf_{r \in \mathcal{R}_n} \|f - r\|_E$$

be the error in best approximation of f by rational functions from $\mathcal{R}_\backslash$. Clearly, since polynomials are rational functions, we have (cf. (5)) $r_n(f) \leq e_n(f)$. A basic theorem regarding the rate of rational approximation was proved by Walsh [17, Ch.IX]. Following is a special case of this theorem.

Theorem 6 (Walsh). *Let E be a single Jordan arc or curve and let f be analytic on a simply connected domain $D \supset E$. Then*

$$\limsup_{n\to\infty} r_n(f)^{1/n} \le \exp\{-1/\mathit{cap}(E,\partial D)\}. \tag{18}$$

The proof of (18) follows the same ideas as the proof of inequality (9). Let Γ be a contour in $D\setminus E$ that is arbitrarily close to ∂D. Let $\mu^* = \mu_1^* - \mu_2^*$ be the equilibrium measure for the condenser (E,Γ). For any n, let $\alpha_1^{(n)},\dots,\alpha_n^{(n)}$ be equally spaced on E (with respect to μ_1^*) and let $\beta_1^{(n)},\dots,\beta_n^{(n)}$ be equally spaced on Γ (with respect to μ_2^*). Then one can show that the rational functions $r_n(z)$ with zeros at the $\alpha_i^{(n)}$'s and poles at the $\beta_i^{(n)}$'s satisfy

$$\left(\frac{\max\limits_E |r_n|}{\min\limits_\Gamma |r_n|}\right)^{1/n} \to \mathrm{e}^{-1/\mathrm{cap}(E,\Gamma)}. \tag{19}$$

Let $R_n = p_{n-1}/q_n$ be the rational function with poles at the $\beta_i^{(n)}$'s that interpolates f at the points $\alpha_i^{(n)}$'s. Then the Hermite formula (cf. (8)) takes the following form:

$$f(z) - R_n(z) = \frac{1}{2\pi\mathrm{i}} \int_\Gamma \frac{r_n(z)}{r_n(t)} \frac{f(t)}{t-z}\mathrm{d}t, \quad z \text{ inside } \Gamma,$$

and it follows from (19) that

$$\limsup_{n\to\infty} r_n(f)^{1/n} \le \limsup_{n\to\infty} \|f - R_n\|_E^{1/n} \le \mathrm{e}^{-1/\mathrm{cap}(E,\Gamma)}.$$

Letting Γ approach ∂D, we get the result.

Remarks.
(a) Unlike in the polynomial approximation, no rate of convergence of $r_n(f)$ to 0 can ensure that a function $f \in C(E)$ is analytic somewhere beyond E.

(b) One can construct a function for which equality holds in (18), so that this bound is sharp. Such a function necessarily has a singularity at every point of ∂D; otherwise f would be analytic in a larger domain, so that the corresponding condenser capacity will become smaller. In view of Theorem 6, this would violate the assumed equality in (18).

Although sharp, the bound (18) is unsatisfactory, in the following sense. Assume, for example, that E is connected and has a connected complement, and let Γ_R, $R > 1$, be a level curve for E. Let f be a function that is analytic in the domain D bounded by Γ_R and such that the equality holds in (18). According to Example 12 we then obtain that

$$\limsup_{n\to\infty} r_n(f)^{1/n} = \frac{1}{R}. \tag{20}$$

By Remark (b) above, such f must have singularities on Γ_R. Hence (recall Remark (b) following Theorem 3) the relation (20) holds with $r_n(f)$ replaced by $e_n(f)$. But the family $\mathcal{R}_n$ contains $\mathcal{P}_n$ and it is much more rich than $\mathcal{P}_n$ — it depends on $2n+1$ parameters while $\mathcal{P}_n$ depends only on $n+1$ parameters. One would expect, therefore, that at least for a subsequence of n's, $r_n(f)$ behaves asymptotically like $e_{2n}(f)$. This was a motivation for the following conjecture.

Conjecture 2 (A.A. Gonchar). Let E be a compact set and f be analytic in an open set D containing E. Then

$$\liminf_{n\to\infty} r_n(f;E)^{1/n} \le \exp\{-2/\mathrm{cap}(E,\partial D)\}. \tag{21}$$

This conjecture was proved by O. Parfenov [9] for the case when E is a continuum with connected complement and in the general case by V. Prokhorov [10]; they used a very different method — the so-called “AAK Theory” (cf. [18]). However this method is not constructive, and it remains a challenging problem to find such a method. Yet, potential theory can be used to obtain bounds like (21) in the stronger form

$$\lim_{n\to\infty} r_n(f;E)^{1/n} = \exp\{-2/\mathrm{cap}(E,\partial D)\}$$

for some important *subclasses* of analytic functions, such as Markov functions [4]) and functions with a finite number of algebraic branch-points [14]).

References

1. T. Bagby, The modulus of a plane condenser, *J. Math. Mech.*, 17:315-329, 1976.
2. G. Freud, On the coefficients in the recursion formulae of orthogonal polynomials, *Proc. Roy. Irish Acad. Sect. A(1)*, 76:1-6, 1976.
3. D. Gaier, *Lectures on Complex Approximation*, Birkhäuser, Boston Inc., Boston, MA, 1987.
4. A.A. Gonchar, On the speed of rational approximation of some analytic functions, *Math USSR-Sb.*, 125(167):117-127, 1984.
5. M. v. Golitschek, Approximation by incomplete polynomials, *J. Approx. Theory*, 28:155-160, 1980.
6. G.G. Lorentz, Approximation by Incomplete Polynomials (problems and results). In E.B. Saff and R.S. Varga, editors, *Padé and Rational Approximations: Theory and Applications*, Academic Press, New York, 289-302, 1977.
7. D.S. Lubinsky, E.B. Saff, Uniform and mean approximation by certain weighted polynomials, with applications, *Constr. Approx.*, 4:21-64, 1988.
8. D.S. Lubinsky, H.N. Mhaskar, E.B. Saff, Freud's conjecture for exponential weights, *Bull. Amer. Math. Soc.*, 15:217-221, 1986.

9. O.G. Parfenov, Estimates of singular numbers of the Carleson embedding operator, *Math. USSR Sbornik*, 59:497-514, 1986.
10. V.A. Prokhorov, Rational approximation of analytic functions, *Mat. Sb*, 184:3-32, 1993, English transl. Russian Acad. Sci. Sb. Math. 78, 1994.
11. T. Ransford, *Potential Theory in the Complex Plane*, Cambridge University Press, Cambridge, 1995.
12. E.B. Saff, V. Totik, *Logarithmic Potentials with External Fields*, Springer-Verlag, New-York, 1997.
13. E.B. Saff, R.S. Varga, The sharpness of Lorentz's theorem on incomplete polynomials, *Trans. Amer. Math. Soc.*, 249:163-186, 1979.
14. H. Stahl, General convergence results for rational approximation, in: *Approximation Theory VI*, volume 2, C.K. Chui et al. (eds.), Academic Press, Boston, 605-634.
15. G. Szegő, *Orthogonal Polynomials*, volume 23 of *Colloquium Publications*, Amer. Math. Soc., Providence, R.I., 1975.
16. V. Totik, Weighted polynomial approximation for convex external fields, *Constr. Approx.*, 16:261-281, 2000.
17. J.L. Walsh, *Interpolation and Approximation by Rational Functions in the Complex Plane*, volume 20 of *Colloquium Publications,* Amer. Math. Soc., Providence, R.I., 1960.
18. N. Young, *An Introduction to Hilbert Space*, Cambridge University Press, Cambridge, 1988.

Good Bases

Jonathan R. Partington

School of Mathematics,
University of Leeds,
Leeds LS2 9JT, U.K.
J.R.Partington@leeds.ac.uk

1 Introduction

There are two standard approaches to finding rational approximants to a given function. The first approach, which we shall review in this paper, is to employ a “basis” of possible functions (interpreted in a fairly loose sense) such that the possible rational approximants are linear combinations of the basis functions, and thus given by a simple parametrization. In this situation it is required to choose the most appropriate parameters or coordinates. An alternative, which we shall not discuss, is the situation when the possible approximants are not linearly parametrized: this is seen in Padé approximation, Hankel-norm approximation, and similar schemes.

Thus the theme of this paper is to describe some families of bases that have been found to be particularly useful in problems of approximation, identification, and analysis of data. The techniques employed are mostly Hilbertian; even in a comparatively simple Banach space such as the disc algebra (the space of functions continuous on the closed unit disc and analytic on the open disc), the technical problems involved in constructing bases well-adapted to the given norm are much more complicated. In this case the functions constructed also tend to have a much less natural appearance, and seem to be of mainly theoretical interest.

Our material divides naturally into two sections. In Section 2 we shall explore situations when we have an orthonormal basis of rational functions and can use inner-product space techniques, such as least squares. Then in Section 3 we review the theory of wavelets, where the basis functions are obtained by translation and dilation of one fixed function: under these circumstances rational approximation is most usefully achieved in the context of frames, which are a convenient generalization of orthonormal bases.

J.-D. Fournier et al. (Eds.): Harm. Analysis and Ratio. Approx., LNCIS 327, pp. 95–102, 2006.

2 Orthogonal polynomials and Szegö bases

2.1 Functions in the unit disc

We recall that the Hardy space $H^2(\mathbb{D})$ is the Hilbert space consisting of all functions $f(z) = \sum_{n=0}^\infty a_n z^n$ analytic in the unit disc $\mathbb{D}$, such that $\|f\|^2 = \sum_{n=0}^\infty |a_n|^2 < \infty$. These functions have L^2 boundary values on the unit circle $\mathbb{T}$, and we have

$$\|f\|^2 = \frac{1}{2\pi}\int_0^{2\pi} |f(\mathrm{e}^{\mathrm{i}t})|^2 \, \mathrm{d}t.$$

The inner product is

$$\left\langle \sum_{n=0}^\infty a_n z^n, \sum_{n=0}^\infty b_n z^n \right\rangle = \sum_{n=0}^\infty a_n \overline{b_n}.$$

The functions $\{1, z, z^2, z^3, \dots\}$ provide a simple orthonormal basis of $H^2(\mathbb{D})$ and of course they connect very well with the theory of Fourier series, since a 2π-periodic function $g(t)$ may be identified with a function on the circle, by writing $g(t) = f(\mathrm{e}^{\mathrm{i}t})$.

Now let w be a positive continuous (weight) function defined on $\mathbb{T}$. Then it is possible to define a new inner product on $H^2(\mathbb{D})$, by

$$\langle f, g\rangle_w = \frac{1}{2\pi}\int_0^{2\pi} f(\mathrm{e}^{\mathrm{i}t})\overline{g(\mathrm{e}^{\mathrm{i}t})} w(\mathrm{e}^{\mathrm{i}t}) \, \mathrm{d}t.$$

With a view to analysing certain questions of weighted approximation, one can construct a new sequence $(g_n)_{n\geq 0}$ of polynomials, defined to be orthonormal with respect to the inner product $\langle\, .\, ,\, .\, \rangle_w$. These can be obtained by applying the Gram–Schmidt procedure to $\{1, z, z^2, z^3, \dots\}$.

It is clear that $\deg g_n = n$ for all n, and that for any $f \in H^2(\mathbb{D})$, the polynomial $p = p_N$ of degree N that minimizes the quantity

$$\frac{1}{2\pi}\int_0^{2\pi} |f(\mathrm{e}^{\mathrm{i}t}) - p(\mathrm{e}^{\mathrm{i}t})|^2 w(\mathrm{e}^{\mathrm{i}t}) \, \mathrm{d}t$$

is simply $p_N = \sum_{k=0}^N \langle f, g_k\rangle_w g_k$.

These orthogonal functions are sometimes known as *Szegö polynomials*. Indeed, it was Szegö who made the first systematic study of the asymptotic properties of such polynomials; he also looked at the convergence of expansions of analytic functions in orthogonal polynomials, and studied the location of the zeroes of such polynomials (in the situation described above, all the zeroes of g_n lie in the open unit disc). Moreover, Szegö's work goes further and includes an analysis of the behaviour of functions orthogonal with respect to a line integral along a general curve in the plane.

One particular case in which the orthogonal polynomials can be calculated very simply is when $v(\mathrm{e}^{\mathrm{i}t}) := w(\mathrm{e}^{\mathrm{i}t})^{-1}$ is a positive trigonometric polynomial. In that case, by a theorem of Fejér and Riesz, v has a *spectral factorization* as $v(\mathrm{e}^{\mathrm{i}t}) = |h(z)|^2$, where h is a polynomial in $z = \mathrm{e}^{\mathrm{i}t}$ having no zeroes in the unit disc. It can easily be verified that

$$g_n(z) = z^n \overline{h}(z^{-1}) \qquad \text{for} \quad n \geq \deg h,$$

where $\overline{h}$ denotes the polynomial whose coefficients are the complex conjugates of the coefficients of h; thus we have an explicit expression for all but a finite number of the g_n. The remaining ones are easy to calculate as well.

We shall now consider bases of more general rational functions in $H^2(\mathbb{D})$. Let $(z_n)_{n=1}^{\infty}$ be a sequence of distinct points in the unit disc satisfying

$$\sum_{n=1}^{\infty}(1 - |z_n|) = \infty,$$

which implies that the only function $f \in H^2(\mathbb{D})$ such that $f(z_n) = 0$ for all n is the identically zero function. (The negation of this condition is called the *Blaschke condition.*)

We define the *Malmquist basis* $(g_n)_{n=1}^{\infty}$ in $H^2(\mathbb{D})$ by

$$g_1(z) = \frac{(1 - |z_1|^2)^{1/2}}{1 - \overline{z}_1 z},$$

and

$$g_n(z) = \frac{(1 - |z_n|^2)^{1/2}}{1 - \overline{z}_n z} \prod_{k=1}^{n-1} \frac{z - z_k}{1 - \overline{z}_k z}, \qquad \text{for} \quad n \geq 2.$$

Note that each g_n has zeroes in the disc at $z_1, \ldots, z_{n-1}$ and poles outside the disc. In fact the functions (g_n) form an orthonormal basis for $H^2(\mathbb{D})$. The Fourier coefficients of a function f with respect to this basis are given by interpolation at the points (z_n), since

$$f(z_m) = \sum_{n=1}^{\infty} \langle f, g_n \rangle g_n(z_m),$$

for each m, and we observe that $g_n(z_m) = 0$ if $n > m$, and so we have the following formulae, which are a form of multi-point Padé approximant:

$$\begin{aligned}
f(z_1) &= \langle f, g_1 \rangle g_1(z_1),\\
f(z_2) &= \langle f, g_1 \rangle g_1(z_2) + \langle f, g_2 \rangle g_2(z_2),\\
f(z_3) &= \langle f, g_1 \rangle g_1(z_3) + \langle f, g_2 \rangle g_2(z_3) + \langle f, g_3 \rangle g_3(z_3),
\end{aligned}$$

and so on. Indeed, the Malmquist basis can also be obtained by applying the Gram–Schmidt procedure to the reproducing kernels $k_{z_n}(z) = 1/(1 - \overline{z}_n z)$, which satisfy $f(z_n) = \langle f, k_{z_n} \rangle$ for $f \in H^2(\mathbb{D})$.

Thus if we want the best rational $H^2(\mathbb{D})$ approximant to f with poles at $1/\overline{z}_1, \ldots, 1/\overline{z}_n$, then the above interpolation procedure explains how to find it.

2.2 Functions in the right half-plane

Recall that $H^2(\mathbb{C}_+)$ consists of all analytic functions $F : \mathbb{C}_+ \to \mathbb{C}$ such that

$$\|F\|_2 := \left(\sup_{x>0} \int_{-\infty}^{\infty} |F(x + \mathrm{i}y)|^2 \mathrm{d}y \right)^{1/2} < \infty,$$

(roughly speaking, functions analytic in the right half-plane, with L^2 boundary values). These are the Laplace transforms of functions in $L^2(0, \infty)$.

We cannot use polynomial approximation this time, since there are no non-constant polynomials in the space. However, two simple rational bases are of interest, namely the Laguerre basis (with poles all at one point), and the Malmquist basis (with poles all at different points).

To construct the Laguerre basis, we fix a number $a > 0$ and write

$$e_k(s) = \sqrt{\frac{a}{\pi}} \frac{(a - s)^k}{(a + s)^{k+1}}, \qquad k = 0, 1, \ldots .$$

These are a natural analogue of $\{1, z, z^2, \ldots\}$ in $H^2(\mathbb{D})$. Note that the functions e_k all have zeroes at a and poles at $-a$. Moreover, they form an orthonormal basis of $H^2(\mathbb{C}_+)$. Their inverse Laplace transforms form an orthogonal basis of $L^2(0, \infty)$ and have the form

$$f_k(t) = p_k(t) \mathrm{e}^{-at},$$

where p_k is a polynomial of degree k. In fact

$$p_k(t) = \sqrt{\frac{a}{\pi}} L_k(2at),$$

where L_k denotes the Laguerre polynomial

$$L_k(t) = \frac{\mathrm{e}^t}{k!} \frac{\mathrm{d}^k}{\mathrm{d}t^k} (t^k \mathrm{e}^{-t}).$$

Alternatively some people use *Kautz functions*, which are more appropriate for approximating lightly damped dynamical systems. These have all their poles at two complex conjugate points: the approximate models have the form

$$\frac{p(s)}{(s^2 + bs + c)^m},$$

where p is a polynomial.

It is also possible to construct Malmquist bases in the half-plane using the reproducing kernel functions for $H^2(\mathbb{C}_+)$. Recall the defining formula for a reproducing kernel, namely

$$f(s_n) = \langle f, k_{s_n} \rangle.$$

In this case the reproducing kernel functions have the formula

$$k_{s_n}(s) = \frac{1}{2\pi(s + \overline{s}_n)}.$$

The Malmquist basis functions for the right half-plane are given by

$$g_1(s) = \sqrt{\frac{1}{\pi}} \frac{(\operatorname{Re} s_1)^{1/2}}{s + \overline{s}_1}$$

and

$$g_n(s) = \sqrt{\frac{1}{\pi}} \frac{(\operatorname{Re} s_n)^{1/2}}{s + \overline{s}_n} \prod_{k=1}^{n-1} \frac{s - s_k}{s + \overline{s}_k}, \qquad \text{for } n \geq 2.$$

In some examples from the theory of linear systems, an approximate location of the poles of a rational transfer function is known, and these techniques enable one to construct models with poles in the required places. In the next section we shall see how wavelet theory enables one to gain further insight into the local behaviour of functions.

3 Wavelets

3.1 Orthonormal bases

One of the purposes of wavelet theory is to provide good orthonormal bases for function spaces such as $L^2(\mathbb{R})$. These basis functions are derived from a single function ψ by taking translated and dilated versions of it, and will be denoted $(\psi_{j,k})_{j,k \in \mathbb{Z}}$, where the parameter j controls the scaling and k controls the positioning. Thus the inner product $\langle f, \psi_{j,k} \rangle$ gives information on f at "resolution" j and "time" k. One may compare classical Fourier analysis, where the Fourier coefficients

$$\hat{f}(k) = \frac{1}{T} \int_0^T f(t) \mathrm{e}^{-2\pi \mathrm{i} kt/T} \, \mathrm{d}t = \langle f, e_k \rangle, \qquad \text{say,}$$

give us information about f at frequency $2\pi k$.

To illustrate this, we construct the *Haar wavelets*. Let $V_0 \subset L^2(\mathbb{R})$ be the closed subspace consisting of all functions f that are constant on all intervals

$(k, k+1)$, $k \in \mathbb{Z}$. Let $\phi(t) = \chi_{(0,1)}(t)$ and $\phi_k(t) = \phi(t-k)$ for $k \in \mathbb{Z}$. Then $(\phi_k)_{k\in\mathbb{Z}}$ is an orthonormal basis of V_0. Any function $f \in V_0$ has the form

$$f = \sum_{k=-\infty}^{\infty} \langle f, \phi_k \rangle \phi_k,$$

and

$$\|f\|^2 = \int_{-\infty}^{\infty} |f(t)|^2 \, \mathrm{d}t = \sum_{k=-\infty}^{\infty} |\langle f, \phi_k \rangle|^2.$$

Next, for $j \in \mathbb{Z}$, let V_j be the space of functions constant on all intervals $(k/2^j, (k+1)/2^j)$, $k \in \mathbb{Z}$. Functions in V_j have steps of width 2^{-j}. Then we have a chain of subspaces,

$$\ldots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset \ldots .$$

Also $\overline{\bigcup V_j} = L^2(\mathbb{R})$ and $\bigcap V_j = \{0\}$. Now we have a rescaling property,

$$f(t) \in V_j \iff f(2^{-j}t) \in V_0,$$

and V_j has orthonormal basis consisting of the functions $2^{j/2}\phi(2^j t - k)$ for $k \in \mathbb{Z}$. Any chain of subspaces with these properties is called a *multi-resolution approximation* or *multi-resolution analysis* of $L^2(\mathbb{R})$.

We cannot directly use the ϕ functions as an orthonormal basis of $L^2(\mathbb{R})$, and one new trick is needed. We build the Haar wavelet, which is a function bridging the gap between V_0 and V_1.

We define the *Haar wavelet* by

$$\psi(t) = \phi(2t) - \phi(2t-1) = \chi_{(0,1/2)}(t) - \chi_{(1/2,1)}(t).$$

The functions $\psi_k(t) = \psi(t-k)$, $k \in \mathbb{Z}$, form an orthonormal basis for a space W_0 such that $V_0 \oplus W_0 = V_1$ (orthogonal direct sum). Then $V_j \oplus W_j = V_{j+1}$, where W_j has orthonormal basis

$$\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k), \qquad k \in \mathbb{Z}.$$

Finally

$$L^2(\mathbb{R}) = \ldots \oplus W_{-2} \oplus W_{-1} \oplus W_0 \oplus W_1 \oplus \ldots$$

and has orthonormal basis $(\psi_{j,k})_{j,k\in\mathbb{Z}}$. Hence, if $f \in L^2(\mathbb{R})$, we have

$$f = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \langle f, \psi_{j,k} \rangle \psi_{j,k},$$

converging in L^2 norm.

In the construction sketched above, the $\psi_{j,k}$ are very simple functions, but they are all discontinuous. By working harder, one may obtain wavelets that are better adapted to approximation problems.

Here is a list of the wavelets most commonly seen in the literature. To obtain good properties of ψ and its Fourier transform $\hat{\psi}$ is not straightforward, and the following are listed in (approximately) increasing order of difficulty.[1]

Wavelet	Properties of $\psi(t)$	Properties of $\hat{\psi}(w)$
Haar	C.S., discontinuous	$O(1/w)$, C^∞
Littlewood–Paley	$O(1/t)$, C^∞	C.S., discontinuous
Meyer	Rapidly-decreasing, C^∞	C.S., can be C^∞
Battle–Lemarié	Rapidly-decreasing, C^k	$O(1/w^k)$, C^∞
Daubechies	C.S., C^k	$O(1/w^k)$, C^∞

3.2 Frames

For rational approximation, orthogonal wavelets are not so useful, and we settle for something weaker. A *frame* $(\psi_{j,k})$ in a Hilbert space H is a sequence for which there are constants A, $B > 0$ such that

$$A\|f\|^2 \leq \sum_{j,k} |\langle f, \psi_{j,k}\rangle|^2 \leq B\|f\|^2 \qquad \text{for all } f \in H.$$

This is a weaker notion than an orthonormal basis (in a finite-dimensional vector space it would correspond to a finite spanning set), but an element $f \in H$ can be reconstructed from its *frame coefficients*, the numbers $\langle f, \psi_{j,k}\rangle$. Namely, there exist dual functions $(\phi_{j,k})$ such that every $f \in H$ has the representation

$$f = \sum_{j,k} \langle f, \psi_{j,k}\rangle \phi_{j,k}.$$

The $(\phi_{j,k})$ also form a frame, the *dual frame.*

There is a general condition, due to Daubechies, which shows that the following two examples, and many others, produce frames.

- If we take

  $$\psi(t) = (1 - t^2)\mathrm{e}^{-t^2/2},$$

  the *Mexican hat* function (the puzzled reader is invited to sketch it), then the functions $\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k)$, with j, $k \in \mathbb{Z}$, form a frame for $L^2(\mathbb{R})$; these were used by Morlet in the analysis of seismic data.

[1] In the table C.S. is short for "Compact support"

- An example involving rational functions can be found by taking $\psi(t) = (1 - \mathrm{i}t)^{-n}$, with $n \geq 2$, which leads to frames in $H^2(\mathbb{C}^+)$, the Hardy space of the upper half-plane (and thus, by conformal mapping, one obtains rational frames in $H^2(\mathbb{D})$). We easily obtain rational frames in $H^2(\mathbb{C}_+)$, i.e., functions of the form
$$2^{j/2}(1 + 2^j t + \mathrm{i}k)^{-n}.$$
Their poles lie on a dyadic lattice in the right half-plane, and they have been used for approximation purposes in linear systems theory.

One advantage of such frames over orthonormal bases is their built-in redundancy; they represent a function that is, in general, less sensitive to errors or perturbations in the frame coefficients. It is this property that makes frames so useful in problems of reconstruction, as well as in certain approximation problems. Certainly non-orthogonal expansions ("non-harmonic Fourier series") have shown themselves to be an efficient alternative to more traditional methods within the last few years; nowadays, a familiarity with both methods is essential in many branches of analysis and its applications.

References

1. I. Daubechies, *Ten lectures on wavelets.* SIAM, 1992.
2. P.J. Davis, *Interpolation and approximation*, Dover, 1975.
3. G. Kaiser, *A friendly guide to wavelets.* Birkhäuser, 1994.
4. J.R. Partington, *Interpolation, identification, and sampling.* The Clarendon Press, Oxford University Press, 1997.
5. G. Szegö, *Orthogonal polynomials.* American Mathematical Society, New York, 1939.
6. J.L. Walsh, *Interpolation and approximation by rational functions in the complex domain.* American Mathematical Society, New York, 1935.

Some Aspects of the Central Limit Theorem and Related Topics

Pierre Collet

Centre de Physique Théorique
CNRS UMR 7644 Ecole Polytechnique
F-91128 Palaiseau Cedex (France)
collet@cpht.polytechnique.fr

1 Introduction

Very often the observation of natural phenomena leads to an average trend with fluctuations around it. One of the most well known example is the observation by Brown and others of a pollen particle in water. The particle is subject to many collisions with water molecules and an average behaviour follows by the law of large numbers. Here the average velocity of the particle is zero, and the particle should stay at rest. However the observation reveals an erratic motion known as Brownian motion. The goal of the central limit theorem (abbreviated below as CLT) and the related results is to study these fluctuations around the average trend.

The CLT is historically attributed to De Moivre and then to Laplace for a more rigorous study. The original argument is interesting for its relation to Statistical Mechanics and we will come back to this approach several times. I will therefore briefly present this argument although it is not the most efficient approach nowadays.

Consider a game of head or tail. One performs independent flips of a coin which has a probability p to display head and $q = 1 - p$ to display tail. We assume $0 < p < 1$ and leave to the reader the discussion of the extreme cases. One performs a large number n of independent flips and records the number $N(n)$ of times the coin displayed head. This is equivalent to a simple model of Statistical Mechanics of n uncoupled $1/2$ spins in a magnetic field. The law of large numbers gives the average behaviour of $N(n)$ for large n. Namely, with probability one

$$\lim_{n\to\infty} \frac{N(n)}{n} = p \; .$$

In other words, if the number of flips n is large, we typically observe np times the coin displaying head and $n(1-p)$ times the coin displaying tail. However the law of large numbers only tells us that $(N(n) - np)/n$ tends to zero with

J.-D. Fournier et al. (Eds.): Harm. Analysis and Ratio. Approx., LNCIS 327, pp. 105–127, 2006.

probability one. This does not say anything about the size of $N(n) - np$, namely the fluctuations.

Since the flips are independent, the probability that a sequence of n flips gives r heads (and hence $n - r$ tails) is $p^r q^{n-r}$. Therefore we obtain

$$\mathbf{P}\big(N(n) = r\big) = \binom{n}{r} p^r q^{n-r} \; . \tag{1}$$

In particular, since the events $\{N(n) = r\}$ are mutually exclusive and one of them is realized, we have

$$1 = \sum_{r=0}^{n} \mathbf{P}\big(N(n) = r\big) = \sum_{r=0}^{n} \binom{n}{r} p^r q^{n-r} \; .$$

It turns out that relatively few terms contribute to this sum. By the law of large numbers, only those terms with $r \approx np$ contribute. More precisely, using Stirling's approximation, one gets the following result for $r - np = \mathrm{O}\big(\sqrt{n}\big)$

$$\mathbf{P}\big(N(n) = r\big) = \frac{\mathrm{e}^{-(r-np)^2/(2npq)}}{\sqrt{2\pi npq}} + \mathrm{O}\left(\frac{1}{n}\right) \; . \tag{2}$$

We will discuss later on in more detail the case $|r - np| \gg \mathrm{O}\big(\sqrt{n}\big)$ and it will turn out that the event

$$|N(n) - np| \gg \mathrm{O}\big(\sqrt{n}\big)$$

has a probability which tends to zero when n tends to infinity.

Coming back to formula (2), we see that since the Gaussian function decays very fast, the number $r - np$ should be of order $\sqrt{npq}$. In other words, we now have the speed of convergence of $N(n)/n$ toward p (of the order of $1/\sqrt{n}$), or equivalently the size of the fluctuations of $N(n) - np$ (of the order of $\sqrt{n}$). This can be measured more precisely by the variance of $N(n)$ which is equal to npq

$$\mathbf{Var}\big(N(n)\big) = \sum_r r^2\, \mathbf{P}\big(N(n) = r\big) - \left(\sum_r r\mathbf{P}\big(N(n) = r\big)\right)^2 = npq \; .$$

Notice that for p fixed, with a probability very near to 1 (for large n), the observed sequence of heads and tails satisfies

$$r = np + \mathrm{O}\big(\sqrt{npq}\big) \; .$$

All these sequences have about the same probability e^{-nh} and their number is about e^{nh} where h is the entropy per flip

$$h = -\big(p \log p + q \log q\big) \; .$$

Using formula (1), the reader can give a rigorous proof of these results (see also [21]).

A more modern and more efficient approach to the CLT is due to Paul Lévy. This approach is based on the notion of characteristic function. Before we present this method, we will briefly recall some basic facts in probability theory and introduce some standard notations.

2 A short elementary probability theory refresher

We will give in this section a brief account of probability theory mostly to fix notations and to recall the main definitions and results. Fore more details we refer the reader to [13], [24], [25], [36], [17] and to the numerous other excellent books on the subject.

A real random variable X is defined by a positive measure on $\mathbb{R}$ with total mass one. In other words, for any (measurable) subset A of $\mathbb{R}$, we associate a number (weight) $\mathbf{P}(A)$ which is the probability of the event $\{X \in A\}$ ("X falls in A"). We refer to the previously mentioned references for a discussion of the interpretation of a probability. If A and B are two disjoint subsets of $\mathbb{R}$, we have $\mathbf{P}\left(A \cup B\right) = \mathbf{P}(A)+\mathbf{P}(B)$. This means that the probability of occurrence of one or the other of the two events is the sum of their probabilities (it is important that they are mutually exclusive for this formula to hold, namely that A and B are disjoint). If f is a function of a real variable, $f(X)$ is also a random variable, and we have

$$\mathbf{P}\big(f(X) \in A\big) = \mathbf{P}\big(f^{-1}(A)\big)$$

where since f may not be invertible, $f^{-1}(A)$ is defined as the set of points whose image by f is in A, namely

$$f^{-1}(A) = \big\{x \,\big|\, f(x) \in A\big\} \; .$$

The expectation of $f(X)$ denoted by $\mathbf{E}\big(f(X)\big)$ is defined by

$$\mathbf{E}\big(f(X)\big) = \int f(x) \mathrm{d}\mathbf{P}(x) \; .$$

In particular

$$\mathbf{P}\big(f(X) \in A\big) = \mathbf{E}\left(\mathbf{1}_{f(X) \in A}\right\} \; ,$$

where $\mathbf{1}_C(y)$ is the function equal to 1 if $y \in C$ and zero otherwise. The variance of $f(X)$ denoted by $\mathbf{Var}\big(f(X)\big)$ is defined by

$$\mathbf{Var}\left(f(X)\right) = \mathbf{E}\big(f(X)^2\big) - \mathbf{E}\big(f(X)\big)^2 = \mathbf{E}\bigg(\Big(f(X) - \mathbf{E}(f(X))\Big)^2\bigg) \; .$$

Observe that the variance is always non-negative, and it is equal to zero if the random variable $f(X)$ is constant except on a set of measure zero. The standard deviation is the square root of the variance. For $f(x) = x$ we obtain the average and the variance of the random variable X.

The law of X (also called its distribution) is the function $F_X(x) = \mathbf{P}((-\infty, x])$. It is equivalent to know the probability $\mathbf{P}$ or the function F_X. The characteristic function of X is defined by

$$\varphi_X(t) = \mathbf{E}\left(\mathrm{e}^{\mathrm{i}tX}\right) \ .$$

It is nothing else than the Fourier transform of $\mathbf{P}$.

We say that a property of X is true almost surely if it holds except on a set of probability zero. For example consider the Lebesgue measure on $[0,1]$ and the random variable $X(x) = x$. Then the value of X is almost surely irrational.

We recall that the measure $\mathbf{P}$ can be of different nature. It can be a finite (or countable) combination of Dirac (atomic) point mass measures. In this case X takes only a finite (or countable) number of different values. The measure $\mathbf{P}$ can also be absolutely continuous with respect to the Lebesgue measure, namely have a density h which should be a nonnegative integrable function (of integral one)

$$\mathrm{d}\mathbf{P}(x) = h(x)\mathrm{d}x \ .$$

There are other types of measures called singular continuous (like the Cantor measure described in section 7), which are neither atomic nor absolutely continuous with respect to the Lebesgue measure. The general case is a combination of these three types of measure but we emphasize that a probability measure is always normalized to have total mass one (and is always a positive measure).

We often have to consider several real random variables $X_1, \ldots, X_n$ at the same time. In order to describe their joint properties, we need to extend slightly the above definition. One considers a set Ω (equipped with a measurable structure) and a positive measure $\mathbf{P}$ of total mass one on this set (we refer to [15] for all measurability questions). In order to complete the link with the previous definition, the above simple definition corresponds to $\Omega = \mathbb{R}$. We now define as a real random variable Y as a (measurable) real valued function on Ω. Define a measure ν on $\mathbb{R}$ by

$$\nu\bigl(A\bigr) = \mathbf{P}\bigl(Y^{-1}(A)\bigr)$$

where $Y^{-1}(A)$ is as before the set of points ω in Ω such that $Y(\omega)$ belongs to A. We leave to the reader the easy exercise to check that ν is a probability measure, and if Y is the function $Y(x) = x$ we recover the above definition.

The set Ω can be chosen in various ways, for n real random variables $X_1, \ldots, X_n$ (and not more) it is convenient to take $\Omega = \mathbb{R}^n$, $\mathbf{P}$ is now a measure of total mass one on this space. Two real random variables X_1 and X_2 are said to be independent if for any pair A and B of (measurable) subsets of $\mathbb{R}$ we have

$$\mathbf{P}\left(\{X_1 \in A\} \cap \{X_2 \in B\}\right) = \mathbf{P}\left(\{X_1 \in A\}\right) \mathbf{P}\left(\{X_2 \in B\}\right) .$$

Equivalently, X_1 and X_2 are independent if for any pair of real (measurable) functions f and g we have

$$\mathbf{E}\left(f(X_1)\, g(X_2)\right) = \mathbf{E}\left(f(X_1)\right) \mathbf{E}\left(g(X_2)\right) .$$

3 Another proof of the CLT

Consider a sequence $X_1, X_2, \ldots$ of real random variables independent and identically distributed (with the same law). This is often abbreviated by i.i.d. Denote by μ the common average of these random variables and by σ their common standard deviation (both assumed to be finite and $\sigma > 0$).

As in section 1, by the law of large numbers, the sum

$$S_n = \sum_{j=1}^{n} X_j$$

behaves like $n\mu$ for large n. It is therefore natural to subtract this dominant term and to consider the sequence of random variables $S_n - n\mu$. In order to understand the behaviour of this random variable for large n, it is natural to look for a normalization (a scale), namely for a sequence of numbers (a_n) such that $(S_n - n\mu)/a_n$ stabilizes to something nontrivial (i.e. non-zero and finite). Of course, if (a_n) diverges too fast, $(S_n - n\mu)/a_n$ tends to zero, while if (a_n) diverges too slowly the limit will be almost surely infinite.

The method of characteristic functions to prove the CLT is based on the asymptotic behaviour of the sequence of functions

$$\varphi_n(s) = \mathbf{E}\left(\mathrm{e}^{\mathrm{i}s\left(S_n - n\mu\right)/a_n}\right) .$$

A very important result of Paul Lévy is the following.

Theorem 1. *If for any real number s, we have $\varphi_n(s) \to \varphi(s)$, and the function $\varphi(s)$ is continuous at $s = 0$, then φ is the Fourier transform of a probability measure ν (on $\mathbb{R}$), and*

$$\mathbf{P}\left(\frac{S_n - n\mu}{a_n} \le x\right) \overset{n\to\infty}{\longrightarrow} \nu\left((-\infty, x]\right) .$$

We refer to [24] or other standard probability books for a proof. We will say that a sequence $\big(\varrho_n\big)$ of probabilities on $\mathbb{R}$ converges in law to the probability ν if for any real number x we have

$$\lim_{n\to\infty} \varrho_n((-\infty, x]) = \nu((-\infty, x]) \ .$$

This implies (see [31]) that for any (measurable) set B such that $\nu(\partial B) = 0$,

$$\lim_{n\to\infty} \mu_n(B) = \nu(B) \ .$$

In other words, Lévy's Theorem relates the convergence in law to the convergence of the characteristic functions.

In order to be able to apply Lévy's Theorem, we have to understand the behaviour of $\varphi_n(s)$ for large n. Since the random variables $X_1, X_2, \ldots$ are independent, we have

$$\varphi_n(s) = \psi(s/a_n)^n$$

where

$$\psi(s) = \mathbf{E}\left(\mathrm{e}^{-\mathrm{i}s\left(X_j-\mu\right)}\right) \ .$$

If we assume that the numbers a_n diverge with n, we have for any fixed s

$$\psi(s/a_n) = 1 - \frac{s^2\sigma^2}{2a_n^2} + o\left(\frac{1}{a_n^2}\right) \ .$$

This estimate follows from the Lebesgue dominated convergence theorem. It also follows more easily if the fourth order moment is finite. We now see that except for a fixed change of scale, there is only one choice of the sequence $\big(a_n\big)$ (more precisely of its asymptotic behaviour) for which we obtain a non-trivial limit, namely $a_n = \sqrt{n}$. Indeed, with this choice we have

$$\varphi_n(s) = \left(1 - \frac{s^2\sigma^2}{2a_n^2} + o\left(\frac{1}{a_n^2}\right)\right)^n \overset{n\to\infty}{\longrightarrow} \mathrm{e}^{-s^2\sigma^2/2} \ .$$

We now observe that

$$\mathrm{e}^{-s^2\sigma^2/2} = \frac{1}{\sqrt{2\pi}\ \sigma}\int_{-\infty}^{\infty} \mathrm{e}^{-ix}\mathrm{e}^{-x^2/(2\sigma^2)}\mathrm{d}x$$

and we can apply the above Lévy theorem which proves the following version of the CLT.

Theorem 2. *Let $\big(X_j\big)$ be a sequence of i.i.d. real random variables with mean μ (finite) and standard deviation σ (finite and non-zero). Then, for any real number x,*

$$\lim_{n\to\infty} \mathbf{P}\left(\frac{S_n - n\mu}{n\sigma} \leq x\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \mathrm{e}^{-y^2/2}\mathrm{d}y \ .$$

4 Some extensions and related results

Unless otherwise stated, we will all along this section consider a sequence of i.i.d. real random variables $\big(X_j\big)$. We will denote by μ their common average (assumed to be finite), and by σ^2 their common variance (assumed to be finite and non-zero). The sequence S_n of partial sums is defined by

$$S_n = \sum_{j=1}^{n} X_j \ .$$

4.1 Other proofs of the CLT

We have already seen the combinatorial proof of De Moivre and the proof using characteristic functions. There are many other proofs, for example the proof due to Lindeberg based on a semi-group idea (see [13]), the proof of Kolmogorov also based on a semi-group idea (see [6]), the so-called Stein method (see [32]), and many others. A useful extension deals with the case of independent random variables but with different distributions. In this context one has the well known Lindeberg-Feller theorem (see [13]).

Theorem 3. *Let $\big(X_j\big)$ be a sequence of independent real random variables with averages $\big(\mu_j\big)$ (finite) and variances $\big(\sigma_j^2\big)$ (finite and non-zero). In other words*

$$\mathbf{E}\big(X_j\big) = \mu_j , \qquad \sigma_j = \sqrt{\mathbf{Var}\big(X_j\big)} = \sqrt{\mathbf{E}\left(X_j^2\right) - \mathbf{E}\left(X_j\right)^2} \ .$$

Let

$$s_n^2 = \sum_{j=1}^{n} \sigma_j^2 \ .$$

If for any $t > 0$

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{j=1}^{n} \mathbf{E}\left(X_j^2 \mathbf{1}_{\left\{\left|X_j\right| > t s_n\right\}} \right) = 0 \ ,$$

then $\big(S_n - \sum_{j=1}^{n} \mu_j\big)/s_n$ converges in law to a Gaussian random variable with zero mean and unit variance.

4.2 Rate of convergence in the CLT

One can control the rate of convergence if something is known about moments higher than the second one. A classical result is the Berry-Esseen theorem for i.i.d. real random variables.

Theorem 4. *Let $\varrho = \mathbf{E}\big(\big|X_j\big|^3\big) < \infty$, then for any real number x and for any integer n*

$$\sup_x \left|\mathbf{P}\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} \le x\right) - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \mathrm{e}^{-s^2/2}\mathrm{d}s\right| \le \frac{33}{4}\frac{\varrho}{\sigma^3\sqrt{n}}\ .$$

This result can be used for finite n if one has information about the three numbers μ, σ and ϱ. If one assumes that higher order moments are finite, one can construct higher order approximations. They involve Hermite functions (Edgeworth expansion). We refer the reader to [13] and [4] for more details.

4.3 Other types of convergence

A first result is the so-called local CLT which deals with the convergence of probability densities (if they exist). The simplest version is as follows.

Theorem 5. *If the common characteristic function ψ of the real i.i.d. random variables $\big(X_j\big)$ is summable (its modulus is integrable), then for any integer n the random variable $\big(S_n - n\mu\big)/(\sigma\sqrt{n})$ has a density f_n (its distribution has a density with respect to the Lebesgue measure), and we have uniformly in x*

$$\lim_{n\to\infty} f_n(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}\ .$$

Stronger versions of this result have been proved recently. We refer the reader to [2] for the convergence in the sense of Fischer information and in the sense of relative entropy.

One may wonder if a stronger form of convergence may hold. For example one could be tempted (as is often seen in bad texts) to formulate the CLT by saying that there is a Gaussian random variable ξ with zero average and variance unity such that

$$\frac{S_n - n\mu}{\sqrt{n}} = \sigma\xi + \varepsilon_n \tag{3}$$

with $\varepsilon_n \to 0$ when n tends to infinity. This is not true in general, one can consider for example the case of i.i.d. Gaussian random variables and use the associated Hilbert space representation. This only holds in the weaker sense of distributions as stated above. There is however a so-called almost sure version of the CLT. In some sense it accumulates all the information gathered for the various values of n. A simple version is as follows.

Theorem 6. *For any real number x, we have almost surely*

$$\lim_{n\to\infty} \frac{1}{\log n}\sum_{j=1}^{n}\frac{1}{j}\vartheta\left(x - \frac{S_j - j\mu}{\sigma\sqrt{j}}\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-s^2/2}\mathrm{d}s\ .$$

where $\vartheta(y)$ is the Heaviside function which vanishes for $y < 0$ and equals 1 for $y > 0$. We refer to [3] for references and a review of the results in this domain. We only emphasize that $1/j$ is essentially the unique weight for which the result holds.

4.4 Bounds on the fluctuations

A classical theorem about fluctuations is the law of iterated logarithms which gives the asymptotic size of the fluctuations.

Theorem 7. *Assume that for some $\delta > 0$, $0 < \mathbf{E}\big([X_j|^{2+\delta}\big) < \infty$. Then we have almost surely*

$$\limsup_{n\to\infty} \frac{S_n - n\mu}{\big(n\sigma^2 \log\log\big(n\sigma^2\big)\big)^{1/2}} = 1 \ .$$

There is of course an analogous result for the $\liminf$. We refer the reader to [33] for a proof and similar results.

4.5 Brownian motion

It is also quite natural to study the sequence $S_n - n\mu$ as a function of n and to ask if there is a normalization of the sequence and of the time (n) such that one obtains a non-trivial limit. Let $\xi_n(t)$ be the sequence of random functions of time (t) defined by

$$\xi_n(t) = \frac{1}{\sigma\sqrt{n}} \sum_{j=1}^{[nt]} \big(X_j - \mu\big) \ ,$$

where $[\,\cdot\,]$ denotes the integer part. This function is piecewise constant and has discontinuities for some rational values of t. One can also interpolate linearly to obtain a continuous function. Note that this is a random function since it depends on the random variables $\big(X_j\big)$. More generally, a random function on $\mathbb{R}^+$ (or $\mathbb{R}$) is called a stochastic process.

An important result is that this sequence of processes converges to the Brownian motion in a suitable sense. We recall that the Brownian motion $B(t)$ is a real valued Gaussian stochastic process with zero average and such that

$$\mathbf{E}\big(B_t B_s\big) = \min\{t, s\} \ .$$

We refer the reader to [5] or [14] for the definition of convergence and the proof. We refer to [37] for the description of the original experimental observation by Brown.

A related result is connected with the question of convergence of the sequence of random variables $\big(S_n - n\mu\big)/\sqrt{n}$ to a Gaussian random variable. This is the almost sure invariance principle.

Theorem 8. *For any sequence $\big(X_j\big)$ of i.i.d. real random variables with zero average, non-zero variance σ^2, and such that for some $\delta > 0$,*

$$\mathbf{E}\big([X_1|^{2+\delta}\big) < \infty \ ,$$

there exists another (enriched) probability space with a sequence $\big(\tilde{S}_n\big)$ or real valued random variables having the same joint distributions as the sequence $\big(S_n\big)$, a Brownian motion $\tilde{B}(t)$, and two constants $C > 0$ and $0 < \lambda < 1/2$ such that almost surely

$$\Big|\big(\tilde{S}_n - n\mu\big) - \sigma\tilde{B}(n)\Big| \le Cn^{1/2-\lambda} .$$

In other words, there exists on this other probability space an integer valued random variable $\tilde{N}$ such that for any $n > \tilde{N}$ the above inequality is satisfied. Using the scaling properties of the Brownian motion, we have also (with $C' = C/\sigma$)

$$\left|\frac{\tilde{S}_n - n\mu}{\sigma\sqrt{n}} - \tilde{B}(1)\right| \le C'n^{-\lambda} .$$

We see how this result escapes from the difficulty mentioned about the formulation (3) by constructing in some sense a larger probability space which contains the limit. We refer the reader to [28] or [33] for a proof.

We also stress an important consequence of the central limit theorem which explains the ubiquity of the Brownian motion. A stochastic process (random function) is called continuous if its realizations are almost surely continuous.

Theorem 9. *Any continuous stochastic process with independent increment has Gaussian increments.*

We refer to [14] for a proof. It is also possible to express any such process in terms of the Brownian motion. Indeed, if $\xi(t)$ is a continuous stochastic process (with $\xi(0) = 0$) with independent increments, there are two (deterministic) functions $e(t)$ and $\sigma(t)$ such that

$$\xi(t) = e(t) + \int_0^t \sigma(s)\mathrm{d}B_s .$$

The integral in the above formula has to be defined in a suitable way since the function B_s is almost surely not differentiable. We refer to [14] for the details. In the physics literature, the derivative of B (in fact a random distribution) is known as a white noise. The independence of the increments reflects the fact that the system is submitted to a noise which is renovating at a rate much faster than the typical rate of evolution of the system. We also refer to [37] for more discussions on this subject.

4.6 Dependent random variables

There are many extensions of the CLT and of the above mentioned results to the case of dependent random variables under different assumptions. A first difficulty is that even for non-trivial random variables, the asymptotic variance may vanish. Indeed, let (Y_j) be a sequence of real i.i.d. random variables with finite non-zero variance σ^2, and consider the sequence (X_j) given by

$$X_j = Y_{j+1} - Y_j \ .$$

It is easy to verify that $\mathbf{E}\big(X_j\big) = 0$ and the common variance is $2\sigma^2 > 0$. However, we have

$$S_n = \sum_{j=1}^{n} X_j = X_{n+1} - X_1$$

which implies that $S_n/\sqrt{n}$ converges in law to zero. Also, the variance of $S_n/\sqrt{n}$ is equal to $2\sigma^2/n$ and tends to zero when n tends to infinity. We refer to [19] for a general discussion around this phenomenon.

For a sequence of non-independent random variables (X_j), the variance of $S_n/\sqrt{n}$ involves the correlation functions

$$C_{i,j} = C(X_i, X_j) = \mathbf{E}\left(X_i X_j\right) - \mathbf{E}\left(X_i\right)\mathbf{E}\left(X_j\right) \ .$$

If we moreover assume that the sequence is stationary (i.e. the joint distributions of $X_{i_1}, \ldots, X_{i_k}$ are equal to those of $X_{i_1+l}, \ldots, X_{i_k+l}$ for any k, $i_1, \ldots, i_k$ and any $l > -\min\{i_1, \ldots, i_k\}$), then $C_{i,j}$ depends only on $|i-j|$. In this case, if as a function of $|i-j|$, $|C_{i,j}|$ is summable, then

$$\lim_{n\to\infty} \mathbf{E}\left(\frac{\big(S_n - n\mu\big)^2}{n}\right) = \mathbf{E}\left(\big(X_1-\mu\big)^2\right) + 2\sum_{j=2}^{\infty} \mathbf{E}\left((X_1-\mu)((X_j-\mu)\right) \ . \tag{4}$$

Under slightly stronger assumptions on the decay of correlations (and some other technical hypothesis) one can prove a CLT and many other related results. We refer to [30] for the precise hypothesis, proofs and references.

A situation where non-independent random variables appear naturally is the case of dynamical systems. Consider for example the map of the unit interval $f(x) = 2x \pmod 1$. It is easy to verify that the Lebesgue measure is invariant $\big(\mu\big(f^{-1}(A)\big) = \mu(A)\big)$ and ergodic (the law of large numbers holds). It is a probability measure on the set $\Omega = [0,1]$. If g is a real function, one defines a sequence of identically distributed real random variables $\big(X_j\big)$ by

$$X_j(x) = g\big(f^{j-1}(x)\big)$$

where f^n denotes the n^{th} iterate of f. Namely, $f^0(x) = x$, and for any integer n, $f^n(x) = f\big(f^{n-1}(x)\big)$. In general the random variables $\big(X_j\big)$ are not independent. Note that here the randomness is only coming from the choice of the initial condition x under the Lebesgue measure. In this context it is natural to ask about the asymptotic fluctuations of ergodic sum

$$S_n(x) = \frac{1}{n}\sum_{j=1}^{n} g\big(f^{j-1}(x)\big) = \sum_{j=1}^{n} X_j \ ,$$

and to wonder if there is a central limit theorem. In order to ensure that the asymptotic variance does not vanish, one has to impose that g is not of the form $u - u \circ f$ and with this assumption one can prove a CLT. We refer to [16] or [8] for the details and to [38] for more general cases.

Of course it may happen that even though the $\big(X_j\big)$ have a finite variance, the quantity (4) diverges. This is for example the case for some observables in a second order phase transition in Statistical Mechanics. One should use a non-trivial normalization to understand the fluctuations. Some non-Gaussian limiting distributions may then show up. We refer to [18] for a review of this question in connection with probability theory.

5 Statistical Applications

The CLT is one of the main tool in statistics. For example it allows to construct confidence intervals for statistical tests. We refer to [7] for a detailed exposition and many other statistical applications. There are also many results about fluctuations of empirical distributions, we refer to [35] for more on this subject.

6 Large deviations

The CLT describes the fluctuations of order $\sqrt{n}$ of a sum of n random variables having zero average. One can also ask what would be the probability of observing a fluctuation of larger (untypical) size. For example, a giant fluctuation (large deviation) which would provide a wrong estimate of the average (i.e. an anomaly in the law of large numbers). There are many results in this direction starting with Chernov's exponential bound. We will give some ideas for the i.i.d. case, and refer to the literature for deeper results.

We will assume that for any real s, the random variable $\exp(sX_j)$ is integrable (existence of exponential moments). One can then define the sequence of functions

$$Z_n(s) = \mathbf{E}\left(\mathrm{e}^{sS_n}\right) \ . \tag{5}$$

Using the i.i.d. property, it follows immediately that

$$Z_n(s) = \big(\mathbf{E}\left(\mathrm{e}^{sX_1}\right)\big)^n \ . \tag{6}$$

Therefore we immediately conclude the existence of the limit

$$\lim_{n\to\infty} \frac{1}{n} \log Z_n(s) = P(s) = \log \mathbf{E}\left(\mathrm{e}^{sX_1}\right) \ .$$

We now come back to (5). If we want to know (estimate) the probability of the event $\big\{S_n > n\alpha\big\}$ ($\alpha > 0$), we can obtain an upper bound as follows using Chebishev's inequality. Starting from (5) we have for any s

$$Z_n(s) \geq \mathbf{E}\left(\mathrm{e}^{sS_n}\mathbf{1}_{\big\{S_n > n\alpha\big\}}\right) \geq \mathrm{e}^{ns\alpha}\mathbf{P}\left(\big\{S_n > n\alpha\big\}\right) \ .$$

Using (6) we have

$$\mathbf{P}\left(\big\{S_n > n\alpha\big\}\right) \leq \mathrm{e}^{-n\big(s\alpha - P(s)\big)} \ .$$

α being kept fixed, we now choose s optimally. In other words, we take the value of s minimizing $s\alpha - P(s)$. In doing so there appears the so called Legendre transform of the function P defined by

$$\varphi(\alpha) = \sup_s \big(s\alpha - P(s)\big) \ . \tag{7}$$

If the function P is differentiable, the optimal s is obtained by solving the equation

$$\alpha = \frac{\mathrm{d}P}{\mathrm{d}s}(s) \ . \tag{8}$$

One may wonder (and should wonder) if the solution is unique. It is easy to see that P is a convex function. We leave to the reader the interesting exercise of computing $P'(s)$ and $P''(s)$ and to interpret the results in particular for $s = 0$. The solution of the problem (7) is unique unless P has affine pieces. This occurs in statistical mechanics in the presence of phase transitions (we refer to [22] for more details). Finally, we have

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbf{P}\left(\big\{S_n > n\alpha\big\}\right) \leq -\varphi(\alpha) \ .$$

With some more work, one can also obtain a lower bound. The following result is due to Plachky and Steinebach.

Theorem 10. *Let $\big(W_j\big)$ be a sequence of real random variables and assume that there exists a number $T > 0$ such that*

i)

$$Z_n(t) = \mathbf{E}\left(\mathrm{e}^{tW_n}\right) < \infty$$

for any $0 \leq t < T$ and any integer n.

ii)

$$P(t) = \lim_{n\to\infty} \frac{1}{n} \log Z_n(t)$$

exists for any $0 < t < T$, is differentiable on $(0, T)$, and P' is strictly monotone on the interval $(0, T)$.

Then for any

$$\alpha \in \big\{P'(t) \,\big|\, t \in \big(0, T\big)\big\}$$

we have

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbf{P}\left(W_n > n\alpha\right) = -\varphi(\alpha)$$

where

$$\varphi(\alpha) = \sup_{t\in(0,T)} \big(\alpha t - P(t)\big) \, .$$

We refer to [29] for a proof. In the present context, one applies this result with $W_n = S_n - n\mu$, or $W_n = -S_n + n\mu$ to obtain information on the large deviations in the other direction. In the case where $\mu = 0$, it is an interesting exercise to compute the first and second derivatives of $\varphi(\alpha)$ in $\alpha = 0$ and to relate at least intuitively the above result to the CLT.

We now give an application to the (easy) case of the game of head or tail discussed in the introduction. Formula (1) already solves the problem in this case, namely one gets easily for $q > x > 0$ using Stirling's formula

$$\mathbf{P}\big(N(n) > n(p+x)\big) \sim \frac{\mathcal{O}(1)}{\sqrt{n}} \mathrm{e}^{(q-x)\log(q-x)+(p+x)\log(p+x)-(q-\alpha)\log q-(p+x)\log p} \, .$$

In other words,

$$\varphi(p+x) = (q-x)\log(q-x) + (p+x)\log(p+x) - (q-x)\log q - (p+x)\log p \, . \tag{9}$$

A similar formula holds for the large deviations below np. Let us recover this expression using the large deviation formalism (this is essentially the original Chernov's bound). We first have to compute the partition function

$$Z_n(s) = \mathbf{E}\left(\mathrm{e}^{s(N(n)-np)}\right) = \mathrm{e}^{-nps} \sum_{r=0}^{n} \mathrm{e}^{sr} \mathbf{P}\Big(N(n) = r\Big) = \mathrm{e}^{-nps}\left(p\mathrm{e}^{s} + q\right)^{n} \, .$$

This immediately implies

$$P(s) = \log\left(p\mathrm{e}^{s} + q\right) - ps \, .$$

We therefore get

$$\alpha = \frac{p\mathrm{e}^{s}}{p\mathrm{e}^{s} + q} - p \, .$$

After easy manipulations we obtain

$$\varphi(\alpha) = (\alpha + p)\log(\alpha + p) + (q - \alpha)\log(q - \alpha) - (\alpha + p)\log p - (q - \alpha)\log q$$

which is identical to (9).

The initial paper by Lanford [22] is still a fundamental reference. In particular, it makes the connection with the formalism of Statistical Mechanics. One can also refer to [29], [11], [10] and many other books and articles.

Note that these results are formulated in terms of the asymptotics of

$$\frac{1}{n}\log \mathbf{P}\left(\{S_n > n\alpha\}\right) \ .$$

In other words, they don't say anything on the behaviour of $e^{n\varphi(\alpha)}\times \mathbf{P}\left(\{S_n > n\alpha\}\right)$ except that it should be sub-exponential. One can compare for example with the more precise formula (9). For results in this direction one can refer to [27], or [26]. This question also falls in the realm of recent results concerning the so called concentration phenomenon (see [34]).

7 Multifractal measures

One among the numerous applications of large deviations is the analysis of the multifractal behaviour of measures. We first introduce briefly this notion. In order to simplify the discussion we will restrict ourselves to (positive) measures on the unit interval, the extension to higher dimension being more or less immediate. The driving question in the multifractal analysis of a (positive) measure μ is what is the measure of a small interval. The simplest behaviour that immediately comes to mind is that the measure of any interval of length r could be proportional to r. More precisely, we will say that a measure is monofractal if there is a number $\delta > 0$ and two positive numbers $C_1 < C_2$ such that for any point $x \in [0,1]$ belonging to the support of μ and for any $r > 0$ small enough

$$C_1 r^{\delta} \le \mu\big(B_r(x)\big) \le C_2 r^{\delta} \tag{10}$$

where $B_r(x)$ is the interval $[x-r, x+r]$ (or more precisely $[x-r,x+r]\cap[0,1]$). The Lebesgue measure satisfies this property with $\delta = 1$ (with $C_2 = 2$ and $C_1 = 1$ because of the boundary points). The number δ is intuitively related to a dimension. If one considers the Lebesgue measure in dimension two, one gets a similar relation with exponent two, and this extends immediately to any dimension. We will say more about this below. Another interesting case is the Cantor set K. This set can be defined easily as the set of real numbers in $[0,1]$ whose triadic expansion does not contain one. In other words

$$K = \left\{ \sum_{j=1}^{\infty} \eta_j 3^{-j} \,\middle|\, \eta_j \in \{0,2\} \right\} \ .$$

It is well known (see [12] or [23]) that this set has dimension $\log 2/\log 3$. This set can also be defined as the intersection of a decreasing sequence of finite unions of closed intervals. Namely for any $n \geq 1$ and for any finite sequence $\eta_1, \ldots, \eta_n$ of numbers 0 or 2, let

$$x_{\eta_1, \ldots, \eta_n} = \sum_{j=1}^{n} \eta_j 3^{-j} ,$$

and

$$I_{\eta_1, \ldots, \eta_n} = \left[x_{\eta_1, \ldots, \eta_n}, x_{\eta_1, \ldots, \eta_n} + 3^{-n} \right] .$$

It is left to the reader to verify that

$$K = \bigcap_{n} \bigcup_{\eta_1, \ldots, \eta_n} I_{\eta_1, \ldots, \eta_n} .$$

Since each interval $I_{\eta_1, \ldots, \eta_n}$ has length 3^{-n}, and there are 2^n such intervals, this almost immediately leads to the above mentioned fact that the (Hausdorff) dimension of K is $\log 2/\log 3$. Let us now define a measure μ on K (the Cantor measure) by imposing

$$\mu\big(I_{\eta_1, \ldots, \eta_n}\big) = 2^{-n} .$$

There are various ways to prove that this indeed defines a probability measure supported by K. We refer to [12] or [23] for the details. We now check (10). For $x \in K$ and a given $r > 0$ ($r < 1/3$), let n be the unique integer such that $3^{-n} \leq r \leq 3^{-n+1}$. It is easy to check that there is a finite sequence $\eta_1, \ldots, \eta_n$ of numbers equal to 0 or 2 such that

$$I_{\eta_1, \ldots, \eta_n} \subset B_r(x) \quad \text{and} \quad B_r(x) \cap K \subset I_{\eta_1, \ldots, \eta_{n-1}} .$$

Therefore

$$2^{-n} \leq \mu\big(B_r(x)\big) \leq 2^{-n+1}$$

and we obtain an estimate (10) with $\delta = \log 2/\log 3$ (it is left to the reader to compute the two constants C_1 and C_2). We see again a relation between δ and the dimension. This is a general fact discovered by Frostman, namely if (10) holds, the dimension of any set of non-zero μ measure is at least δ. There is a converse to this result known as Frostman's Lemma. We refer to [20] for the complete statement and a proof. We only sketch the proof of the direct (easy) part. Recall (see [20], [12] or [23]) that the Hausdorff dimension of a set A is defined as follows. Let $B_{r_j}\big(x_j\big)$ be a sequence of balls covering A, namely

$$A \subset \bigcup_{j} B_{r_j}\big(x_j\big) .$$

For a numbers $d > 0$ and $\varepsilon > 0$, define

$$\mathrm{H}_d(\varepsilon) = \inf_{A\subset\cup_j B_{r_j}(x_j)\,,\,\sup_j r_j\le\varepsilon} \sum_j r_j^d \;.$$

This is obviously a non-increasing function of ε which may diverge when ε tends to zero. Moreover, if the limit when $\varepsilon \searrow 0$ is finite for some d, it is equal to zero for any larger d. This limit is also non-increasing in d. Moreover, if it is finite and non-zero for some d, it is infinite for any smaller d. The Hausdorff dimension of A, denoted below by $d_{\mathrm{H}}(A)$ is defined as the infimum of the set of d such that the limit vanishes (for this special d the limit may be infinite). This is also the supremum of the set of d such that the limit is infinite. Coming back to (10), let $B_{r_j}(x_j)$ be a sequence of balls covering a set A. Using (10) we have immediately

$$\mu(A) \le \sum_j \mu\big(B_{r_j}(x_j)\big) \le C_2 \sum_j r_j^\delta \;.$$

Therefore, if $\mu(A) > 0$,

$$\lim_{\varepsilon\searrow 0} \mathrm{H}_\delta(\varepsilon) > 0 \;,$$

and hence the Hausdorff dimension of A is at least δ.

Monofractal measures are relatively simple objects and one often encounters more complicated situations. The notion of multifractal measures originated from theoretical investigations on turbulence. There a whole spectrum of values for the exponent δ is allowed. The multifractal analysis is devoted to understanding the characteristics of the sets where the exponent has a given value. For $\delta > 0$ we define the set

$$E_\delta = \left\{ x \;\middle|\; \lim_{r\to 0} \frac{\log \mu\big(B_r(x)\big)}{\log r} = \delta \right\} \;. \tag{11}$$

Roughly speaking, if $x \in E_\delta$, then $\mu\big(B_r(x)\big) \sim r^\delta$. One way to say something interesting about these sets is to compute their (Hausdorff) dimension. Note that for monofractal measures, these sets are all empty except one which is the support of the measure. In the simplest generalization, all these sets have measure zero except one which has full measure. The corresponding δ is called in this case the dimension of the measure. We warn the reader that one can construct wilder examples of measures μ where the set of δ with $\mu\big(E_\delta\big) > 0$ is of cardinality larger than one, and even wilder examples where the sets E_δ are not well defined.

A way to obtain the Hausdorff dimension of the sets E_δ is to use the thermodynamic formalism. This method works under some assumptions on the measure μ for which we refer the reader to the literature (see for example [9] or [1]). We first consider a sequence $\big(\mathrm{A}_n\big)$ of partitions of the support of μ with atoms of decreasing diameter. The simplest case is to use a partition with atoms of equal size, for example p dyadic partition. We also assume that the cardinality of A_n grows exponentially fast with n, more precisely that there is a number $\gamma > 0$ such that

$$\lim_{n\to\infty} \frac{1}{n} \log \#\big(\mathrm{A}_n\big) = \gamma$$

where $\#(\,\cdot\,)$ denotes the cardinality. From now on we will only consider this case. We then consider the sequence of partition functions at inverse temperature β defined by

$$Z_n(\beta) = \frac{1}{\#\big(\mathrm{A}_n\big)} \sum_{I\in \mathrm{A}_n} \mu(I)^{\beta} \ .$$

The parameter β may be chosen positive or negative (this is a difference with standard Statistical Mechanics where temperature which is proportional to the inverse of β is non-negative). We now define the pressure function $P(\beta)$ (if it exists) by

$$P(\beta) = \lim_{n\to\infty} \frac{1}{n} \log Z_n(\beta) \ .$$

Note that

$$-\frac{\mathrm{d}\log Z_n}{\mathrm{d}\beta}(1) = -\sum_{I\in \mathrm{A}_n} \mu(I) \log \mu(I)$$

which is the entropy of μ with respect to the partition A_n. For $\beta \neq 1$, the quantity $P(\beta)/(\beta-1)$ is often referred to as the Renyi entropy. At this point, provided the function P is non-trivial, we can make the link with large deviations. For this purpose, we first define a sequence (W_n) of random variables. The values of W_n are the numbers $\beta \log \mu(I)$ $(I \in \mathrm{A}_n)$ and the corresponding probability is $1/\#\big(\mathrm{A}_n\big)$. Of course, we should take care that $\mu(I) \neq 0$. However, if $\mu(I) = 0$ we can simply ignore the atom I since it does not belong to the support of μ. We now immediately see that our partition function $Z_n(\beta)$ is exactly the expectation appearing in hypothesis i) of Theorem 10. Therefore, if the second hypothesis ii) of this Theorem is also satisfied, we get

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{\#\big\{I \in \mathrm{A}_n \bigm| \mu(I) \sim \mathrm{e}^{n\alpha}\big\}}{\mathrm{e}^{n\gamma}} = -\varphi(\alpha) \ .$$

Here we assume a little more than the conclusion of Theorem 10, namely that instead of having information about those atoms I for which $\mu(I) > \mathrm{e}^{n\alpha}$, we have information for $\mu(I) \sim \mathrm{e}^{n\alpha}$. This follows easily if $\varphi(\alpha)$ is differentiable with non-zero derivative for this value of α.

From this result we can come back to the Hausdorff dimension of the sets E_δ. For this purpose, we will assume that there is a number $\varrho \in (0,1)$ such that all atoms of A_n are intervals of length ϱ^n (uniform partition). Therefore using definition (11), we need about $\mathrm{e}^{n\gamma}\mathrm{e}^{-n\varphi(\delta\log\varrho)}$ balls of radius ϱ^n to cover E_δ. In other words

$$\inf_{E_\delta\subset\cup_j B_{r_j}(x_j)\,,\,\sup_j r_j\le\varrho^n}\sum_j r_j^d\le\varrho^{nd}\mathrm{e}^{-n\varphi(\delta\log\varrho)}\mathrm{e}^{n\gamma}\ .$$

If $d>(\varphi(\delta\log\varrho)-\gamma)/\log\varrho$, the above quantity tends to zero when n tends to infinity and we conclude that $d_{\mathrm{H}}\big(E_\delta\big)\le(\varphi(\delta\log\varrho)-\gamma)/\log\varrho$. Under some bounded distortion properties, one can prove that this is also a lower bound (see [9] and [1]).

As an easy example consider on the Cantor set K the measure μ defined for $0<p<1$ (and $q=1-p$) by

$$\mu\big(I_{\eta_1,\dots,\eta_n}\big)=p^{\sum_{j=1}^n\eta_j/2}q^{n-\sum_{j=1}^n\eta_j/2}\ .$$

When $p=q=1/2$ we get the Cantor measure defined above which has a trivial mono-fractal structure. From now on we assume $p\neq q$. Consider the sequence of partitions $\mathrm{A}_n=\big\{I_{\eta_1,\dots,\eta_n}\big\}$. It is easy to prove that the multifractal formalism applies to this measure using the large deviation results previously established. One gets since $\gamma=\log 2$

$$P(\beta)=\log\big(p^\beta+q^\beta\big)-\log 2\ ,$$

and assuming for example $p>q$ (the other case is left to the reader) it follows that

$$s=\frac{\log\big(\log p-\alpha\big)-\log\big(\alpha-\log q\big)}{\log q-\log p}\ ,$$

and one deduces immediately

$$\varphi(\alpha)=\frac{\big(\log p-\alpha\big)\log\big(\log p-\alpha\big)}{\log p-\log q}+\frac{\big(\alpha-\log q\big)\log\big(\alpha-\log q\big)}{\log p-\log q}$$

$$+\log 2-\log\big(\log p-\log q\big)\ .$$

Therefore, since $\varrho=1/3$ we get $\alpha=-\delta\log 3$ (see the definition (11)) and

$$d_{\mathrm{H}}\big(E_\delta\big)=-\frac{\big(\log_3 p+\delta\big)\log\big(\log_3 p+\delta\big)}{\log_3 p-\log_3 q}-\frac{\big(-\delta-\log_3 q\big)\log\big(-\delta-\log_3 q\big)}{\big(\log_3 p-\log_3 q\big)}$$

$$+\log_3\big(\log_3 p-\log_3 q\big)\ .$$

When we consider the graph of $d_{\mathrm{H}}\big(E_\delta\big)$ in figure 1, there are four particularly interesting points. First of all there are the two extreme points

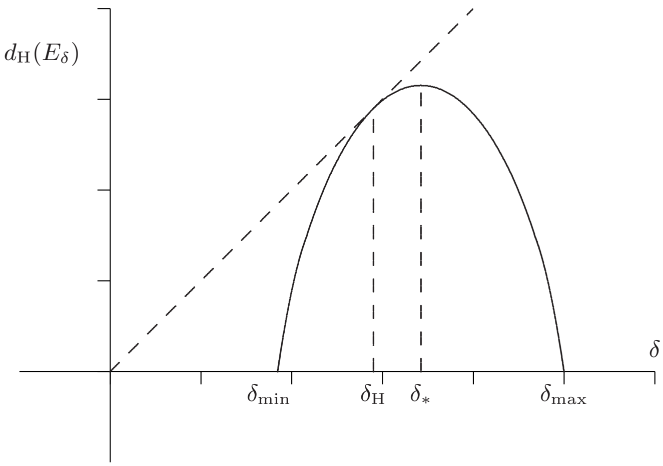


Fig. 1. The Hausdorff dimension $d_{\mathrm{H}}\big(E_\delta\big)$ as a function of δ for $p = 2/3$.

$\delta_{\min} = -\log_3 p$ and $\delta_{\max} = -\log_3 q$ where $d_{\mathrm{H}}\big(E_\delta\big)$ vanishes (recall that $p > q$). These correspond respectively to the largest and smallest measure of atoms of fixed size, namely for an atom I of A_n we have $q^n \le \mu(I) \le p^n$, and there is only one atom reaching each bound. The constraint $-\log_3 p \le \delta \le -\log_3 q$ can also be deduced from the formula

$$\alpha = \frac{p^s \log p + q^s \log q}{p^s + q^s} \;.$$

There is a unique maximum for $d_{\mathrm{H}}\big(E_\delta\big)$ at $\delta_* = -\big(\log_3 p + \log_3 q\big)/2$ which gives $d_{\mathrm{H}}\big(E_{\delta_*}\big) = \log 2/\log 3$, namely the Hausdorff dimension of the support K of μ. Note that it corresponds to the value of δ such that E_δ is covered by the largest number of atoms of A_n. However the total contribution of these atoms to the weight of μ is asymptotically negligible, namely

$$\sum_{I \in \mathrm{A}_n \;,\; \mu(I) \sim \mathrm{e}^{-n\delta_* \log 3}} \mu(I) \sim 2^n \mathrm{e}^{-n\delta_* \log 3} \mathrm{e}^{-n\varphi(-\delta_* \log 3)} \sim (2\sqrt{pq})^n$$

which tends to zero when n tends to infinity since for $p \neq 1/2$ one has $p(1-p) < 1/4$.

Finally there is the point $\delta_H = -p \log_3 p - q \log_3 q$ where the slope is equal to one. Since $\varphi'(\alpha) = s$, this gives $s = 1$. Note also that by the normalization of the probability measure μ, we have for each δ

$$\sum_{I \in \mathrm{A}_n \;,\; \mu(I) \sim \mathrm{e}^{-n\delta \log 3}} \mu(I) \sim 2^n \mathrm{e}^{-n\delta \log 3} \mathrm{e}^{-n\varphi(-\delta \log 3)} \sim 2^n \mathrm{e}^{nP\big(\beta_\delta\big)} \le 1 \;,$$

where $\beta_\delta = \varphi'\big(-\delta\log 3\big)$. In particular we obtain immediately

$$d_{\mathrm{H}}\big(E_\delta\big) \le \delta\ .$$

We can have equality only if $P\big(\beta_\delta\big) = \log 2$, and this must be a tangency. This equation leads immediately to $p^\beta + q^\beta = 1$, and since $p + q = 1$ we deduce $\beta = 1$. From $\alpha = P'(\beta)$ we deduce $\alpha = p\log p + q\log q$, and an easy calculation leads indeed to $\varphi'\big(-\delta_H\log 3\big) = 1$. The number $d_{\mathrm{H}}\big(E_{\delta_H}\big) = \delta_H$ is called the dimension of the measure μ. This quantity is defined in general as the infimum of the dimensions of sets of measure one. We see here that because of the multifractal behaviour, this dimension is strictly smaller than the dimension of the support (since $p \neq q$). We also see again a phenomenon reminiscent of Statistical Mechanics. If we consider atoms of a given size only a very small percentage contribute to the total mass of the measure. Moreover, these atoms have about the same measure (roughly equal to the inverse of their number).

For $p = q = 1/2$, the curve collapses to one point and we recover the monofractal Cantor measure.

We refer to the literature (see for example [9] and [1]) for other examples and extensions.

We also mention that although most of the sets E_δ have zero μ measure, they may be of positive (and even full) measure for another measure. This is why in certain situations they may become important and in fact observable.

References

1. L. Barreira, Y. Pesin, J. Schmeling. On a general concept of multifractality: multifractal spectra for dimensions, entropies, and Lyapunov exponents. Multifractal rigidity. Chaos 7:27-38 (1997).
2. A. Barron, O. Johnson. Fisher information inequality and the central limit theorem. *http://arXiv.org/abs/math/0111020*.
3. I. Berkes, E. Csáki. A universal result in almost sure central limit theory. Stochastic Process. Appl. 94:105-134 (2001).
4. R.N. Bhattacharya, R. Ranga Rao. *Normal approximations and asymptotic expansion*. Krieger, Melbourne Fla. 1986.
5. P. Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, New York 1968.
6. A. Borovkov. Boundary-value problems, the invariance principle, and large deviations. Russian Math. Surveys 38:259-290 (1983).
7. A. Borovkov. *Statistique Mathématique*. Editions Mir, Moscou 1987.
8. P. Collet. Ergodic properties of maps of the interval. In *Dynamical Systems*. R. Bamon, J.-M. Gambaudo & S. Martínez éditeurs, Hermann, Paris 1996.
9. P. Collet, J. Lebowitz, A. Porzio. The dimension spectrum of some dynamical systems. J. Statist. Phys. 47:609-644 (1987).
10. A. Dembo, O. Zeitouni. *Large Deviation Techniques and Applications*. Jones and Bartlett, Boston 1993.

11. R.S. Ellis. *Entropy, Large Deviations, and Statistical Mechanics*. Springer, Berlin 1985.
12. K.J. Falconer. *The geometry of fractal sets*. Cambridge Tracts in Mathematics, 85. Cambridge University Press, Cambridge, 1986.
13. W. Feller. *An introduction to Probability Theory and its Applications I, II*. John Wiley & Sons, New York, 1966.
14. I. Guikhman, A. Skorokhod. *Introduction à la Théorie des Processus Aléatoires*. Editions Mir, Moscou 1980.
15. P. Halmos. *Measure Theory*. D. Van Nostrand Company, Inc., New York, N. Y., 1950.
16. F. Hofbauer, G. Keller. Ergodic properties of invariant measures for piecewise monotonic transformations. Math. Zeit. 180:119-140 (1982).
17. E.T. Jaynes. *Probability Theory, The Logic of Science*. Cambridge University Press, Cambridge 2004.
18. G. Jona-Lasinio. Renormalization group and probability theory. Physics Report 352:439-458 (2001).
19. A. Kachurovskii. The rate of convergence in ergodic theorems. Russian Math. Survey 51:73-124 (1996).
20. J.-P. Kahane. *Some random series of functions*. Cambridge University press, Cambridge 1985.
21. A.I. Khinchin. *Mathematical Foundations of Statistical Mechanics*. Dover, New York 1949.
22. O.E. Lanford III. Entropy and equilibrium states in classical statistical mechanics. In *Statistical Mechanics and Mathematical Problems*. A. Lenard editor, Lecture Notes in Physics 20, Springer, Berlin 1973.
23. P. Mattila. *Geometry of sets and measures in Euclidean spaces. Fractals and rectifiability*. Cambridge Studies in Advanced Mathematics, 44. Cambridge University Press, Cambridge, 1995.
24. M. Métivier. *Notions fondamentales de la théorie des probabilités*. Dunod, Paris 1968.
25. J. Neveu. *Calcul des Probabilités*. Masson, Paris 1970.
26. P. Ney. Notes on dominating points and large deviations. Resenhas 4:79-91 (1999).
27. V.V. Petrov. *Limit Theorems of Probability Theory. Sequences of independent random variables*. Clarendon Press, Oxford 1995.
28. W. Philipp, W. Stout. *Almost sure invariance principles for partial sums of weakly dependent random variables*. Memoirs of the AMS, 161:1975.
29. D. Plachky, J. Steinebach. A theorem about probabilities of large deviations with an application to queuing theory. Periodica Mathematica 6:343-345 (1975).
30. E. Rio. *Théorie asymptotique des processus aléatoires faiblement dépendants*. Springer, Berlin 2000.
31. L. Schwartz. *Cours d'Analyse de l'Ecole Polytechnique*. Hermann, Paris 1967.
32. C. Stein. *Approximate Computations of Expectations*. IMS, Hayward Cal. 1986.
33. W. Stout. *Almost Sure Convergence*.Academic Press, New York 1974.
34. M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. Inst. Hautes Études Sci. Publ. Math. 81:73-205 (1995).
35. W. van der Vaart, J. Wellner. *Weak convergence and empirical processes : with applications to statistics*. Springer, Berlin 1996.

36. H. Ventsel. *Théorie des Probabilités*. Editions Mir, Moscou 1973.
37. N. Wax. *S*elected Papers on Noise and Stochastic Processes. Dover, New York 1954.
38. L.S. Young. Recurrence times and rates of mixing. Israel J. Math. 110:153-188 (1999).

Distribution of the Roots of Certain Random Real Polynomials

Bénédicte Dujardin

Département Artémis, Observatoire de la Côte d'Azur,
BP 4229, 06304 Nice Cedex 4, France.
dujardin@obs-nice.fr

1 Introduction

Random polynomials appear naturally in different fields of physics, like quantum chaotic dynamics, where one has to study the statistical properties of wavefunctions of chaotic systems and the distribution of their zeros [2, 11]. Our personal interest lies rather in their connection with noisy data analysis, especially in the context of the linear parametric modelization of random processes [5, 3, 16] and the problem of the resonance recognition, i.e. the identification of the poles of rational estimators of the power spectrum, computed from a measured sample of the signal.

In this contribution we address a probabilistic question concerning the real and complex roots of certain classes of random polynomials, the coefficients of which are random real numbers. The roots of such polynomials are random variables, real or complex conjugates, and one is interested in the mathematical expectation of their distribution in the complex plane, according to the degree of the polynomial and the statistics of its coefficients.

Because of mathematical simplicity, we study in section 2 the statistics of the real roots of polynomials with real random Gaussian coefficients. This material is taken from the historical papers by Kac [8, 9] and subsequent works [6, 4, 12]. In section 3 are introduced several directions of generalization; we investigate the statistics of the roots in the whole complex plane, and introduce the notion of generalized monic polynomials. We just give an outline of the derivation of the density of complex roots by recalling the passage from the real case to the complex one, the proof of which can be found in [13]. We use this result in order to understand and characterize the behavior of the roots in the two extreme cases, homogeneous random polynomials on the one hand, monic polynomials with weak disorder on the other hand. We briefly look at the particular class of self-inversive random polynomials [2], whose roots have an interesting behavior on the unit circle; the case of random complex coefficients is just mentioned in the conclusion.

J.-D. Fournier et al. (Eds.): Harm. Analysis and Ratio. Approx., LNCIS 327, pp. 129–143, 2006.

2 Real roots

The first problem about random polynomials is the question of the average number of real roots of a polynomial of degree n and was solved by Kac in the 40's [8] in the simple case of coefficients a_k independent identically distributed (i.i.d.) with Gaussian probability density function (pdf) $\mathrm{N}(0,1)$ of average zero and variance 1. Let

$$P_n(z) = \sum_{k=0}^{n} a_k z^k \tag{1}$$

be a random polynomial, and N_n the number of different real roots, called $t_k^{(n)}$, of P_n. In the following it is assumed that the probability to have multiple roots is negligible. N_n is the integral on the real axis of the counting measure

$$\sigma_n(t) \equiv \sum_{k=0}^{N_n} \delta(t - t_k^{(n)}) = |P_n'(t)|\; \delta(P_n(t)), \tag{2}$$

the Jacobian being due to the change of independent variable in the Dirac distribution. $\sigma_n(t)$ is actually the exact density of the roots for a given realization, and one can calculate its mathematical expectation

$$\varrho_n(t) \equiv \mathbb{E}(\sigma_n(t)) = \int_{\mathbb{R}^2} \mathrm{d}P\; \mathrm{d}P'\; \mathrm{P}(P, P')\; |P'|\; \delta(P) \tag{3}$$

considering, for any fixed $t \in \mathbb{R}$, $P_n(t)$ and $P_n'(t)$ as two correlated random variables written P and P'. Let us now make the hypothesis that the a_k are i.i.d. $\mathrm{N}(0,1)$; since $P_n(t)$ and $P_n'(t)$ are linear combinations of the a_k, they are themselves Gaussian variables with zero mean and joint pdf

$$\mathrm{P}(P, P') = \frac{1}{2\pi\sqrt{\det(\boldsymbol{C})}} \exp\left\{ -\frac{1}{2}(P, P')\boldsymbol{C}^{-1} \begin{pmatrix} P \\ P' \end{pmatrix} \right\}, \tag{4}$$

$$\text{given the correlation matrix} \quad \boldsymbol{C} = \begin{pmatrix} \mathbb{E}(P^2) & \mathbb{E}(PP') \\ \mathbb{E}(PP') & \mathbb{E}(P'^2) \end{pmatrix}.$$

Equations (3) and (4) lead to the average density

$$\varrho_n(t) = \frac{1}{2\pi\sqrt{\Delta}} \int_{\mathbb{R}} \mathrm{d}P'\; |P'| \exp\left\{ -\frac{1}{2\Delta}\mathbb{E}(P^2)P'^2 \right\} = \frac{\sqrt{\Delta}}{\pi \mathbb{E}(P^2)}, \tag{5}$$
$$\text{where } \Delta = \mathbb{E}(P^2)\mathbb{E}(P'^2) - \mathbb{E}(PP')^2.$$

Replacing P and P' by their respective values $\sum_0^n a_k\, t^k$ and $\sum_0^n k\; a_k\; t^{k-1}$ we obtain the exact formula of the average density of real roots

$$\varrho_n(t) = \frac{1}{\pi}\left\{\frac{1}{(1-t^2)^2} - \frac{(n+1)^2 t^{2n}}{(1-t^{2n+2})^2}\right\}^{1/2}. \qquad (6)$$

Figure 1 shows the distribution of real roots for $n = 10$ and 100. The dotted line is the theoretical density $\varrho_n(t)$ given by (6); it has two peaks centered near ±1 and these peaks get narrower when the degree of the polynomial n increases. The black line is an histogram over 1000 realizations of the real roots of random polynomials as defined by (1), and we see that the simulations match the theory quite well.

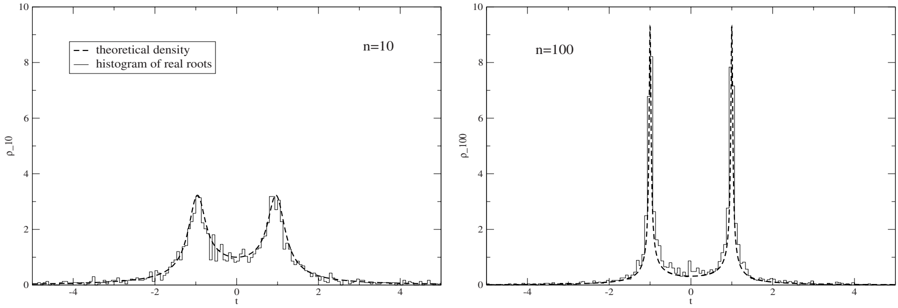


Fig. 1. Density of real roots and histograms over 1000 realizations for polynomials of degrees $n = 10$ and $n = 100$.

The number of real roots is then

$$N_n = \int_{\mathbb{R}} \mathrm{d}t\ \sigma_n(t). \qquad (7)$$

Since the integrals on t and on P, P' can be inverted, the average number of real roots is

$$\mathbb{E}(N_n) = \int_{\mathbb{R}} \mathrm{d}t\ \varrho_n(t) = \frac{1}{\pi}\int_{\mathbb{R}} \mathrm{d}t\ \left\{\frac{1}{(1-t^2)^2} - \frac{(n+1)^2 t^{2n}}{(1-t^{2n+2})^2}\right\}^{1/2}. \qquad (8)$$

This integral is bounded by

$$\frac{2}{\pi}\{\ln n + \ln(2 - \frac{1}{n})\} \ \le\ \mathbb{E}(N_n) \le\ \frac{2}{\pi}\{\ln n + \ln 2 + 4\sqrt{3}\} \quad \forall n \in \mathbb{N}, \qquad (9)$$

so for large degrees the main term of $\mathbb{E}(N_n)$ varies like $\frac{2}{\pi}\ln n$, as illustrated in Fig. 2 with numerical simulations.

Several other works have been carried out on this problem, relaxing the hypothesis of independence or gaussianity of the coefficients a_k. Littlewood and Offord [12] worked with uniform and bimodal distributions of the a_k, and proved that for large degrees the order of magnitude of $\mathbb{E}(N_n)$ kept on growing like $\frac{2}{\pi}\ln n$.

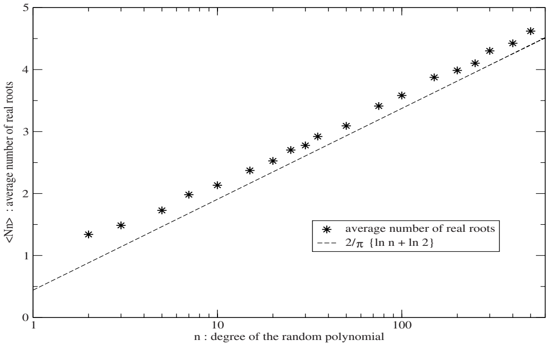


Fig. 2. Average number of real roots of P_n over 1000 realizations, as a function of the degree of the polynomial n.

Edelman and Kostlan [4] developed another method based on geometrical considerations that led them to Kac's formula (6). Their method, just like the calculation above, can be generalized to correlated and non-centered Gaussian coefficients, as long as the joint pdf $\mathrm{P}(P, P')$ is Gaussian. The average density is not so easy to write, mainly because of the emergence of an error function *erf*, but the asymptotic limit for $\mathbb{E}(N_n)$ remains generally valid. Ibragimov and Maslova [6] extended this result to any i.i.d. centered variables whose probability laws belong to the basin of attraction of the normal law according to the central limit theorem.

The logarithmic growth of the average number of real roots as a function of the degree is thus a common feature of random real polynomials. Although its Lebesgue measure is zero, the real axis of the complex plane, which is a symmetry axis for the set of the roots, is a singularity for the distribution of roots.

3 Complex roots

3.1 Complex roots

We are now interested in the average distribution in the whole complex plane of the roots of the random polynomial P_n, at least in the limit $n \gg 1$. The same argument as in section 2 can be applied so we get an integral formula for the roots density, with slight modifications. The counting measure in the plane is

$$\sigma_n(z) = \sum_k \delta^{(2)}(z - z_k^{(n)}) = |P_n'(z)|^2 \; \delta^{(2)}(P_n(z)), \quad z \in \mathbb{C}. \tag{10}$$

The change between formulæ (2) and (10) is due to the transition from a 1-dimensional space to a 2-dimensional space, and to the holomorphy of P_n. The notation is the same as in section 2, P and P' being the random variables that are the polynomial and its derivative at point z. Those quantities are complex, since z is complex, so we actually have 4 random variables, the real and imaginary parts of P and P', that we must take in account when writing the average density of roots

$$\varrho_n(z) = \int \mathrm{d}^2 P' \ |P'|^2 \ \mathrm{P}(0, 0, \mathrm{Re}(P'), \mathrm{Im}(P')). \tag{11}$$

The average number of roots in a domain $\Omega \subset \mathbb{C}$ is the integral

$$\mathbb{E}(N_n(\Omega)) = \int_\Omega \mathrm{d}^2 z \ \varrho_n(z). \tag{12}$$

3.2 Generalized monic polynomials

Motivated by the remarkable properties of the roots of Szegö polynomials, Mezincescu et al. [13] became interested in the case of monic polynomials, and more widely, in the class of random generalized monic polynomials, defined as

$$P_n(z) = \Phi(z) + \sum_{k=0}^{n} a_k f_k(z), \quad z \in \mathbb{C}. \tag{13}$$

Φ, $f_0, \ldots, f_n$ are holomorphic functions, and the a_k are the real random coefficients. We will later focus on two cases of particular interest: taking $\Phi = 0$ and $f_k = z^k$ returns an homogeneous random polynomial as studied in section 2; taking $\Phi(z)$ as a polynomial of degree n and $f_k = z^k$, we get a monic – in the classical sense – polynomial.

With the hypothesis that the joint pdf of P and P' is Gaussian, computing the integral over P' becomes possible and leads to the density of complex roots. Let us suppose that the a_k are i.i.d. $\mathrm{N}(0, 1)$; this hypothesis is not restrictive, since a judicious choice of Φ and f_k allows one to reduce systematically the coefficients to zero-mean and same variance random variables.

Working with 4 random variables instead of 2 makes the calculations more mathematically intensive but does not change the principle, so we just give the final result. Let us first introduce some notations adapted to the problem [13, 14, 16]. Let $\boldsymbol{v}$ and $\boldsymbol{w}$ be two complex vectors of dimension n, $(\boldsymbol{v}, \boldsymbol{w})$ is defined as the 2×2 matrix

$$(\boldsymbol{v}, \boldsymbol{w}) \equiv \begin{pmatrix} \mathrm{Re}(\boldsymbol{v}) \cdot \mathrm{Re}(\boldsymbol{w}) & \mathrm{Re}(\boldsymbol{v}) \cdot \mathrm{Im}(\boldsymbol{w}) \\ \mathrm{Im}(\boldsymbol{v}) \cdot \mathrm{Re}(\boldsymbol{w}) & \mathrm{Im}(\boldsymbol{v}) \cdot \mathrm{Im}(\boldsymbol{w}) \end{pmatrix} \tag{14}$$

The transposed column vector of $(f_0(z), \ldots, f_n(z))$ and its derivative are written $\boldsymbol{f}$ and $\boldsymbol{f'} \cdot \Phi$ is considered as a 2-dimensional vector $(\mathrm{Re}(\Phi(z)), \mathrm{Im}(\Phi(z)))$ of derivative Φ'. With those notations, the mathematical expectation of the counting measure of the complex roots is, at the points where $\det(\boldsymbol{f}, \boldsymbol{f}) \neq 0$,

$$\mathbb{E}(\varrho_n) = \frac{1}{2\pi}\frac{1}{\sqrt{\det(\boldsymbol{f},\boldsymbol{f})}}\ \exp^{-\frac{1}{2}\Phi\cdot(\boldsymbol{f},\boldsymbol{f})^{-1}\Phi} \times \tag{15}$$
$$\times\left\{\mathrm{Tr}[(\boldsymbol{f}',\boldsymbol{f}') - (\boldsymbol{f}',\boldsymbol{f})(\boldsymbol{f},\boldsymbol{f})^{-1}(\boldsymbol{f},\boldsymbol{f}')] + \|\Phi' - (\boldsymbol{f}',\boldsymbol{f})(\boldsymbol{f},\boldsymbol{f})^{-1}\Phi\|^2\right\}$$

3.3 Strong disorder limit: classical homogeneous polynomial

When $\Phi = 0$ and $f_k = z^k$, the polynomial is an homogeneous random polynomial of degree n. The average pdf of its complex roots can be explicitly computed for any $z = r\mathrm{e}^{\mathrm{i}\theta} \in \mathbb{C} \setminus \mathbb{R}$ according to

$$\frac{1}{2\pi}\frac{1}{\sqrt{\det(\boldsymbol{f},\boldsymbol{f})}}\,\mathrm{Tr}[(\boldsymbol{f}',\boldsymbol{f}') - (\boldsymbol{f}',\boldsymbol{f})(\boldsymbol{f},\boldsymbol{f})^{-1}(\boldsymbol{f},\boldsymbol{f}')]. \tag{16}$$

This expression is a function of r, θ and n, plotted in Fig. 3 for $n = 10$ and 100. As one can see, the area of the plane where this function is not negligible is an annulus around the unit circle that becomes narrower when the degree n increases. This result is not valid on the real axis, where $\det(\boldsymbol{f},\boldsymbol{f}) = 0$; function (16) is equal to zero, as it appears on the plot, but we have already seen in section 2 that the global measure has a singular component on the real axis given by (6).

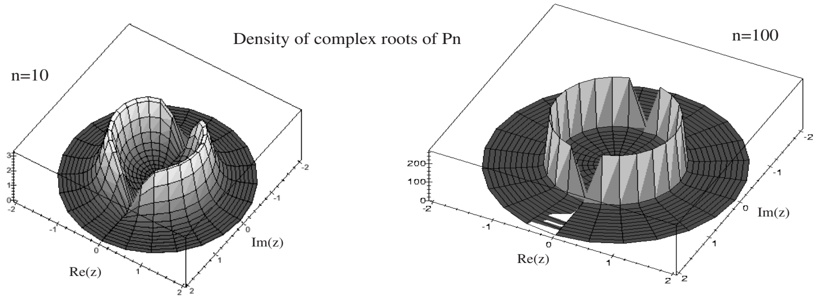


Fig. 3. 3-dimensional representation of the density of complex roots of an homogeneous random polynomial of degree n, for $n = 10$ (left) and 100 (right).

In the domain of the plane defined by

$$\left|\frac{1}{\ln r}\ \frac{1 - r^{2n+2}}{1 + r^{2n+2}}\ \sin\theta\right| \gg 1, \quad z = r\mathrm{e}^{\mathrm{i}\theta}, \tag{17}$$

i.e. close to the unit circle and far enough from the real axis, as shown in Fig. 4 for $n = 10$, and which corresponds to the interesting area of strong density, the average density can be approximated by

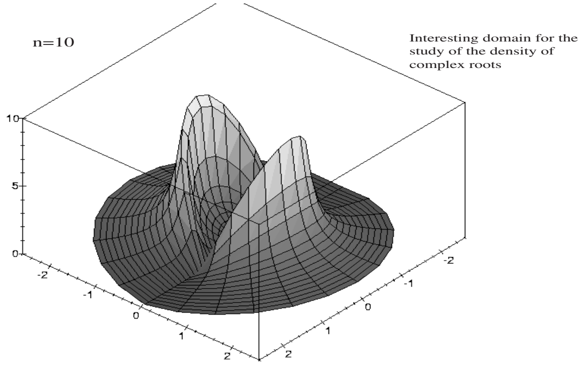


Fig. 4. 3-dimensional representation of the function (17) for $n = 10$.

$$\varrho_n(r\mathrm{e}^{\mathrm{i}\theta}) \simeq \frac{1}{\pi}\left\{\frac{1}{(\ln r^2)^2} - \frac{(n+1)^2 r^{2n+2}}{(1-r^{2n+2})^2}\right\}. \tag{18}$$

With this approximation, $\varrho_n(z)$ is a function of the radial variable r only and is independent of the angular variable θ, which implies a certain uniformity of the angular distribution of the roots. The curve, plotted in Fig. 5 for $n = 10$ and 100, has a peak for $r = 1$.

This behavior is characterized by two asymptotic results in the limit $n \gg 1$, still valid in the case of an α-stable distribution of the coefficients [7]. The first theorem concerns the fraction of roots in a disc of radius R

$$\frac{1}{n}\mathbb{E}(N_n(\mathrm{D}(0,R))) \overset{n\to\infty}{\longrightarrow} 0 \quad \text{in probability} \quad \forall R < 1. \tag{19}$$

Since the a_k are i.i.d., the statistics of the distribution of complex roots is invariant under the transformation $z \mapsto z^{-1}$, and (19) implies that most of the roots are present in a neighborhood of the unit circle. The other result concerns the fraction of roots in an angular sector $[\alpha, \beta]$

$$\frac{1}{n}\mathbb{E}(N_n(\alpha,\beta)) \overset{n\to\infty}{\longrightarrow} \frac{|\alpha-\beta|}{2\pi} \quad \text{in probability} \quad \forall\ [\alpha,\beta] \subset\ ]0,\pi[. \tag{20}$$

Because of the symmetry with regards to the real axis, and apart from is singularity, formula (20) implies an angular uniform distribution. In Fig. 5 are plotted, on the left, the 10000 roots of 1000 random polynomials of degree 10. We observe the strong concentration of points around the unit circle, and the singularity of the real axis. On the right are plotted histograms of the moduli of the complex (non-real) roots for $n = 10$ and 100, and with dotted lines the asymptotic curves given by (18); for the angular distribution, see Fig. 10.

Let us now call Λ the order of magnitude of the parameters of the deterministic term Φ. In the limit $\Lambda \ll 1$, i.e. for a strong disorder, the governing

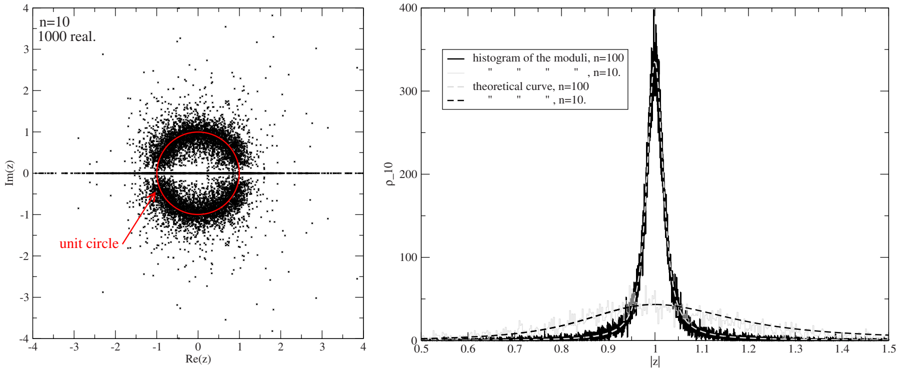


Fig. 5. On the left, the location in the complex plane of the roots of an homogeneous random polynomial of degree 10, for 1000 realizations. On the right, the radial distribution of complex roots and histograms of the moduli for $n = 10$ and 100.

term of P_n is its random part; the resulting density of roots is similar to the density for the homogeneous polynomial, and this behavior remains true as long as $\Lambda = \mathrm{O}(1)$.

3.4 Weak disorder limit: monic polynomials

In the weak disorder limit $\Lambda \gg 1$, P_n is dominated by $\Lambda\Phi(z)$; the main contributions to the density of roots (given by formula (15)) come from

$$\frac{\Lambda^2}{\sqrt{\det(\boldsymbol{f},\boldsymbol{f})}} \|\Phi' - (\boldsymbol{f}',\boldsymbol{f})(\boldsymbol{f},\boldsymbol{f})^{-1}\Phi\|^2 \times \exp\Big\{-\frac{1}{2}\Lambda^2\,\Phi\cdot(\boldsymbol{f},\boldsymbol{f})^{-1}\Phi\Big\}. \tag{21}$$

Let Φ be a polynomial of degree n, and z_0 a root of Φ of multiplicity M. In a neighborhood of z_0, $\Phi(z)$ can be written as $\Phi_0^{(M)}(z-z_0)^M + \mathrm{O}((z-z_0)^{M+1})$. The first function grows then like $\Lambda^2 M^2 \|z-z_0\|^{2M-2}$, when the other one is a decreasing exponential of the shape $\exp\{-\Lambda^2 C\|z-z_0\|^{2M}\}$; C is a positive quantity depending on z and z_0, of order 0 in $\|z-z_0\|$.

If the root is simple, for $M = 1$, the result is simply Gaussian, and the density has a peak centered on z_0, with a width of order Λ^{-1}.

If $M > 1$, the areas of strong density result from a balance between the positive power of $\|z - z_0\|$ and the decreasing exponential $\exp\{-\Lambda^2 C\|z-z_0\|^{2M}\}$. The average modulus of $(z-z_0)$ is not zero but is such that $\Lambda^2 C\|z-z_0\|^{2M} = \mathrm{O}(1)$, as illustrated in Fig. 6. The roots are then located in an annulus centered on z_0, with radius of order $C^{-1/(2M)}\Lambda^{-1/M}$.

A rough estimate of the integral of the density in a domain surrounding z_0 shows that the relative weight of the annulus circling a root of multiplicity M, with regards to the weight of the peak corresponding to a simple root,

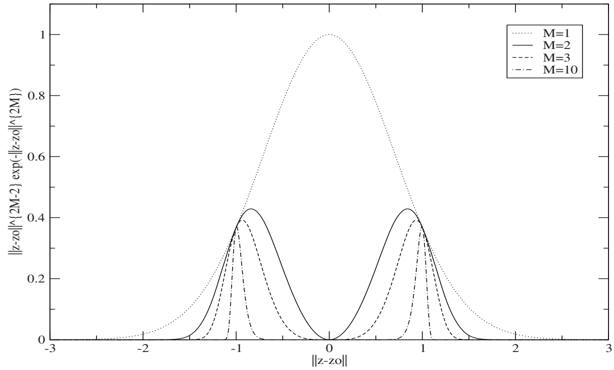


Fig. 6. Function $\|z-z_0\|^{2M-2}\exp\{-\|z-z_0\|^{2M}\}$ for different values of M.

is of order M. In other words, each peak in density centered on a root of multiplicity M does actually contain the equivalent of M roots.

Furthermore, the dependence of C with regards to the argument of $(z-z_0)$ causes some directions to possibly be more favorable. This phenomenon is strongly related to the expression of Φ, so nothing more can be said at this level of generality, but we will observe the emergence of an angular structure with the example $\Phi(z)=z^n$.

In Fig. 7 are plotted the 10 roots of the random polynomial

$$P_{10}(z) = 20\ (z^2+4)\ (z+\frac{1}{2}-\frac{\mathrm{i}}{2})(z+\frac{1}{2}+\frac{\mathrm{i}}{2})\ (z-1-\mathrm{i})^3(z-1+\mathrm{i})^3+\sum_0^{10} a_k z^k \tag{22}$$

for 500 realizations. We observe the presence of four (very) sharp peaks of density around the four simple roots ±2i and $-0.5\pm0.5\mathrm{i}$. The 2×3 remaining roots are located on two circles surrounding the roots of order $M=3$, $1\pm\mathrm{i}$.

Let us now consider the particular case of a root of multiplicity n at the origin by taking $\Phi(z)=z^n$. The density has then the shape

$$\Lambda^2 r^{2n-2}\ \exp\Big\{-\frac{1}{2}\ \Lambda^2 r^{2n-2}\ \frac{\sin^2 n\theta+\mathrm{O}(r^2)}{\sin^2\theta+\mathrm{O}(r^2)}\Big\},\qquad z=r\mathrm{e}^{\mathrm{i}\theta}\in\mathbb{R}\setminus\mathbb{C}\ ; \tag{23}$$

it is of order 1 as long as $\Lambda^2 r^{2n-2}\sin^2 n\theta\ \lesssim 1$, which means on a circle, the radius of which varies like $\Lambda^{-1/(n-1)}$, more particularly in the directions $\theta=k\pi/n$, for $k=0,\dots,2n-1$. Outside of those areas, the density of roots is negligible. This result can be recovered using dimensional analysis. Let us introduce the rescaled variable $y\equiv\Lambda^{1/n}|a_0|z$. The roots of P_n are given by the roots of the new polynomial

$$\tilde{P}_n(y) = y^n+\sum_{k=1}^{n-1} a_k\ |a_0|^{\frac{k}{n}-1}\ \Lambda^{-\frac{k}{n}}\ y^k+\mathrm{sgn}(a_0). \tag{24}$$

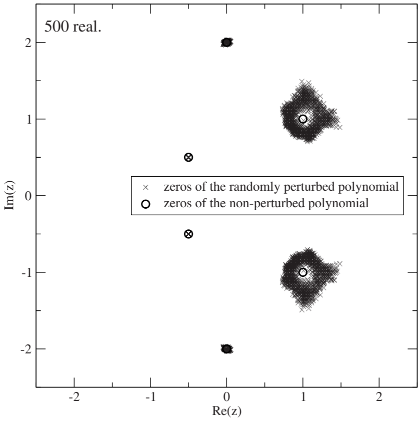


Fig. 7. Positions of the roots of 500 realizations of a random polynomial of degree 10. The deterministic part Φ has 4 simple roots in $\pm 2\mathrm{i}$, $-0.5 \pm 0.5\mathrm{i}$, and 2 roots of multiplicity 3 in $1 \pm \mathrm{i}$, indicated by a circle, and is of order $\Lambda = 20$.

Since $\Lambda \gg 1$ we neglect all the negative powers of Λ. The roots of $\tilde{P}_n(y)$ are then the n^{th} roots of ± 1, depending on the sign of the random variable a_0, and the roots of $P_n(z)$ are located on a circle of average radius

$$\Lambda^{-1/n}\mathbb{E}(a_0^{-1/n}) = \Lambda^{-1/n}\ 2^{1/2n}\ \frac{1}{\sqrt{\pi}}\ \Gamma(\frac{n+1}{2n}), \tag{25}$$

$$\text{where } \Gamma(t) = \int_0^\infty \mathrm{d}z\ z^{t-1}e^{-z} \text{ is the Euler function,}$$

at the favored angles $2k\pi/n,\ k = 0, \dots, n-1$, if $a_0 \geq 0$, and $(2k+1)\pi/n$ if $a_0 \leq 0$, which happens with equal probability since a_0 is $\mathrm{N}(0,1)$.

Figures 8, 9 and 10 demonstrate this particular behavior of the distribution of the roots of monic polynomials. The roots form what is called a “quasi-crystal” [1, 5], with a specific angular structure, located at a distance of the origin that depends on the order of magnitude of the deterministic part and the degree n. On the left are located in the plane the roots of 1000 monic polynomials with $\Phi(z) = z^n$ for $n = 4$ (Fig. 8) and $n = 10$ (Fig. 9). On the right of Fig. 8 is plotted a function of the shape given by (23), for $n = 4$. We observe 8 peaks of density at the regular angular intervals $k\pi/4$, $k = 0, \dots, 7$.

The radial behavior is studied in Fig. 9. On the right are plotted histograms of the moduli of complex (non-real) roots of monic polynomials of degree $n = 10$ and 100. As n increases, the peak gets sharper and its location tends to 1.

The radial and angular behaviors of the density of roots for respectively homogeneous and monic random polynomials are characterized and compared in Fig. 10. On the left we observe the evolution of the average modulus of complex roots as a function of the degree of the polynomial n. In the ho-

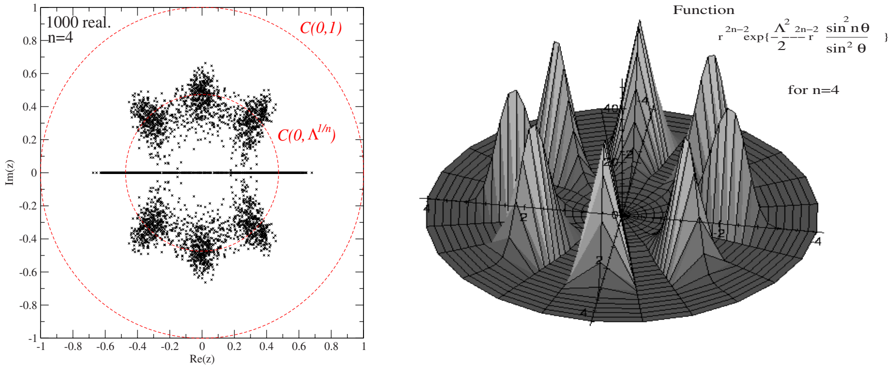


Fig. 8. Left, position in the plane of the roots of 1000 monic random polynomials with $\Phi(z) = z^4$ and $\Lambda = 20$. Right, a 3-dimensional plot of function (23) for $n = 4$.

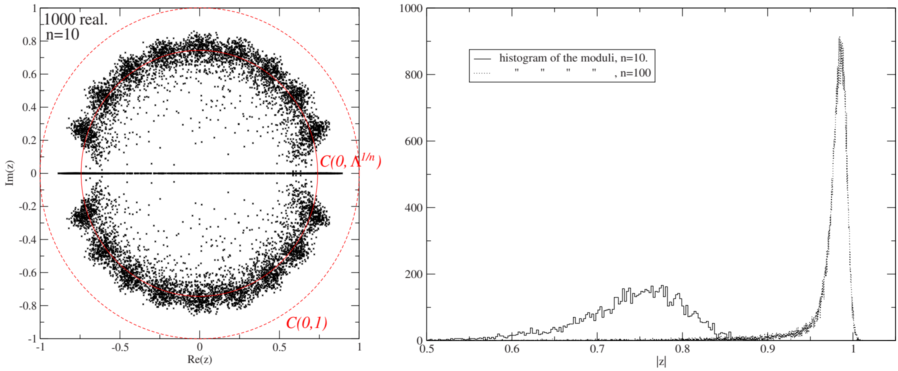


Fig. 9. On the left, position in the plane of the roots of 1000 monic random polynomials with $\Phi(z) = z^{10}$ and $\Lambda = 20$. On the right, histograms of the moduli of the non-real roots of monic polynomials for $n = 10$ and 100.

mogeneous case, the position of the peak is almost constant, close to 1; in the monic case, we have an exponential law of the inverse of n, corresponding to the order of magnitude $\Lambda^{-1/n}$. On the right, we study the angular distribution with histograms of the arguments of the complex (non real) roots. We observe the appearance of an angular structure in the monic case, while the angular distribution is quite uniform in the homogeneous case.

3.5 Self-inversive polynomials

Let us finally mention the particular case when the polynomial P_n has the self-inverse symmetry, which means that its coefficients have the reflective property $a_k = a_{n-k}$, $k = 0, \ldots, n$, with the consequence that the set of its roots is invariant through the transformation $z \mapsto z^{-1}$.

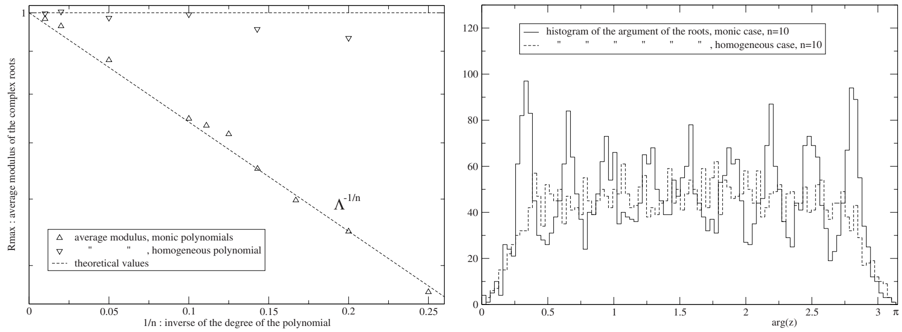


Fig. 10. Left, the average moduli of the complex roots of homogeneous and monic ($\Lambda = 20$) random polynomials, as a function of the inverse of the degree $1/n$, compared to their theoretical values (the constant 1 and $\Lambda^{-1/n}$). Right, histograms of the positive arguments of the complex roots of homogeneous and monic ($\Lambda = 20$, $\Phi = z^n$) random polynomials of degree $n = 10$.

Such polynomials have been studied by Bogomolny et al. [1, 2] and we just recall here their main result, which more generally holds for complex coefficients. The roots of a self-inversive random polynomial P_n have this remarkable property that they are not only concentrated in the neighborhood of the unit circle, but that a macroscopic fraction of them stands precisely on the circle, a fraction equal on average to

$$\frac{1}{n}\mathbb{E}(N_n(\{|z| = 1\})) = \sqrt{\frac{n-2}{3n}}, \qquad (26)$$

which tends to $1/\sqrt{3} \simeq 0.577$ in the limit of large degrees. In contrast to the fraction of real roots of real random polynomials, this fraction does not tend to zero as n tends to infinity.

This behavior is illustrated in Fig. 11. On the left are plotted the roots of 1000 self-inversive real polynomials of degree $n = 10$. Apart from the singularity of the real axis, we observe that a certain number of them is located exactly along the circle. On the right is plotted the average fraction of roots located exactly on the unit circle, as a function of the degree of the polynomial. As n tends to infinity, this fraction tends to the asymptotic value $1/\sqrt{3}$.

4 Conclusion

We have discussed two classes of random real polynomials, according to the order of magnitude of the random part with regards to the deterministic component. In the strong disorder case, the roots are concentrated around the unit circle. The second class concerns random monic polynomials in the

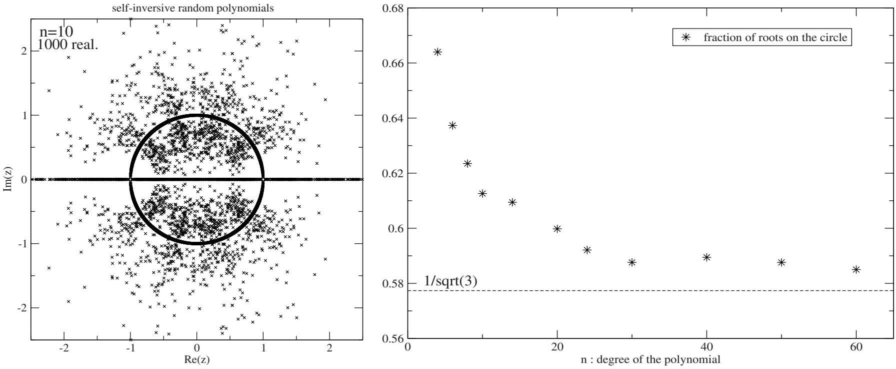


Fig. 11. Left, roots of 1000 self-inversive polynomials of degree $n = 10$. Right, evolution of the average fraction of roots on the unit circle as a function of the degree of the polynomial, tending towards to an asymptotic value of $1/\sqrt{3}$ (dashed line). The average number is computed over 1000 realizations.

presence of weak disorder. Their roots are located in the neighborhood of the roots of the non perturbed polynomial if those roots are simple, and in the case of multiple roots, on a quasi-crystal centered on the root.

We have seen two examples of polynomials whose roots have a certain symmetry, with regards to the real axis when the coefficients are real, or with regards to the unit circle in the self-inversive case. In both situations, the symmetry line “attracts” a certain fraction of the roots.

We have not studied here the case of random polynomials with complex coefficients. Yet, many studies have been carried out concerning this problem [1, 2, 15]. For high degrees, the roots are located in an annulus around the unit circle, with a uniform angular distribution, as one can see in Fig. 12 with the roots of 1000 random polynomials of degree 10 with complex coefficients whose real and imaginary parts are i.i.d. Gaussian variables. Histograms of the moduli and arguments complete this illustration.

The accumulation of the roots around the unit circle is related to the existence of a natural boundary of analyticity on this circle of the random series [10]

$$\sum_{k=0}^{\infty} a_k z^k. \tag{27}$$

The zeros of homogeneous random polynomials, i.e. partial sums of this series, are located in the neighborhood of the boundary [15].

It is possible to pursue the study of the statistical properties of the zeros of random polynomials with the determination of the k-point correlation functions $\varrho_k(z_1, \dots, z_k)$ using the same method [14]. Taking $k = 1$ returns the density of zeros, and $\varrho_2(z_1, z_2)$ characterizes the correlation between the roots.

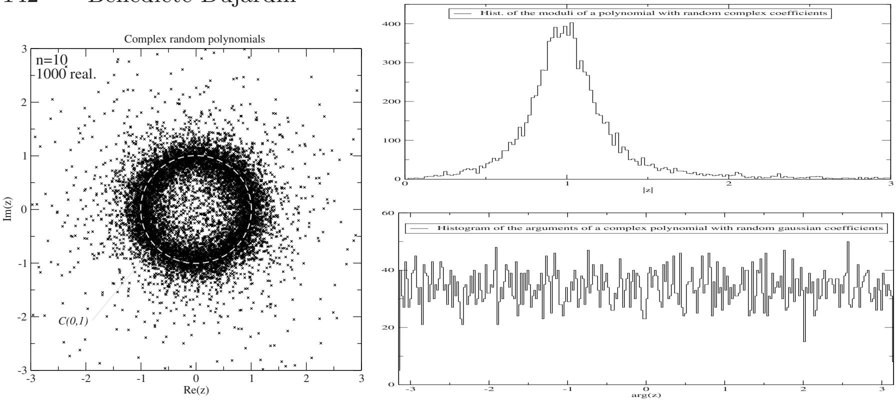


Fig. 12. Left, roots of 1000 complex random polynomials of degree $n = 10$. The real and imaginary parts of the coefficients are i.i.d. Gaussian $\mathrm{N}(0, 1)$ random variables. Right, histograms of the moduli and of the arguments of the roots.

Acknowledgements: I want first of all to thank Juliette and Jean-Daniel for directing the organization of the school and the edition of these proceedings. In addition to the questions some people asked me during the school, and the help of J.-D., Jonathan and David, I appreciated the constructive remarks and advice of my anonymous referee, and tried to take all of these suggestions into account. As usual, I am infinitely grateful to PB for his sympathetic ear and his olympian calm. I found that botching up this lecture was particularly profitable.

References

1. E. Bogomolny, O. Bohigas, P. Leboeuf, "Distribution of roots of random polynomials", *Phys. Rev. Lett.*, 68(18):2726-2729, 1992.
2. E. Bogomolny, O. Bohigas, P. Leboeuf, "Quantum Chaotic Dynamics and Random Polynomials", *J. Stat. Phys.*, 85:639-679, 1996.
3. B. Dujardin, J.-D. Fournier, "Coloured noisy data analysis using Padé approximants", *submitted.*
4. A. Edelman, E. Kostlan, "How many zeros of a random polynomial are real?", *Bull. Amer. Math. Soc.*, 32:1-37, 1995.
5. J.-D. Fournier, "Complex zeros of random Szegö polynomials", *Computational Methods and Function Theory*, pp. 203-223, 1997.
6. I.A. Ibragimov, N.B. Maslova, "On the expected number of real zeros of random polynomials I. Coefficient with zero means", *Theory Probab. Appl.* 16:228-248, 1971.
7. I.A. Ibragimov, O. Zeitouni, "On Roots of Random Polynomials", *Trans. Amer. Math. Soc.*, 349(6):2427-2441, 1997.
8. M. Kac, "On the average number of real roots of a random algebraic equation", *Bull. Amer. Math. Soc.*, 49:314-320, 1943.

9. M. Kac, “Probabilities & Related Topics in Physical Sciences”, *Lectures in Applied Mathematics Vol. 1A, Am. Math. Soc.*, 1959.
10. J.-P. Kahane, “Some random series of function”, *Cambridge University Press*, 1968.
11. P. Leboeuf, P. Shukla, “Universal fluctuations of zeros of chaotic wavefunctions”, *J. Phys. A. : Math. gen.* 29:4827-4835, 1996.
12. J.E. Littlewood, A.C. Offord, “On the number of real roots of a random algebraic equation”, *J. London Math. Soc.*, 13:288-295, 1938.
13. A. Mezincescu, D. Bessis, J.-D. Fournier, G. Mantica, F. Aaron, “Distribution of Roots of Random Real Generalized Polynomials”, *J. Stat. Phys.*, 86:675-705, 1997.
14. T. Prosen, “Exact statistics of complex zeros for Gaussian random polynomials with real coefficients”, *J. Phys. A. : Math. gen.,* 29:4417-4423, 1996.
15. B. Schiffman, S. Zelditch, “ Equilibrium distribution of zeros of random polynomials”, *Int. Math. Res. Not.*, 2003.
16. R. Schober, W.H. Gerstacker, “The zeros of random polynomials : Further results and applications”, *IEEE transactions on communications*, 50(6):892-896, 2002.

Rational Approximation and Noise

Maciej Pindor

Instytut Fizyki Teoretycznej,
Uniwersytet Warszawski ul.Hoża 69,
00-681 Warszawa, Poland.

1 Introduction

In the previous lecture I discussed (and advertised) a special type of rational approximation to functions of the complex variable – the one that can be constructed when the information on the function approximated is given in the form of coefficients of its power (favorably Taylor) expansion. The knowledge of the Taylor series coefficients specifies a function completely and, as we have seen, one can construct from a finite number of coefficients a rational function which (almost everywhere) approximates this function better and better when we take into account more and more coefficients. There are however other interesting sequences of rational approximants and I shall first say few words about them. They use the information on a behaviour of the function at several points. In either case, every application of this or other approximation scheme encounters in practice the additional difficulty: all the information we want and can use to construct an approximating rational function is biased by errors – we can know either expansion coefficients or function values with finite accuracy only. Consequences of this fact, fundamental in all practical applications, will be discussed in later sections.

2 Rational Interpolation

As we know an analytic function $f(z)$ can also be uniquely specified by an infinite number of its values at points contained in a compact set, on which the function is analytic. Construction of a sequence of rational functions having the same values as $f(z)$ on a given finite set of points is known as the rational interpolation problem. It is the classical problem of the numerical analysis and was studied long ago. My exposition will be partially based on [10]. Before we discuss the convergence of sequences of rational interpolants let us comment on the problem of their existence. Let there be a sequence of points in the complex plane $\{z_i\}_{i=0}^{N}$ (which we shall call *the nodes*) and a

J.-D. Fournier et al. (Eds.): Harm. Analysis and Ratio. Approx., LNCIS 327, pp. 145–156, 2006.

sequence of complex numbers $\{f_i\}_{i=0}^{N}$ – we call them henceforth *the data* – such that

$$f(z_i) = f_i \qquad i = 0, \dots, N\,. \tag{1}$$

We are looking for a rational function $r_{m,n}(z)$ of degrees m in the numerator and n in the denominator such that $r_{m,n}(z_i) = f_i$, $i = 0, \dots, N$. If we call the numerator and the denominator of $r_{m,n}$, $T_m(z)$ and $B_n(z)$ respectively, and treat their coefficients as unknowns we get the system of equations for these coefficients

$$\frac{T_m(z_i)}{B_n(z_i)} = f_i \qquad i = 0, \dots, N \tag{2}$$

and we can expect a unique solution if $m + n \leq N$.

These equations are nonlinear, but can be linearized to the form

$$T_m(z_i) = B_n(z_i) f_i \qquad i = 0, \dots, N \tag{3}$$

however (3) is not strictly equivalent to (2) – all solutions of the latter are solutions of the former, but not vice versa. The situation seems analogous to the one encountered in the construction of Padé approximants, but its origin is even easier to comprehend here. It is obvious that if (3) is satisfied and $B_n(z)$ does not vanish on any node, then we can divide the equation by $B_n(z_i)$ and (2) is also satisfied. We conclude that if a solution of (3) does not satisfy (2) then $B_n(z)$ must vanish on some subset of (say k) nodes. Then, however, $T_m(z)$ must also vanish there. Therefore both $T_m(z)$ and $B_n(z)$ contain a common factor – the polynomial of degree k vanishing on these nodes – let it be $w_k(z)$. In this case (3) looks as

$$\tilde{T}_{m-k}(z_i) w_k(z_i) = \tilde{B}_{n-k}(z_i) w_k(z_i) f_i \qquad i = 0, \dots, N \tag{4}$$

and it means that there exists a rational function of degrees $m - k$ and $n - k$ respectively, such that it interpolates our data on a subset of $N + 1 - k$ nodes. Vice versa, it is easy to see that if (4) is satisfied then there is no rational function of degrees m and n respectively and with relatively prime numerator and denominator that interpolates all our data – the k nodes at which $w_k(z)$ vanishes are called *unattainable*. All the details of the problem are studied in depth in [8]. We can say that the problem is the one of degeneracy and we shall not be concerned with it in the rest of the lecture.

Before I talk about convergence let me first point you out that rational interpolants we discussed above and Padé approximants are not entirely alien to each other. Actually, they are rather extreme cases of general rational interpolants. To see that we consider an *interpolation scheme*, i.e. a triangular matrix of *interpolation nodes* $a_{i,j} \in \overline{\mathbb{C}}$ defined as follows

$$\mathcal{A} := \begin{pmatrix} a_{00} & & & \\ \cdots & \cdots & & \\ a_{0n} & \cdots & a_{nn} & \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix} \tag{5}$$

Each row of the matrix $\mathcal{A}$

$$\mathcal{A}_n = (a_{0n}, \cdots, a_{nn}) \tag{6}$$

defines an *interpolation set* of $n+1$ nodes. We allow here some or all nodes in the set to be identical.

To each interpolation set $\mathcal{A}$ we assign the polynomial

$$w_n(z) = \prod_{x \in \mathcal{A}} (z - x) = \prod_{i=0}^{n} (z - a_{in}). \tag{7}$$

We say now that $r_{m,n}(z)$ is is the (generalized) rational interpolant of the function $f(z)$ on the set $\mathcal{A}_{m+n}$ (where the function is assumed to be analytical) if it is the rational function of degrees at most m in the numerator and at most n in the denominator, such that

$$\frac{f(x) - r_{m,n}(x)}{w_{m+n}(x)} \text{ is bounded at each } x \in \mathcal{A}_{m+n}. \tag{8}$$

$r_{m,n}(z)$ is also called Hermite type (sense) rational interpolant, or multi-point Padé approximant. This latter name can be understood if we observe that when all the points in the interpolation set are identical then $r_{m,n}$ is just $[m/n]$ Padé approximant. On the other side, if all of them are distinct we have the ordinary rational interpolant. In the intermediate cases $r_{m,n}$ and its derivatives $r_{m,n}^{(k)}$ are identical with f and its derivatives $f^{(k)}$ at $x \in \mathcal{A}_{m+n}$ up to an order corresponding to a number of occurrences of x in $\mathcal{A}_{m+n}$ – which are our *data* in this situation.

As is the case for the classical rational interpolant, after introducing the numerator and the denominator of $r_{m,n}$, P_m and Q_n, respectively, we can substitute (8) by the linearized version

$$\frac{Q_m(x)f(x) - P_n(x)}{w_{m+n}(x)} \text{ is bounded at each } x \in \mathcal{A}_{m+n}. \tag{9}$$

Again, not every rational interpolant with the numerator and the denominator satisfying (8) satisfies (9), but the latter always has a solution. If however there exists a pair of polynomials satisfying (9) and $Q_n(z) \neq 0$ on any of the points of $\mathcal{A}_{m+n}$, then the problem (8) is also soluble and the solution is

$$r_{mn,n}(x) = \frac{P_m(z)}{Q_n(z)}.$$

Many algebraic problems connected with existence of multipoint Padé approximants have been studied in [5] and it is known that "blocks" appearing in the table do not need to be of square shape.

A special intermediate case is the one called Two-Point Padé Approximants. In this case the interpolation set consists of only two distinct points:

zero and the point at infinity, appearing in $\mathcal{A}_{m+n}$ altogether $m+n+1$ times. This type of rational approximation appears sometimes in physics or technology when we are interested in the function (assumed or postulated to be analytical) of some variable having special meaning at zero and infinity and we know some number of coefficients of its expansion around these two points. For example the function can be the dielectric constant of the composite of two different materials and the variable, the ratio of their contents [11].

3 Convergence

The convergence problem is in many respect analogous to the one of Padé approximants, though we have here the additional dependence on the asymptotic distribution of the interpolation nodes a_{in} in the interpolation scheme $\mathcal{A}_{m+n}$. There is no place here to discuss it in detail, but we can summarize the results by saying that rational interpolants converge in capacity for holomorphic functions and, apart of the set of cuts they choose, also for functions with branchpoints, but depending on the localisation of the interpolation scheme and the set of branchpoints it can happen that in different areas of the complex plane the rational interpolants will converge to different branches of the function.

4 Rational Interpolation with Noisy Data

As I have mentioned in the introduction, practical application of rational interpolation encounters the serious obstacle in the fact that the data (function values and expansion coefficients) are always known with finite accuracy only. This poses the problem when they are supposed to be used to construct an approximation for an analytical function. The analytical function is the "stiff" object – any, even the smallest, modification of it at some place may result in an arbitrarily large change at a finite distance from the place at which we made the modification. Look at the simplest possible, even naive example: assume that we have two sets of values at nodes $\{z_i\}_{i=0}^{N}$ – $\{d_i\}_{i=0}^{N}$ and $\{d_i + \varepsilon/(1+z_i)\}_{i=0}^{N}$. They can differ arbitrarily little, but if the first set comes from a function $f(z)$, the second one comes from $f(z)+\varepsilon/(1+z)$ which differs from $f(z)$ arbitrarily much at $z=-1$ (assuming that $f(z)$ is regular there).

In the following I shall take for granted that the interpolation set is contained in the real line, that functions studied are real on the real line and that perturbations of data are also real. This assumptions are inessential in all algebraic considerations and I shall comment below when they influence presented results.

The problem has been observed when the first applications of Padé approximants in physics appeared. In physics Padé approximants have been used to "sum" so called perturbation series – rather to estimate the sum of the series from finite number (usually small, unfortunately) of coefficients. Calculation of those coefficients is generally a serious task, involving numerical calculation of multiple integrals, and usually physicists must accept substantial limitations in accuracy with which they can know such coefficients. Very soon it was found that varying these coefficients within limits of the accuracy with which they were known, resulted in "wild" variations of singularities of Padé approximants constructed from the coefficients. In this situation Marcel Froissart [4] made very simple, but highly enlightening numerical experiments. He took just the geometrical series and perturbed randomly coefficients of its power series in the following manner

$$1 + x + x^2 + \dots \quad \Rightarrow \quad 1 + \varepsilon r_0 + (1 + \varepsilon r_1)x + (1 + \varepsilon r_2)x^2 + \dots \tag{10}$$

with some small ε and random r_i's taken from same distribution. Obviously all Padé approximants to the series on the left are equal to [0/1], i.e. to the function itself. On the other hand, all (almost, except for the set of measure zero on the event space) Padé approximants to the series on the right are different and, if we concentrate first on the sequence $[n-1/n]$, they have $n-1$ different zeros and n different poles – both randomly distributed. At first this seems a catastrophe – independently of how small ε is, Padé approximants to the perturbed series seem to have nothing in common with the function represented by the original series! However, when one looks where zeros and poles of these Padé approximants are, an amazing phenomenon can be seen. Look at zeros and poles of [4/5] with some choice of (real, normally distributed) r_i's

ε	zeros	poles	
.00001	$-\ .57471 \pm .64809$i	$-\ .57472 \pm .64809$i	1.00000098
	$-.091740,\ 3.1348$	$-.091740,\ 3.1349$	
.01	$-.57034 \pm .64907$i	$-.57468 \pm .64812$i	1.00099
	$-.091958,\ 3.0223$	$.091958,\ 3.1384$	

(11)

First you see that there is a pole close to 1 – the place where the function represented by the original series has one. Next you see that all the other zeros and poles – which represent only *noise* – come in tight pairs. The smaller ε is, the tighter they are – quite natural, because we want that at $\varepsilon = 0$ we return to the original series! Of course you must remember that positions of all these zeros and poles are random and for any finite ε both the separation of the pairs as well as the distance of the "unpaired" pole from 1 can be arbitrarily large, but we expect that at $\varepsilon \to 0$ they will both vanish. You can see it clearly in the next example where I took different choice of $r_i's$

$$\begin{array}{cccc}
\varepsilon & \text{zeros} & \text{poles} & \\
.00001 & 395.688,\ \ -.55084, & -387.376,\ \ -.55084, & 1.000000097 \\
 & -.013502 \pm 1.48561\mathrm{i} & -.013471 \pm 1.48566\mathrm{i} & \\
.0001 & -490.299,\ \ -.55084, & -387.370,\ \ -.55084, & 1.00000097 \\
 & -.013776 \pm 1.48518\mathrm{i} & -.013471 \pm 1.48566\mathrm{i} & \\
.001 & 356.883,\ \ -.55083, & -387.311,\ \ -.55084, & 1.0000097 \\
 & -.016466 \pm 1.48084\mathrm{i} & -.013471 \pm 1.48566\mathrm{i} &
\end{array} \tag{12}$$

When ε grows, the pair at large negative x's separates so strongly that at $\varepsilon = 001$ there is no pair at all.

This phenomenon of "pairing" of noise induced zeros and poles would not be interesting at all if it manifested itself only for geometrical series, but it appeared to be universal and got the name of *Froissart phenomenon* and the pairs are known as *Froissart doublets.*

For other sequences of Padé approximants, when there is a "surplus" of zeros or poles, it appeared that these extra zeros or poles escape to infinity when $\varepsilon \to 0$.

Let us now see at Fig. 1 what happens when n grows – as you see Froissart doublets are distributed in a close vicinity of the unit circle! This phenomenon would be even more pronounced if we took n larger, on the other hand one would always find some doublets, even for large n at a "finite" distance from the circle – like those inside the circle.

To show you what happens if we perturb – in the same way as before – a series representing a function with branchpoints, let us consider the function

$$f(z) = \sqrt{z+1}\sqrt{2z+1} + \frac{2}{z-1}. \tag{13}$$

It has branchpoints at -1 and $-1/2$, the pole at 1, but also zeros at 1.604148754 and -1.391926826. Moreover it behaves like $\sqrt{2z}$ when $z \to \infty$. We know already that for this function we should expect approximants $[n+1/n]$ be the best suited. Below you have zeros and poles of [6/5] "exact" i.e. $\varepsilon = 0$ and also $\varepsilon = .001$ and $\varepsilon = .00001$. We clearly see that the "exact" approximants give zeros and the pole very close to zeros and the pole of the function while the remaining zeros and poles of the approximant simulate the cut $(-1, -1/2)$. It is also interesting to note that [6/5] behaves for $z \to \infty$ like $1.4142139z$ ($\sqrt{2} \approx 1.4142136$). For approximants to perturbed series we see that zeros and the pole of the function are reproduced much worse, but they are there. Behaviour at infinity is also perturbed, but makes sense. Finally we also see Froissart doublets.

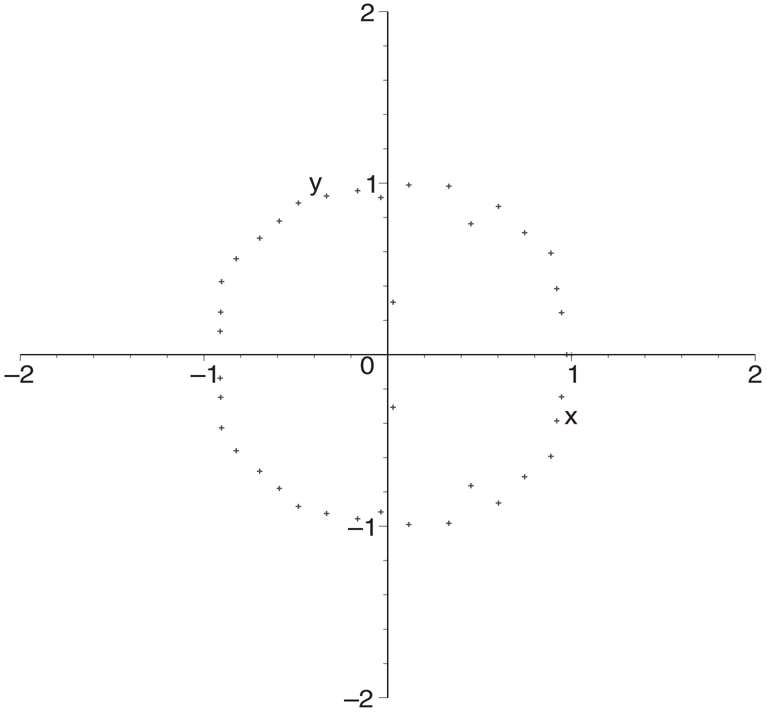


Fig. 1. Zeros and poles of [39/40] to perturbed geometrical series; $\varepsilon = .001$

ε	zeros	poles	p_6/q_5
0	$-1.3918660,\ -.89206841,$ $-.71781989,\ -.59123097,$ $-.52175271,\ 1.60414873$	$-\ .90621750,\ -.73283509,$ $-.59816026,\ -.52337768,$ $.999999998$	1.4142139
10^{-5}	$-1.391698,\ -.854144,$ $-.651406,\ -.535246,$ $-.1948449670,\ 1.6041319$	$-.873892,\ -.665806,$ $-.538656,\ -.1948449666,$ $.9999996$	1.414239
.001	$-1.38207,\ -.712257,$ $-.23399823 \pm .63417286i$ $-.545154,\ 1.602343$	$-.737541,\ -.550235,$ $-.23399385 \pm 63418746i,$ $.999951$	1.41749
.01	$-1.3608,\ -.587831$ $-.358697 \pm .433745i,$ $-.0517020759,\ 1.58744$	$-.602766,\ .999609,$ $-.358691 \pm .433605i,$ $-.051702075$	1.43968

(14)

But what happens with Froissart doublets when n grows? Look at Fig. 2. Now it seems that Froissart doublets are “attracted” by the circle of the radius

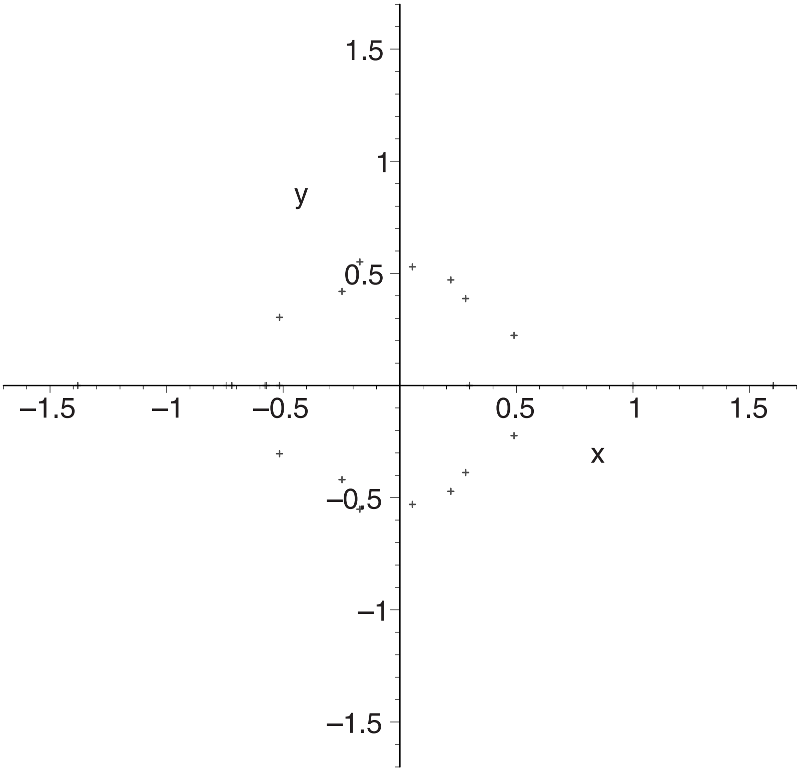


Fig. 2. Zeros and poles of [20/19] to perturbed series of the function discussed in the text; $\varepsilon = .001$

1/2 – yes, this is what takes place. But what is so special about 1/2? – it is the distance to the closest (with respect to the point of expansion) singularity.

Let us summarize our observations: Padé approximants to perturbed series exhibit the Froissart phenomenon, i.e. part of the zeros and poles form doublets that are tighter and tighter when perturbation becomes smaller. How large is this part also depends on a size of perturbation – when it is small most of the zeros and poles of the approximant are only slightly perturbed. When it grows more and more zeros and poles leave the neighborhood of "exact" zeros and poles and from Froissart doublets. For growing degrees of the numerator and of the denominator (more and more terms of the series used) and a "size" of the perturbation kept constant, Froissart doublets become attracted by the circle of a radius of the closest singularity.

Before I give you some explanation of this behaviour, let me show you what happens for other types rational interpolant. For this end I calculate 12 values of our function at equidistant interpolation nodes on $(2, 4)$ and calculate the (6/5) rational interpolant, first from "exact" function values and then for values perturbed in the way analogous to the way I perturbed coefficients of the series

$$f(z_i) \to f(z_i)(1 + \varepsilon r_i) \tag{15}$$

where z_i's belong to the interpolation set and r_i are independent random numbers from the same distribution.

In the table below you see that there appear here also "noise induced doublets", but they seem to be attracted by the interpolation interval! On the other hand, even for the largest ε this pairing of zeros and poles produced by noise is so strong that we can guess that there are somewhere two "real" zeros and one "real" pole, though their positions came out very badly. This example demonstrates the main characteristic feature of the result of rational interpolation made out of "noisy" data: Froissart doublets appear on the interpolation interval, or in the very close vicinity of it.

ε	zeros	poles	p_6/q_5
0	−1.3919268, 1.604148754,	.9999999999,	1.41421356
	−.940143, −.799964,	−.947363, −.814864,	
	−.649026, −.539497	−.660840, −.543340	
10^{-8}	−1.38538, 1.604150,	−1.00004	1.414230
	−.640039, 2.541938,	−.671489, 2.541938,	
	3.016366, 3.341408	3.016366, 3.341408	
10^{-6}	−1.34710, 1.604116,	.998614,	1.41544
	2.51337413, 3.1022690,	2.51337430, 3.1022689,	
	3.4848785, 3.7939914	3.4848786, 3.7939911	
10^{-4}	−1.33677, 1.604231,	1.000956,	1.40403
	2.523215, 2.590149,	2.523219, 2.590160,	
	3.095483, 3.683980	3.095482, 3.6839740	

(16)

Summing up what has been observed in many numerical experiments and what I demonstrated to you on simple examples: the main effect of the noise in the data used for rational approximation is the appearance of doublets of zeros and poles separated by a distance roughly proportional to the "size" (relative with respect to the "deterministic" part of the data) of the noise. What is the distribution of these *Froissart doublets* depends on a specific type of the rational approximation we use. We have seen that for Padé approximants they coalesced on a circle of a radius of the closest singularity, for classical rational interpolants they sticked to the interpolation interval. Except for the close vicinity of such doublets, rational approximation reproduces values of the function with accuracy specified by the "size" of the noise. Can we find an explanation of this phenomenon?

5 Froissart Polynomial

If we assume that $f(z)$ – the function responsible for our unperturbed data – can be approximated well by a sequence of rational approximants then we can consider our perturbed data as perturbed data produced by some member of this sequence. Let us, therefore, assume that, for given m and n such that $m+n+1+2k=M$, there exists a rational function

$$r_{m,n}(z)=\frac{T_m(z)}{B_n(z)} \tag{17}$$

approximating $f(z)$ with some accuracy on some vicinity of the set of interpolation nodes $\{z_i\}_{i=0}^{M}$ where $M>m+n+1$. We are given data $\{d_i\}_{i=0}^{M}$ at these nodes; let me recall you that if some – even all – nodes appear with multiplicity $m_i>1$ then the data at this (multiple) node are the value of $f(z)$ and derivatives of $f(z)$ up to the (m_i-1)th one which are perturbed randomly with some "scale" as in (10) or (15)

$$d_i=d_i^{(0)}(1+\varepsilon r_i)\quad i=0,\dots,M \tag{18}$$

where $d_i^{(0)}$ are the exact data. Introducing $\{d_i^{(r)}\}_0^M$ – data of the same type as d_i's but coming from $r_{m,n}(z)$, we can write

$$\begin{aligned} d_i &= d_i^{(r)}(1+\varepsilon r_i)+(d_i^{(0)}-d_i^{(r)})(1+\varepsilon r_i)\quad i=0,\dots,M\\ &= d_i^{(r)}+\varepsilon\varrho_i \end{aligned} \tag{19}$$

introducing arbitrarily some ε as a scale of deviations of $d_i^{(r)}$ from d_i. It summarizes the effect of random perturbations and of differences $d_i^{(0)}-d_i^{(r)}$.

It can be proved using elementary algebra, but with some effort [2] that the rational interpolant of degrees $m+k$ and $n+k$, $R_{m+k,n+k}(z)$ constructed from data d_i has the form

$$R_{m+k,n+k}(z)=\frac{T_m(z)G_k(z)+\sum_{l=1}^{n+1}\varepsilon^l U_{m+k}^{(l)}(z)}{B_n(z)G_k(z)+\sum_{l=1}^{n}\varepsilon^l V_{n+k}^{(l)}(z)}. \tag{20}$$

Actually this formula is almost obvious – it says that if coefficients of a system of linear equations and right sides of the system are polynomials of the first degree in some ε, then the solution is a polynomial in ε of degree equal to the size of the system. If ε vanishes we must also get $r_{m,n}$, therefore the free terms of the numerator and of the denominator must be proportional to $T_m(z)$ and $B_n(z)$ correspondingly, with the same coefficient. What is not obvious is that this coefficient must be a polynomial of degree k in z.

If we now look at this formula with attention, we see that it perfectly explains the appearance of Froissart doublets – if both ε and $d_i^{(0)}-d_i^{(r)}$ are "sufficiently" small, i.e. ε is small, then zeros of the numerator and of the denominator are close to zeros of $T_m(z)G_k(z)$ and of $B_n(z)G_k(z)$. It is then $G_k(z)$, depending on ϱ_i's, that governs where the Froissart doublets appear – distances of zeros of numerator and of the denominator from zeros of $G_k(z)$ will be $O(\varepsilon)$. If $f(z)$ itself was a rational function $r_{m,n}(z)$ or it differed neglibly from $r_{m,n}(z)$ then $G_k(z)$ would depend only on perturbations r_i's and on $r_{m,n}$. In that case we shall call it the *Froissart polynomial* and use the symbol $F_k(z)$. This is the manageable situation and we can say a lot about $F_k(z)$ ([6], [7], [3]).

Before I discuss the Froissart polynomial let me point you out that as seen from (20) and (19), the zero-pole doublets appearing for perturbed data coming form arbitrary function will "behave" like Froissart doublets, i.e. will be distributed randomly with their mutual distance being $O(\varepsilon)$, only when ε is definitely larger than $d_i^{(0)} - d_i^{(r)}$. We can formulate it this way: Froissart doublets will be observed in a rational approximant constructed from perturbed data of a function when perturbations of the data are much larger than differences between exact data from this function and exact data from the best rational approximation of the same type but lower degrees, to the function of interest.

To say where the Froissart doublets go, we would have to study the distribution of zeros of $F_k(z)$. For this end one needs a formula expressing coefficients of this polynomial by r_i's and this formula depend on what type of rational interpolant we deal with. From considerations in [2] one can only say that they will be linear combinations of all possible products of k different r_i's from a set of M of them. The only thing that was possible to find from this very general information was the asymptotic behaviour of the pdf of zeros of F_k for $|z| \to \infty$ [1].

As was shown in [9] pdf of zeros of polynomials with random but real coefficients has two components: pdf of real zeros (called the singular component) and pdf of complex zeros (called the regular component). One can show that the pdf of the singular component of $F_k(z)$, which we denote $\Phi_s(x)$ behaves like $1/x^2$ as $x \to \infty$, while pdf of the regular component – $\Phi_r(z)$ – falls of like $1/|z|^4$ for $|z| \to \infty$, except for k directions along which it falls of like $1/|z|^3$. This behaviour means that whatever is a locus of coalescence of Froissart doublets when their number grows, their distribution has the long tail – the behaviour observed in numerical experiments.

Up to now it was possible to find the exact form of the pdf of zeros of the Froissart polynomial only for $k = 1$ – in that case coefficients of the polynomial are linear in r_i's therefore a pdf of the polynomial and of its derivative, necessary to calculate pdf of zeros according to formulae in [9], are very simple. The very interesting result came out for classical rational interpolation on equidistant nodes inside of a real interval [3]: the pdf of zeros of $F_1(x)$ (they are all real, here) has a maximum on the interpolation interval and also the probability of finding the zero on the interpolation interval is larger than the probability of finding it outside. It means that the Froissart doublets will appear rather inside the interpolation interval than outside, i.e. the extrapolation will be less affected by noise in data than the interpolation!

6 Conclusions

My goal was to convince you that rational functions are a very powerful tool in deciphering (or if you prefer: making a sophisticated guess about) an analytical structure of a function known from a finite set of "data". This explains why they are so good in approximating values of functions most economically – no wonder they are exploited in your pocket calculators and in internal compiler routines for transcendental functions. Moreover, the rational approximation of the form I discussed, has the amazing property of being "practically stable" with respect to perturbation of these data – noise in data goes mainly into Froissart doublets that almost annihilate themselves. This is one more reason, I think, why rational functions have much more potential in applications than generally recognised.

References

1. J.D. Fournier, M. Pindor. in preparation.
2. J.D. Fournier, M. Pindor. On multi-point Padé approximants to perturbed rational functions. *submitted to Constr. Math. and Funct. Th.*
3. J.D. Fournier, M. Pindor. Rational interpolation from stochastic data: A new froissart phenomenon. *Rel. Comp.*, 6:391–409, 2000.
4. M. Froissart. private information. J. Gammel, see also J. Gilewicz, Approximants de Padé LNM 667, Springer Verlag 1976 ch 6.4.
5. M.A. Galluci, W.B. Jones. Rational approximations corresponding to Newton series (Newton-Padé approximants). *J. Appr. Th.*, 17:366–372, 1976.
6. J. Gilewicz, M. Pindor. Padé approximants and noise: A case of geometric series. *JCAM*, 87:199–214, 1997.
7. J. Gilewicz, M. Pindor. Padé approximants and noise: rational functions. *JCAM*, 105:285–297, 1999.
8. J. Meinguet. On the solubility of the Cauchy interpolation problem. In *Proc. of the University of Lancaster Symposium on Approximation Theory and its Applications*, pages 137–164. Academic Press, 1970.
9. G.A. Mezinescu, D. Bessis, J.-D. Fournier, G. Mantica, F. D. Aaron. Distribution of roots of random real generalized polynomials. *J. Stat Phys.*, 86:675–705, 1997.
10. H. Stahl. Convergence of rational interpolants. Technical Report 299/8-1, Deutsche Forschungsgemeinschaft Report Sta, 2002.
11. S. Tokarzewski, J.J. Telega, M. Pindor, J. Gilewicz. Basic inequalities for multipoint Padé approximants to Stieltjes functions. *Arch. Mech.*, 54:141–153, 2002.
12. H. Wallin. Potential theory and approximation of analytic functions by rational interpolation. In Springer Verlag, editor, *Proc. of the Colloquium on Complex Analysis at Joensuu*, number 747 in LNM, pages 434–450, 1979.

Stationary Processes and Linear Systems

Manfred Deistler

Department of Mathematical Methods in Economics, Research Group
Econometrics and System Theory, Vienna University of Technology
Argentinierstr. 8, A-1040 Wien, Austria
Deistler@tuwien.ac.at

1 Introduction

Time series analysis is concerned with the systematic approaches to extract information from time series, i.e. from observations ordered in time. Unlike in classical statistics of independent and identically distributed observations, not only the values of the observations, but also their ordering in time may contain information. Main questions in time series analysis concern trends, cycles, dependence over time and dynamics.

Stationary processes are perhaps the most important models for time series. In this contribution we present two central parts of the theory of wide sense stationary processes, namely spectral theory and the Wold decomposition; in addition we treat the interface between the theory of stationary processes and linear systems theory, namely ARMA and state-space systems, with an emphasis on structure theory for such systems.

The contribution is organized as follows: In section 2 we give a short introduction to the history of the subject, in section 3 we deal with the spectral theory of stationary processes with an emphasis on the spectral representation of stationary processes and covariance functions and on linear transformations in frequency domain. In section 4, the Wold decomposition and prediction are treated. Due to the Wold decomposition every (linearly) regular stationary process can be considered as a (in general infinite dimensional) linear system with white noise inputs. These systems are finite dimensional if and only if their spectral density is rational and this case is of particular importance for statistical modeling. Processes with rational spectral densities can be described as solutions of ARMA or (linear) state space systems (with white noise inputs) and the structure of the relation between the Wold decomposition and ARMA or state space parameters is analyzed in section 5. This structure is important for the statistical analysis of such systems as is shortly described in section 6.

The intention of this contribution is to present main ideas and to give a clear picture of the structure of fundamental results. The contribution is

J.-D. Fournier et al. (Eds.): Harm. Analysis and Ratio. Approx., LNCIS 327, pp. 159–179, 2006.

oriented towards a mathematically knowledgeable audience. A certain familiarity with probability theory and the theory of Hilbert spaces is required. We give no proofs. The main references are [12], [7], [8], [10] and [11]. For the sake of brevity of presentation, we do not give reference, even to important original literature, if cited in the references listed above; for this reason important and seminal papers by Kolmogorov, Khinchin, Wold, Hannan, Kalman, Akaike and others will not be found in the list of references at the end of this contribution.

2 A Short View on the History

Here we give a short account of the historical development of the subject treated in this contribution. For the early history of time series analysis we refer to [3], for the history of stationary processes to the historical and bibliographic references in [12] and for a recent account to [6].

The early history of time series analysis dates back to the late eighteenth century. At this time more accurate data from the orbits of the planets and the moon become available due to improvements in telescope building. The fact that Kepler's laws result from a two body problem, whereas more than two bodies are in our planet system, triggered the interest in the detection of systematic deviations from these laws, and in particular in hidden periodicities and long term trends in these orbits. Harmonic analysis begins probably with a memoir published by Lagrange in 1772 on these problems. Subsequently the theory of Fourier series has been developed by Euler and Fourier. The method of least squares fitting of a line into a scatter plot was introduced by Legendre and Gauss in the early nineteenth century. Later in the nineteenth century Stokes and Schuster introduced the periodogram as a method for detecting hidden periodicities, to study, among others, sunspot numbers.

The empirical analysis of business cycles was on other important area is early time series analysis. In the seventies and eighties of the nineteenth century the British economist Jevons investigated fluctuations in economic time series.

The statistical theory of linear regression analysis was developed at the turn of the nineteenth to the twentieth century by Galton, Pearson, Gosset and others.

Stochastic models for time series, namely moving average and autoregressive models have been proposed by Yule in the nineteen-twenties, mainly in order to model non-exactly periodic fluctuations such as business cycles. Closely related is the work of Slutzky on the summation of random causes as a source of cyclical processes and Frisch's work on propagation and impulse problems in dynamic economics.

In the thirties and forties of the twentieth century, the theory of stationary processes was developed. The concept of a stationary process was introduced by Khinchin, the spectral representation is due to Kolmogorov, its proof based on the spectral representation of unitary operators was given by Karhunen. The properties of covariance functions were investigated by Khinchin, Wold, Cramer and Bochner; linear transformations of stationary processes appear in Kolmogorov's work. ARMA processes and the Wold representation are introduced in Wold's thesis. The prediction theory was developed by Kolmogorov, the rational case was investigated by Wiener and Doob.

At about the same time, the work of the Cowles Commission, constituting econometrics as a field of its own, came off. Triggered by the great economic depression, starting 1929, economic research activities in describing the "macrodynamics" of an economy were intensified. Questions of quantitative economic policy based on Keynesian theory led to problems of estimating parameters in macroeconomic models. In the work of the Cowles Commission, in particular in the works of Mann and Wald, Haavelmo and Koopmans, identifiability and least squares - and maximum likelihood estimators, in particular their asymptotic properties were investigated for multi-input, multi-output AR(X) systems.

In the late forties and fifties time series analysis, mainly for the scalar case, using non-parametric frequency domain methods was booming, in particular in engineering applications. The statistical properties of the periodogram were derived and the smoothed spectral estimators were introduced and analyzed by Tukey, Grenander and Rosenblatt, Hannan and others; analogously, non-parametric transfer function estimation methods based on spectral estimation were developed.

Almost parallel to the development of non-parametric frequency domain analysis, the parametric time domain counterparts, namely identification of AR(X) and ARMA(X) models, were developed in the forties, fifties and sixties of the twentieth century, mainly for the scalar case, by Mann and Wald, T.W. Anderson, Hannan, and others. For AR(X) models actual identification and the corresponding asymptotic theory turned out to be much simpler compared to the ARMA(X) case. The reason is that in the first case parameterization is simple and ordinary least squares estimators are asymptotically efficient and numerically simple at the same time. For the ARMA(X) case, on the other hand, the maximum likelihood estimator (MLE) has, in general, no explicit representation and is obtained by numerically optimizing the likelihood function. In addition questions of parameterization and the derivation of the asymptotic properties of the MLE are quite involved.

The work of Kalman, which is based on state space representations, triggered a "time domain revolution" in engineering. A particularly important aspect in the context of this paper is Kalman's work on realization and parameterization of, in general, multi-input, multi-output state space systems.

The book by [1] triggered a boom in applications, mainly because explicit instructions for actually performing applications for the scalar case were given. This included rules for transforming data to stationarity, for determining orders, an algorithm for maximizing the likelihood function and procedures for model validation.

A major shortcoming of the Box-Jenkins approach was that order determination had to be done by an experienced modeler in a non-automatic way. Thus an important step was the development and evaluation of automatic model selection procedures based on information criteria like AIC or BIC by Akaike, Hannan, Rissanen and Schwartz.

Identification of multivariate ARMA(X) and state space systems was further developed in the seventies and eighties of the last century, leading to a certain maturity of methods and theory. This is also documented in the monographs on the subject appearing in the late eighties and early nineties, in particular [9], [2], [8], [13] and [11]. However substantial research in this area is still going on.

3 The Spectral Theory of Stationary Processes

For more details, in particular for proofs, concerning results presented in this and the next section we refer to [12] and [7].

3.1 Stationary Processes and Hilbert spaces

Here we give the basic definitions and introduce the Hilbert space setting for stationary processes.
Let $(\Omega, \mathcal{A}, P)$ be a probability space and consider random variables $x_t : \Omega \rightarrow \mathbb{C}^s$ where $\mathbb{C}$ denotes the complex numbers. A stochastic process $(x_t \,|\, t \in T)$ is a family of random variables; here $T \subset \mathbb{R}$, where $\mathbb{R}$ denotes the real numbers and in particular the case $T = \mathbb{Z}$, the integers, is considered. In the latter case we write (x_t) and $\mathbb{Z}$ is interpreted as time axis. A stochastic process (x_t) is said to be *(wide sense) stationary* if

(i) $\mathbb{E}x_t^* x_t < \infty \qquad t \in \mathbb{Z}$
(ii) $\mathbb{E}x_t = m = const$
(iii) $\mathbb{E}x_{t+r}x_t^*$ does not depend on t, for every $r \in \mathbb{Z}$

holds. Here $*$ denotes the conjugate transpose of a vector or a matrix. For a stationary process the first and second moments exist and do not depend on time t; in particular the linear dependence relations between arbitrary one dimensional component variables $x_{t+r}^{(i)}$ and $x_t^{(j)}$; $i, j = 1, \ldots, s$, which are described by the (central) covariances $\mathbb{E}(x_{t+r}^{(i)} - \mathbb{E}x_{t+r}^{(i)})(x_t^{(j)} - \mathbb{E}x_t^{(j)})^*$ do only depend on the time difference r but not on the position in time t. The *covariance function* then is defined by

$$\gamma : \mathbb{Z} \rightarrow \mathbb{C}^{s \times s} : \gamma(r) = \mathbb{E}(x_{t+r} - \mathbb{E}x_{t+r})(x_t - \mathbb{E}x_t)^*$$

Note that here the covariance matrices are defined as being central; this is of no great importance and in many cases m is assumed to be zero.

Stationary processes are appropriate descriptions for many steady state random phenomena. But even in apparently nonstationary situations, such as in the presence of trends, stationary process are often used as models, e.g. for transformed data or as part of an overall model. The first and second moments do not fully describe a stationary process or its probability law, but they contain important information about the process which is sufficient e.g. for forecasting and filtering problems. Here we restrict ourselves to this information.

An arbitrary function $\gamma : \mathbb{Z} \to \mathbb{C}^{s\times s}$ is called *nonnegative - definite* if, for every T, $T = 1, 2, \dots$, the matrices of the form

$$\Gamma_T = \begin{bmatrix} \gamma(0) & \gamma(-1) \dots & \gamma(-T+1) \\ \gamma(1) & \gamma(0) & \vdots \\ \vdots & \ddots & \\ \gamma(T-1) & \dots & \gamma(0) \end{bmatrix}$$

are nonnegative-definite (denoted by $\Gamma_T \geq 0$). The following theorem gives a mathematical characterization of covariance functions of stationary processes:

Theorem 1. *A function $\gamma : \mathbb{Z} \to \mathbb{C}^{s\times s}$ is a covariance function of a stationary process if and only if it is nonnegative-definite.*

Let L_2 denote the Hilbert space of square integrable random variables $x : \Omega \to \mathbb{C}$ (or, to be more precise, of the corresponding P-a.e. equivalence classes), over the complex numbers, with inner product defined by $<x, y> = \mathbb{E}x\bar{y}$ where $\bar{y}$ denotes the conjugate of y. Then the Hilbert space $H_x \subset L_2$, spanned by the one dimensional process variables $x_t^{(i)}$, $t \in \mathbb{Z}$, $i = 1, \dots, s$ is called the *time domain* of the stationary process (x_t) (Note that condition (i) above implies $x_t^{(i)} \in L_2$.)

The stationarity condition (iii), in Hilbert space language, means that for every i, $i = 1, \dots, s$, the lengths $||x_t^{(i)}||$ of the $x_t^{(i)}$ do not depend on t and that the angles between $x_{t+r}^{(i)}$ and $x_t^{(j)}$ also do not depend on t. Note that the lengths are the square roots of the noncentral variances and the angles are noncentral correlations. Thus the operator shifting the process in time does not change lengths and angles.

This motivates the following theorem:

Theorem 2. *For every stationary process (x_t) there is a unique unitary operator $U : H_x \to H_x$ such that*

$$x_t^{(i)} = U^t x_0^{(i)},$$

holds.

We only consider stationary processes where the random variables are $\mathbb{R}^s$-valued; clearly then γ is $\mathbb{R}^{s\times s}$ valued; the complex notation is only used for simplification of formulas for the spectral representation.

Important examples of stationary processes are:

1. *White noise* (ε_t), which is defined by $\mathbb{E}\varepsilon_t = 0$; $\mathbb{E}\varepsilon_s\varepsilon_t' = \delta_{st}\varSigma$, where δ_{st} is the Kronecker Delta, ε_t' is the transpose of ε_t (the same notation is used for matrices) and where $\varSigma \geq 0$ holds. White noise has no linear "memory" (i.e. dependencies) over time.
2. *Moving average (MA) processes* can be represented as:

$$y_t = \sum_{j=0}^{q} b_j\varepsilon_{t-j}, \qquad b_j \in \mathbb{R}^{s\times m} \tag{1}$$

   where (ε_t) is white noise. (y_t) is said to be an MA(q) process if $b_q \neq 0$. A stationary process is an MA(q) process if and only if its covariance function satisfies $\gamma(q + r) = 0$ for some $q > 0$ and for all $r > 0$ and if $\gamma(q) \neq 0$. MA processes have finite linear memory.
3. *Linear - or MA* (∞) *processes* can be represented as

$$y_t = \sum_{j=-\infty}^{\infty} b_j\varepsilon_{t-j}, \qquad b_j \in \mathbb{R}^{s\times m} \tag{2}$$

   where (ε_t) is white noise and where the condition

$$\sum_{j=-\infty}^{\infty} \|b_j\|^2 < \infty \tag{3}$$

   guaranteeing the existence of the infinite sum in (2) in the sense of mean squares convergence, holds. In this paper limits of random variables are always defined in this sense; $\|b_j\|$ denotes a norm. Note that the first and second moments of MA(∞) processes are given by

$$\mathbb{E}y_t = 0$$

   and

$$\gamma(r) = \sum_{j=-\infty}^{\infty} b_j\varSigma b_{j-r}' \ . \tag{4}$$

   From (4) and (3), we see that an MA(∞) process has fading linear memory. The class of MA(∞) processes is a large class of stationary processes; it includes important subclasses, such as the class of *causal or one-sided MA*(∞) *processes*

$$y_t = \sum_{j=0}^{\infty} b_j\varepsilon_{t-j} \tag{5}$$

   or the class of stationary processes with rational spectral density treated in detail in section 5.

4. *Harmonic processes* are of the form

$$x_t = \sum_{j=1}^{h} \mathrm{e}^{\mathrm{i}\lambda_j t} z_j \tag{6}$$

where without loss of generality the (angular) frequencies λ_j are restricted to $(-\pi, \pi]$, $\lambda_1 < \lambda_2 < \ldots < \lambda_h$ and where $z_j : \Omega \to \mathbb{C}^s$ are, in general, genuine complex random variables describing random amplitudes and phases. In order to guarantee stationarity of (x_t) we have to assume

$$\mathbb{E} z_j^* z_j < \infty$$

$$\mathbb{E} z_j = \begin{cases} \mathbb{E} x_t & \text{for } \lambda_j = 0 \\ 0 & \text{for } \lambda_j \neq 0 \end{cases}$$

and

$$\mathbb{E} z_j z_l^* = 0 \qquad \text{for } j \neq l.$$

Since x_t is $\mathbb{R}^s$-valued, in addition we have

$$\lambda_{1+j} = -\lambda_{h-j}, \qquad j = 0, \ldots, h-1$$

and

$$z_{1+j} = \bar{z}_{h-j} \qquad j = 0, \ldots, h-1.$$

A harmonic process has a finite dimensional time domain; actually H_x is spanned by $z_j^{(i)}$, $i = 1, \ldots, s$, $j = 1, \ldots, h$.
The covariance function of a harmonic process is of the form

$$\gamma(r) = \sum_{j=1}^{h} \mathrm{e}^{\mathrm{i}\lambda_j r} F_j; \qquad F_j = \begin{cases} \mathbb{E} z_j z_j^* & \text{for } \lambda_j \neq 0 \\ \mathbb{E}(z_j - \mathbb{E} z_j)(z_j - \mathbb{E} z_j)^* & \text{for } \lambda_j = 0 \end{cases} \tag{7}$$

From this we see, that for (nontrivial) harmonic processes, the memory is not fading. The *spectral distribution function* $F : [-\pi, \pi] \to \mathbb{C}^{s \times s}$ is defined by

$$F(\lambda) = \sum_{j: \lambda_j \leq \lambda} F_j. \tag{8}$$

As is easily seen γ and F are in an one-to-one relation, and thus contain the same information about the underlying process, however this information is displayed in F in a different way. The $k-th$ diagonal element of F_j is a measure of the expected amplitude of the frequency component $\mathrm{e}^{\mathrm{i}\lambda_j t} z_j^{(k)}$ of the $k-th$ component process $(x_t^{(k)} \,|\, t \in \mathbb{Z})$. The (k, l) off-diagonal element of F_j (which is a complex number in general) measures by its absolute value the strength of the linear dependence between the $k-th$ and $l-th$ component process at frequency λ_j and by its phase the expected phase shift.

3.2 The Spectral Representation

In this subsection the Fourier representation for stationary processes and for their covariance functions are described. The main result states that, in a certain sense, every stationary process can be obtained as a limit of a sequence of harmonic processes.

A stochastic process $(z(\lambda) \,|\, \lambda \in [-\pi, \pi])$ where the random variables $z(\lambda) : \Omega \to \mathbb{C}^s$ are complex in general, is said to be a *process with orthogonal increments* if:

1. $\mathbb{E}z^*(\lambda)z(\lambda) < \infty$
2. $z(-\pi) = 0$
3. $\lim_{\varepsilon \downarrow 0} z(\lambda + \varepsilon) = z(\lambda)$, $\lambda \in [-\pi, \pi]$
4. $\mathbb{E}(z(\lambda_4) - z(\lambda_3))(z(\lambda_2) - z(\lambda_1))^* = 0$ for $\lambda_1 < \lambda_2 \leq \lambda_3 < \lambda_4$

holds. A process of orthogonal increments can be considered as a random variable or L_2^s-valued distribution function and thus determines an L_2^s-valued measure on the Borel sets of $[-\pi, \pi]$ and an associated integral (defined in the sense of convergence in mean squares).

By Theorem 2, the shift operator for a stationary process is unitary. From the spectral representation of unitary operators then we obtain:

Theorem 3 (Spectral representation of stationary processes). *For every stationary process (x_t) there is a unique process with orthogonal increments $(z(\lambda) \,|\, \lambda \in [-\pi, \pi])$ satisfying $z(\pi) = x_0$ and $z^{(i)}(\lambda) \in H_x$ such that*

$$x_t = \int_{[-\pi,\pi]} e^{i\lambda t} dz(\lambda) \tag{9}$$

holds.

The importance of the spectral representation (9) is twofold: First, it allows to interpret a stationary process in terms of frequency components. In particular, as has been said already, every stationary process may be obtained as a limit, pointwise in t, of a sequence of harmonic processes. Note that, in general, convergence will not be uniform in t. Second, as will be seen in the next subsection, certain operations are easier to perform and to interpret in frequency domain.

For a general stationary process its *spectral distribution function* $F : [-\pi, \pi] \to \mathbb{C}^{s \times s}$ is defined by

$$F(\lambda) = \mathbb{E}\tilde{z}(\lambda)\tilde{z}^*(\lambda) \qquad \text{where} \qquad \tilde{z}(\lambda) = \begin{cases} z(\lambda) & \text{for} \lambda < 0 \\ z(\lambda) - \mathbb{E}x_t & \text{for } \lambda \geq 0. \end{cases} \tag{10}$$

Theorem 3 implies that the covariance function has spectral representation of the form

$$\gamma(t) = \int_{[-\pi,\pi]} \mathrm{e}^{\mathrm{i}\lambda t} \mathrm{d}F(\lambda) \tag{11}$$

constituting a one-to-one relation between γ and F.

In many cases F is absolutely continuous w.r.t Lebesgue-measure, λ say; then there exists the so-called *spectral density* $f : [-\pi, \pi] \to \mathbb{C}^{s\times s}$ satisfying

$$F(\omega) = \int_{-\pi}^{\omega} f(\lambda)\mathrm{d}\lambda.$$

A sufficient condition for the existence of a spectral density is that

$$\sum_{j=-\infty}^{\infty} ||\gamma(t)||^2 < \infty \tag{12}$$

holds. Clearly a spectral density is uniquely defined only λ-a.e.; analogously to the case of random variables, we do not distinguish between f as function and f as an equivalence class of λ-a.e. identical functions. If (12) holds, then the one-to-one relation between f and γ is given by

$$\gamma(t) = \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}\lambda t} f(\lambda)\mathrm{d}\lambda \tag{13}$$

$$f(\lambda) = (2\pi)^{-1} \sum_{t=-\infty}^{\infty} \gamma(t)\mathrm{e}^{-\mathrm{i}\lambda t} \tag{14}$$

where the infinite sum in (14) corresponds to convergence in the L_2 over $[-\pi, \pi]$ with Lebesgue measure.

As a consequence of Theorem 1, a function $f : [-\pi, \pi] \to \mathbb{C}^{s\times s}$ is a spectral density if and only if

$$f(\lambda) \geq 0 \qquad \lambda - \text{a.e.}$$

and

$$\int_{-\pi}^{\pi} f(\lambda)\mathrm{d}\lambda \qquad (= \gamma(0)) \qquad \text{exists} \tag{15}$$

hold. Since we only consider $\mathbb{R}^s$-valued stationary processes, $\gamma(t) = \gamma(-t)'$ holds and thus, in addition $f(\lambda) = f(-\lambda)'$ has to be satisfied.

(Nontrivial) harmonic processes are examples for stationary processes having no spectral density.

F describes the second moments of $(\tilde{z}(\lambda) \,|\, \lambda \in [-\pi, \pi])$. In particular we have

$$F(\lambda_2) - F(\lambda_1) = \mathbb{E}(\tilde{z}(\lambda_2) - \tilde{z}(\lambda_1))(\tilde{z}(\lambda_2) - \tilde{z}(\lambda_1))^* \qquad \text{for} \qquad \lambda_2 > \lambda_1 \tag{16}$$

and, if the spectral density exists, this is equal to

$$\int_{\lambda_1}^{\lambda_2} f(\lambda)\mathrm{d}\lambda\,. \tag{17}$$

Interpreting the integral in (9) as a limit of a sums of the form (6), we can adopt the interpretation of F, given for harmonic processes, for general stationary processes, and, if f exists, analogously for f. For instance, for the case $s = 1$, the integral (17) is a measure for the expected "amplitudes" in this interval (often called frequency band) (λ_1, λ_2). In a certain sense, peaks of f (to be more precise areas under such peaks) mark the important frequency bands. Equation (15) gives a decomposition of the variance of the stationary process (x_t) into the variance contributions (17) corresponding to different frequency bands. For the case $s > 1$, e.g. the off diagonal elements in (17) (which are complex in general) again convey the information concerning the strength of the linear dependence between different component processes in a certain frequency band and about expected phase shifts there.

3.3 The Isomorphism between Time Domain and Frequency domain. Linear Transformations of Stationary Processes

The spectral representation (9) defines an isomorphism between the time domain H_x and an other Hilbert space introduced here, the so-called frequency domain. For simplicity of notation here we assume $\mathbb{E}x_t = 0$, otherwise $F(\lambda)$ has to be replaced by $\mathbb{E}z(\lambda)z^*(\lambda)$ in this subsection. As shown in this subsection, the analysis of linear transformations of stationary processes has some appealing features in the frequency domain.

We start by introducing the frequency domain: For the one-dimensional (i.e. $s = 1$) case, the frequency domain L_2^F is the L_2 over the measure space $([-\pi, \pi], \mathcal{B} \cap [-\pi, \pi], \mu_F)$, where $\mathcal{B} \cap [-\pi, \pi]$ is the σ-algebra of Borel sets over $[-\pi, \pi]$ and μ_F is the measure corresponding to the spectral distribution function, i.e. $\mu_F((a, b]) = F(b) - F(a)$. The isomorphism $I : H_x \to L_2^F$, given by (9) then is defined by $I(x_t) = \mathrm{e}^{\mathrm{i}\lambda t}$.

For the multivariate ($s > 1$) case, things are more complicated: First consider a measure μ on $\mathcal{B} \cap [-\pi, \pi]$ such that there exists a density $f^{(\mu)}$ for F w.r.t. this measure, i.e. such that

$$F(\lambda) = \int_{[-\pi,\lambda]} f^{(\mu)}\mathrm{d}\mu$$

holds. Such a measure always exists, one particular choice is the measure corresponding to the sum of all diagonal elements of F. Let $\varphi = (\varphi_1, \dots, \varphi_s)$ and $\psi = (\psi_1, \dots, \psi_s)$ denote row vectors of functions $\varphi_i, \psi_i : [-\pi, \pi] \to \mathbb{C}$; we identify φ and ψ if

$$\int_{[-\pi,\pi]} (\varphi - \psi) f^{(\mu)} (\varphi - \psi)^* \mathrm{d}\mu = 0$$

holds. Then the set (of equivalence classes)

$$L_2^F = \{\varphi \,|\, \int_{[-\pi,\pi]} \varphi f^{(\mu)} \varphi^* \mathrm{d}\mu < \infty\}$$

endowed with the inner product

$$<\varphi, \psi> = \int_{[-\pi,\pi]} \varphi f^{(\mu)} \psi^* \mathrm{d}\mu$$

is a Hilbert space; in particular, L_2^F is the *frequency domain* of the stationary process (x_t). As can be shown, the frequency domain does not depend on the special choice of the measure μ and of $f^{(\mu)}$. We have:

Theorem 4. *The mapping $I : H_x \to L_2^F$, defined by $I(x_t^{(j)}) = (0, \dots, \mathrm{e}^{\mathrm{i}\lambda t}, 0, \dots, 0)$, where $\mathrm{e}^{\mathrm{i}\lambda t}$ is in $j-th$ position, is an isomorphism of the two Hilbert spaces.*

Now, we consider *linear transformations* of (x_t) of the form

$$y_t = \sum_{j=-\infty}^{\infty} a_j x_{t-j}; \qquad a_j \in \mathbb{R}^{s \times m}. \tag{18}$$

Here

$$\sum_{j=-\infty}^{\infty} ||a_j|| < \infty \tag{19}$$

is a sufficient condition for the existence of the infinite sum in (18) or, to be more precise a necessary and sufficient condition for the existence of this infinite sum for all stationary inputs (x_t). As can easily be seen, the stationarity of (x_t) implies that $(x_t', y_t')'$ is (jointly) stationary. From (9) we obtain (using an obvious notation):

$$y_t = \int_{[-\pi,\pi]} \mathrm{e}^{\mathrm{i}\lambda t} \mathrm{d}z_y(\lambda) = \int_{[-\pi,\pi]} \mathrm{e}^{\mathrm{i}\lambda t} (\sum_{j=-\infty}^{\infty} a_j \mathrm{e}^{-\mathrm{i}\lambda j}) \mathrm{d}z_x(\lambda). \tag{20}$$

The *transfer function*, defined by

$$k(z) = \sum_{j=-\infty}^{\infty} a_j z^j \tag{21}$$

is in one-to-one relation with the *weighting* function $(a_j \,|\, j \in \mathbb{Z})$.

By definition $y_t^{(j)} \in H_x$ and thus $H_y \subset H_x$ holds. If U is the unitary shift for (x_t) then, by linearity and continuity of U, the restriction of U to H_y is the shift for (y_t). Due to the isomorphism between the time- and the frequency domain of (x_t), $k_j(\mathrm{e}^{-\mathrm{i}\lambda})\mathrm{e}^{\mathrm{i}\lambda t}$, where k_j is the $j-th$ row of the transfer function k, corresponds to $y_t^{(j)}$. Strictly speaking there are two transfer functions. The first is defined under the condition (20), from (21) as a function in the sense of pointwise convergence. The second is a matrix whose rows are elements of the frequency domain of (x_t). In the latter case (19) is not required.

Note that the discrete convolution (18) in time-domain corresponds to multiplication in frequency domain. In a sloppy notation we have from (20)

$$\mathrm{d}z_y(\lambda) = k(\mathrm{e}^{-\mathrm{i}\lambda})\mathrm{d}z_x(\lambda) \tag{22}$$

As a straightforward consequence from (20) we obtain:

Theorem 5. *Let* (x_t) *be stationary with spectral density* f_x *and let* (18) *hold. Then the spectral density* f_y *of* (y_t) *and the cross spectral density* f_{yx} *between* (y_t) *and* (x_t) *(i.e. the upper off-diagonal block in the spectral density matrix of the joint process* $(x_t', y_t')'$*) exist and are given by*

$$f_y(\lambda) = k(\mathrm{e}^{-\mathrm{i}\lambda})f_x(\lambda)k(\mathrm{e}^{-\mathrm{i}\lambda})^* \tag{23}$$

$$f_{yx}(\lambda) = k(\mathrm{e}^{-\mathrm{i}\lambda})f_x(\lambda) \tag{24}$$

where k *is given by* (21).

An analogous (and more general) result holds for spectral distribution functions. As a direct consequence of the above theorem, we see that for a linear process the spectral density exists and is of the form

$$f_y(\lambda) = (2\pi)^{-1}k(\mathrm{e}^{-\mathrm{i}\lambda})\sum k(\mathrm{e}^{-\mathrm{i}\lambda})^*; \quad k(z) = \sum_{j=-\infty}^{\infty} b_j z^j \tag{25}$$

where (3) holds. Note that (3) is more general than (19). The expression (18) shows an input process (x_t) transformed by a (deterministic) linear system (described by its weighting function $(a_j \mid j \in \mathbb{Z})$ or its transfer function k) to an output process (y_t). Such systems are *time invariant*, i.e. the a_j do not depend on t and *stable*, i.e. the input–output operator is bounded.

The effect of the linear transformation (18) can easily be interpreted from (22). For instance for the case $s = m = 1$, where k is scalar, the absolute value of $k(\mathrm{e}^{-\mathrm{i}\lambda})$ shows how the frequency components of (x_t) are amplified (for $|k(\mathrm{e}^{-\mathrm{i}\lambda})| > 1$) or attenuated (for $|k(\mathrm{e}^{-\mathrm{i}\lambda})| < 1$) by passing through the linear system and its phase indicates the phase-shift.

Linear systems with noise are of the form

$$y_t = \hat{y}_t + u_t \tag{26}$$

$$\hat{y}_t = \sum_{j=-\infty}^{\infty} l_j x_{t-j}; \qquad l_j \in \mathbb{R}^{s\times m} \tag{27}$$

$$u_t = \sum_{j=-\infty}^{\infty} k_j \varepsilon_{t-j}; \qquad k_j \in \mathbb{R}^{s\times s} \tag{28}$$

where (x_t) are the observed inputs, (u_t) is the noise on the unobserved outputs $(\hat{y}_t)$; (ε_t) is white noise and finally (y_t) are the observed outputs. We assume that

$$\mathbb{E}x_t u_s' = 0 \qquad \text{for all } s,t \in \mathbb{Z} \tag{29}$$

holds. This is equivalent to saying that $\hat{y}_t^{(j)}$ is the projection of $y_t^{(j)}$ on H_x or, due to the projection-theorem, that $\hat{y}_t^{(j)}$ is the best approximation of $y_t^{(j)} \in L_2$ by an element in H_x. We will then say that $(\hat{y}_t)$ is the best linear least squares approximation of (y_t) by (x_t).

If (x_t) has a spectral density, then we have

$$f_y(\lambda) = l(\mathrm{e}^{-\mathrm{i}\lambda}) f_x(\lambda) l(\mathrm{e}^{-\mathrm{i}\lambda})^* + (2\pi)^{-1} k(\mathrm{e}^{-\mathrm{i}\lambda}) \varSigma k(\mathrm{e}^{-\mathrm{i}\lambda})^* \tag{30}$$

and

$$f_{yx}(\lambda) = l(\mathrm{e}^{-\mathrm{i}\lambda}) f_x(\lambda) \tag{31}$$

where $l(z) = \sum_{j=-\infty}^{\infty} l_j z^j$, $k(z) = \sum_{j=-\infty}^{\infty} k_j z^j$ hold.

Formulas (30), (31) describe the relations between the second moments of observed inputs and outputs on one side and the covariance matrix $\varSigma$ and the two linear systems described by l and k on the other side. If $f_x(\lambda) > 0$, $\lambda \in [-\pi, \pi]$ holds, then l is obtained from the second moments of the observations by the so called Wiener filter formula

$$l(\mathrm{e}^{-\mathrm{i}\lambda}) = f_{yx}(\lambda) f_x(\lambda)^{-1}.$$

An important special case occurs if both transformations (30), (31) are causal, i.e. $l_j = 0$, $j < 0$; $k_j = 0$, $j < 0$ and the transfer functions are $k(z)$ and $l(z)$ are rational, i.e. there exist polynomial matrices

$$a(z) = \sum_{j=0}^{p} a_j z^j, \qquad b(z) = \sum_{j=0}^{q} b_j z^j, \qquad d(z) = \sum_{j=0}^{r} d_j z^j$$

such that $k = a^{-1}b$, $l = a^{-1}d$. In this case the linear system can be represented by an ARMA(X) system (see e.g. [8])

$$a(z) y_t = d(z) x_t + b(z) \varepsilon_t \tag{32}$$

or a state space system

$$s_{t+1} = A s_t + B \varepsilon_t + D x_t \tag{33}$$

$$y_t = C s_t + \varepsilon_t + E x_t. \tag{34}$$

Here z is used for a complex variable as well as for the backward shift $z(x_t \,|\, t \in \mathbb{Z}) = (x_{t-1} \,|\, t \in \mathbb{Z})$, s_t is the state at time t and A, B, C, D, E are parameter matrices. For further details we refer to [8].

4 The Wold Decomposition and Forecasting

The Wold decomposition provides important insights in the structure of stationary processes. These insights are particularly useful for forecasting.

Let (x_t) again be stationary. Linear least squares forecasting is concerned with the best (in the linear least squares sense) approximation of a "future" variable x_{t+h}, $h > 0$ by "past" (and "present") variables x_r, $r \leq t$. By the projection theorem, this approximation, $\hat{x}_{t,h} = (\hat{x}_{t,h}^{(1)}, \dots, \hat{x}_{t,h}^{(s)})'$ say, is obtained by projecting the components $x_{t+h}^{(j)}$ of x_{t+h} on the Hilbert space $H_x(t)$ spanned by the $x_r^{(j)}$; $r \leq t$, $j = 1, \dots, s$, yielding $\hat{x}_{t,h}^{(j)}$. Then $\hat{x}_{t,h}$ is called the *predictor* and $x_{t+h} - \hat{x}_{t,h}$ is called the *prediction error.*

As far as forecasting is concerned, we may distinguish the following two extreme cases:

A stationary process (x_t) is called *(linearly) singular* if $x_{t+h} = \hat{x}_{t,h}$ for one t and $h > 0$, and thus for all t, h, holds. Thus a singular process can be forecasted without error and $H_x(t) = H_x$ holds. Harmonic processes are examples for singular processes.

A stationary process (x_t) is called *(linearly) regular* if

$$\lim_{h \to \infty} \hat{x}_{t,h} = 0$$

for one t and thus for all t holds. White noise is a simple example for a regular process. For a regular process we have $\bigcap_{r \leq t} H_x(r) = \{0\}$.

Theorem 6 (Wold).

1. *Every stationary process (x_t) can be uniquely decomposed as*

$$x_t = y_t + z_t \tag{35}$$

where

$$\mathbb{E} y_s z_t' = 0 \qquad \textit{for all } s, t$$

$y_t^{(j)}$, $z_t^{(j)} \in H_x(t)$, $j = 1, \dots, n$ and where (y_t) is regular and (z_t) is singular.

2. *Every regular process (y_t) can be represented as*

$$y_t = \sum_{j=0}^{\infty} k_j \varepsilon_{t-j}, \qquad \sum_{j=0}^{\infty} ||k_j||^2 < \infty \tag{36}$$

where (ε_t) is white noise and where $H_\varepsilon(t) = H_y(t)$ holds.

From the theorem above we see that $H_x(t)$ is the orthogonal sum of $H_y(t)$ and $H_z(t)$ and thus we can predict the regular and the singular part separately. For a regular process we can split the Wold representation (36) as

$$y_{t+h} = \sum_{j=h}^{\infty} k_j \varepsilon_{t+h-j} + \sum_{j=0}^{h-1} k_j \varepsilon_{t+h-j} \tag{37}$$

The components of the first part of the r.h.s. of (37) are elements of $H_y(t) = H_\varepsilon(t)$ and the components of the second part of the r.h.s. are orthogonal to $H_y(t)$. Thus, by the projection theorem,

$$\hat{y}_{t,h} = \sum_{j=h}^{\infty} k_j \varepsilon_{t+h-j} \tag{38}$$

and the second part on the r.h.s of (37) is the prediction error. Expressing the $\varepsilon_l^{(j)}$ as linear combinations or limits of linear combinations of $y_r^{(j)}$, $r \leq l$ and inserting this in (38) gives the prediction formula, i.e. $\hat{y}_{t,h}$ as a linear function of y_r, $r \leq t$. Thus, for given Wold representation (36) (i.e. for given k_j, $j = 0, 1, \ldots$) the predictor formula can be determined.

From (36) we see that every linearly regular process can be interpreted as the output of a linear system with white noise inputs. Thus the spectral density f_y of (y_t) exists and is of the form (see (25))

$$f_y(\lambda) = (2\pi)^{-1} k(\mathrm{e}^{-\mathrm{i}\lambda}) \Sigma k(\mathrm{e}^{-\mathrm{i}\lambda})^* \tag{39}$$

where

$$k(z) = \sum_{j=0}^{\infty} k_j z^j, \qquad \Sigma = \mathbb{E}\varepsilon_t \varepsilon_t'. \tag{40}$$

5 Rational Spectra, ARMA and State Space Systems

From a statistical point of view, AR(X), ARMA(X) and state space systems are the most important models for stationary processes. The reason is that for such models only finitely many parameters have to be estimated and that a large class of linear systems can be approximated by such models. Here, for simplicity of presentation, we only consider the case of no observed inputs. Most of the results can be extended to the case of observed inputs in a straight forward manner. In this section we investigate the relation between the "internal parameters" (system parameters and possibly the variance covariance matrix Σ of the white noise (ε_t)) and external behavior (described by the second moments of the observations (y_t) or the transfer function $k(z)$) of such systems.

An *ARMA system* is of the form

$$a(z) y_t = b(z) \varepsilon_t \tag{41}$$

where z is the backward shift operator, ε_t is the unobserved white noise, $a(z) = \sum_{j=0}^{p} a_j z^j$, $b(z) = \sum_{j=0}^{q} b_j z^j$, a_j, b_j; $\in \mathbb{R}^{s \times s}$ and (y_t) is the (observed) output process. As is well known, the set of all solutions of a linear difference equation (41) is of the form one particular solution plus the set of all solutions of $a(z)y_t = 0$. We are only interested in stationary solutions; they are obtained by the so called *z-transform*. In solving (41), the equation, in a certain sense, has to be multiplied by the inverse of $a(z)$ from the left. Using the fact that multiplication of power series in the backward shift and in $z \in \mathbb{C}$ is done in the same way, we obtain:

Theorem 7. *Under the assumption*

$$\det a(z) \neq 0 \qquad |z| \leq 1 \tag{42}$$

the causal stationary solution of (41) *is given by*

$$y_t = k(z)\varepsilon_j = \sum_{j=0}^{\infty} k_j \varepsilon_{t-j} \tag{43}$$

where the transfer function is given by

$$k(z) = \sum_{j=0}^{\infty} k_j z^j = a^{-1}(z)b(z) = (\det a(z))^{-1} adj(a(z))b(z) \qquad |z| \leq 1 \tag{44}$$

*Here "*det*" and "adj" denote the determinant and the adjoint respectively.*

Condition (42) is called the *stability condition.* It guarantees that the norms $\|k_j\|$ in the causal solution converge geometrically to zero. Thus the ARMA process has an exponentially fading (linear) memory. For actually determining the k_j, the following block recursive linear equation system

$$a_0 k_0 = b_0, \qquad a_0 k_1 + a_1 k_0 = b_1, \ldots$$

obtained by a comparison of coefficients in $a(z)k(z) = b(z)$, has to be solved.

If in addition the so-called *miniphase condition*

$$\det b(z) \neq 0 \qquad |z| < 1 \tag{45}$$

is imposed, then we have $H_y(t) = H_\varepsilon(t)$ and thus the solution (43) is already the Wold representation (36). Condition (42) sometimes is relaxed to $\det a(z) \neq 0$ for $|z| = 1$. Then there exists a stationary solution $y_t = \sum_{j=-\infty}^{\infty} k_j \varepsilon_{t-j}$, which in general will not be causal.

A state space system (in innovations form) is given as

$$s_{t+1} = A s_t + B \varepsilon_t \tag{46}$$

$$y_t = C s_t + \varepsilon_t \tag{47}$$

Here s_t is the, in general unobserved, n-dimensional state and $A \in \mathbb{R}^{n \times n}$. $B \in \mathbb{R}^{n \times s}$, $C \in \mathbb{R}^{s \times n}$ are parameter matrices.

The stability condition (42) is of the form

$$|\lambda_{\max}(A)| < 1 \tag{48}$$

where $\lambda_{\max}(A)$ denotes an eigenvalue of A of maximum modulus. The steady state solution then is of the form

$$y_t = (C(Iz^{-1} - A)^{-1}B + I)\varepsilon_t. \tag{49}$$

Here the coefficients of the transfer function are given as $k_j = CA^{j-1}B$ for $j > 0$.

The miniphase condition

$$|\lambda_{\max}(A - BC)| \leq 1 \tag{50}$$

then guarantees that (49) corresponds to Wold representation (36). Note that (43) and (49) define causal linear processes, with a spectral density given by (39). Clearly the transfer function of both, ARMA and state space solutions are rational and so are their spectral densities. The following theorem clarifies the relation between rational spectral densities, ARMA and state space systems:

Theorem 8. *1. Every rational and λ-a.e nonsingular spectral density f_y can be uniquely factorized (as in* (39)*) such that $k(z)$ is rational, analytic within a circle containing the closed unit disk,* $\det k(z) \neq 0$ $|z| < 1$*, $k(0) = I$ and $\Sigma > 0$;*

2. For every rational transfer function $k(z)$ with the properties given in (1)*, there is a stable and miniphase ARMA system with $a_0 = b_0$ and conversely, every such ARMA system has a rational transfer function with the properties given in* (1)*;*

3. For every rational transfer function $k(z)$ with the properties given in (1)*, there is a stable and miniphase state space system and conversely, every such state space system has a rational transfer function with the properties given in* (1)*.*

Thus, in particular, (stable and causal) ARMA (with $a_0 = b_0$)- and (stable and causal) state space systems represent the same class of transfer functions or spectral densities.

Now we consider the “inverse problem” of finding an ARMA or state space system from the spectral density, or, equivalently, from the transfer function. From now on we assume throughout that the stability and the miniphase conditions hold. Two ARMA systems (a, b) and $(\tilde{a}, \tilde{b})$ say, are called *observationally equivalent* if they have the same transfer function (and thus for given Σ, the same second moments of the solution) i.e. if $a^{-1}b = \tilde{a}^{-1}\tilde{b}$ holds. Observational equivalence for state space systems is defined analogously. Now, in general, (a, b) is not uniquely determined from $k = a^{-1}b$. Let us assume that (a, b) is relatively left prime, i.e. that every common left (polynomial matrix) divisor u of (a, b) is unimodular, i.e. $\det u(z) = const \neq 0$ holds. Here a polynomial matrix u is called a common left divisor of (a, b), if there exist polynomial matrices $(\tilde{a}, \tilde{b})$ such that $(a, b) = u(\tilde{a}, \tilde{b})$ holds. In a certain sense, relative left primeness excludes redundant ARMA systems.

Then we have:

Theorem 9. *Let (a, b) and $(\tilde{a}, \tilde{b})$ be relatively left prime; then (a, b) and $(\tilde{a}, \tilde{b})$ are observationally equivalent if and only if there exists a unimodular u matrix such that*

$$(a, b) = u(\tilde{a}, \tilde{b}) \tag{51}$$

holds.

A state space system (A, B, C) is called *minimal* if the state dimension n is minimal among all state space systems corresponding to the same transfer function. This is the case if and only if the observability matrix

$$\mathcal{O}_n = (C', A'C', \ldots, (A')^{n-1}C')'$$

and the controllability matrix

$$\mathcal{C}_n = (B, AB, \ldots, A^{n-1}B)$$

both have rank n. Also minimality is a requirement of nonredundancy. We have:

Theorem 10. *Two minimal state space systems (A, B, C) and $(\tilde{A}, \tilde{B}, \tilde{C})$ are observationally equivalent if and only if there exists a nonsingular matrix $T \in \mathbb{R}^{n \times n}$ such that*

$$A = T\tilde{A}T^{-1}, \qquad B = T\tilde{B}, \qquad C = \tilde{C}T^{-1}$$

holds.

A class of ARMA or state space systems is called *identifiable* if it contains no distinct observationally equivalent systems. Of course identifiability is a desirable property, because it attaches to a given spectral density or a given transfer function a unique ARMA or state space system. In general terms, identifiability is obtained by selecting representatives from the classes of observationally equivalent systems. In addition, from an estimation point of view, subclasses of the class of all ARMA or state space systems, leading to finite dimensional parameter spaces and to a continuous dependence of the parameters on the transfer function (for details see e.g. [5]) are preferred.

As an example consider the set of ARMA systems (a, b) where (42), (45) and $a_0 = b_0 = I$ hold, which are relatively left prime and where the degrees of $a(z)$ and $b(z)$ are both p and where (a_p, b_p) has rank s. We denote the set of all corresponding vec $(a_1, \ldots, a_p, b_1, \ldots, b_p)$, where "vec" means stacking the columns of the respective matrix, by $T_{p,p}$. As can be shown, $T_{p,p}$ contains a nontrivial open subset of $\mathbb{R}^{2ps^2}$ and is identifiable as under these conditions, since (51) implies that u must be the identity matrix; thus $T_{p,p}$ is a "reasonable" parameter space. In this setting a system is described by the integer valued parameter p and by the real valued parameters in vec $(a_1, \ldots, b_p)$. For the description of f_y, of course also Σ is needed. Let $U_{p,p}$ denote the set of all transfer functions k corresponding to $T_{p,p}$ via (44). Then due to identifiability there exists a mapping $\rho : U_{p,p} \to T_{p,p}$ attaching to the transfer functions the corresponding ARMA parameters. Such a mapping is called parameterization. A disadvantage of the specific approach described above is that for $s > 1$ not every transfer function corresponding to an ARMA system can be described in this way, i.e. there are k for which there is no p such that $k \in U_{p,p}$.

For a general account on parameter spaces for and parameterizations of ARMA and state space systems we refer to [8], [4] and [5].

6 The Relation to System Identification

In system identification, the task is to find a "good" model from data. The approach we have in mind here is semi-nonparametric in the sense that identification can be decomposed into the following two steps, see [5]:

1. Model selection: Here we commence from the original model class, i.e. the class of all a priori candidate systems, for instance the class of all ARMA systems (41), for given s and for arbitrary p and q. The task then is to find a "reasonable" subclass from the data, such as the class $T_{p,p}$ described in the last section; typically estimation of the subclass consists in estimation of integers, such as p for $T_{p,p}$, e.g. by information criteria such as AIC or BIC see e.g. [8]
2. Estimation of real valued parameters: Here, for a given subclass, the system parameters such as *vec* $(a_1, \ldots, b_p)$ for $T_{p,p}$ and the variance covariance matrix Σ are estimated. As has been mentioned already in the previous section, the subclasses are chosen in a way such that they can be described by finite dimensional parameter spaces.

It should be noted, that in most cases only parameters describing the second moments (the spectral density) of (y_t) are estimated; accordingly estimation of moments of order greater than two, which is of interest in some applications, is not considered here.

For the AR(X) case, i.e. when $b(z) = I$ holds on the r.h.s. of (32), identification is much simpler compared to the ARMA(X) or state space case. This is the reason why AR(X) models still dominate in many applications, despite the fact that they are less flexible, so that more parameters may be needed for modeling. Again we restrict ourselves to the case of no observed inputs. Once the maximum lag p has been determined, assuming $a_0 = I$, the parameter space $T_p = \{vec(a_1, \dots, a_p) \in \mathbb{R}^{s^2 p} \mid \det a(z) \neq 0 \quad |z| \leq 1\}$ is identifiable. Ordinary least squares type estimators (such as Yule-Walker estimators) can be shown to be consistent and asymptotically efficient under general conditions on the one hand and are easy to calculate on the other hand.

For ARMA and state space systems identification is more complicated for two reasons:

1. The maximum likelihood estimators (for the real valued parameters) are in general not explicitly given, but have to be determined by numerical optimization procedures.
2. Parameter spaces and parameterizations are more complicated.

In this case, for a full understanding of identification procedures an analysis of topological and geometric properties of parameter spaces and parameterizations is needed. For instance as shown in [8] the MLE's of the transfer function are consistent; thus continuity of the parametrization guarantees consistency of parameter estimators. The importance of such structural properties for identification is discussed in detail in [8], [4] and [5].

References

1. G. Box, G. Jenkins, *Time Series Analysis. Forecasting and Control*, Holden Day, San Francisco, 1970
2. P. E. Caines, *Linear Stochastic Systems*, John Wiley & Sons, New York, 1988.
3. H. Davis, *The Analysis of Economic Time Series*, Principia Press, Bloomington, 1941.
4. M. Deistler, Identification of Linear Dynamic Multiinput/Multioutput Systems, *in* D. Pena et al. (ed.), *A Course in Time Series Analysis*, John Wiley & Sons, New York, 2001.
5. M. Deistler, System Identification - General Aspects and Structure, *in* G. Goodwin (ed.), *System Identification and Adaptive Control*, (Festschrift for B.D.O. Anderson), Springer, London, pp. 3 – 26, 2001.
6. M. Deistler, System Identification and Time Series Analysis: Past, Present and Future., *in* B. Pasik-Duncan (ed.), *Stochastic Theory and Control*, Springer, Kansas, USA, pp. 97 – 108. (Festschrift for Tyrone Duncan), 2002.
7. E. Hannan, *Multiple Time Series*, Wiley, New York, 1970.
8. E. Hannan, M. Deistler, *The Statistical Theory of Linear Systems*, John Wiley & Sons, New York, 1988.
9. L. Ljung, *System Identification: Theory for the User*, Prentice Hall, Englewood Cliffs, 1987.
10. M. Pourahmadi, *Foundations of Time Series Analysis and Predicton Theory*, Wiley, New York, 2001.
11. G. Reinsel *Elements of Multivariate Time Series Analysis*, Springer Verlag, New York, 1993.

12. Y. A. Rozanov, *Stationary Random Processes*, Holden Day, San Francisco, 1967.
13. T. Söderström, P. Stoica, *System Identification*, Prentice Hall, New York, 1989.

Parametric Spectral Estimation and Data Whitening

Elena Cuoco

INFN, Sezione di Firenze, Via G. Sansone 1, 50019 Sesto Fiorentino (FI),
present address: EGO, via Amaldi, Santo Stefano a Macerata, Cascina (PI)
cuoco@fi.infn.it, elena.cuoco@ego-gw.it

Abstract

The knowledge of the noise Power Spectral Density is fundamental in signal processing for the detection algorithms and for the analysis of the data. In this lecture we address both the problem of identifying the noise Power Spectral Density of physical system using parametric techniques and the problem of the whitening procedure of the sequence of data in time domain.

1 Introduction

In the detection of signals buried in noisy data, it is necessary to know the Power Spectral Density (PSD) $S(\nu)$ of the noise of the detector in such a way to be able to perform the Wiener filter [9]. By the theory of optimal filtering for signal buried in stationary and Gaussian noise [9, 10], if we are looking for a signal of known wave-form with unknown parameters, the optimal filter is given by the Wiener matching in filter domain

$$C(\theta) = \int_{-\infty}^{\infty} \frac{x(\nu)h(\nu,\theta)^*}{S(\nu)}, \tag{1}$$

where $h(\nu,\theta)$ is the template of the signal we are looking for, θ are the parameters of the waveform and $x(\nu)$ is the Fourier Transform of our sequence of data $x[n]$.

We can implement the Wiener filter in the frequency domain and, supposing the noise is stationary, we can estimate the PSD, for example, as a windowed averaged periodogram:

$$P_{\mathrm{PER}} = \frac{1}{N}\left|\sum_{n=0}^{N-1} x[n]\exp(-2\mathrm{i}\pi\nu n)\right|^2 . \tag{2}$$

J.-D. Fournier et al. (Eds.): Harm. Analysis and Ratio. Approx., LNCIS 327, pp. 181–191, 2006.

$$P_{\text{medio}} = \frac{1}{K} \sum_{m=0}^{K-1} P_{\text{PER}}^{(m)}(\nu) = S(\nu). \tag{3}$$

Sometimes it could be useful to implement the Wiener filter in time domain [1, 3], in this case what we perform in time domain is the so called "whitening procedure", i.e. we estimate the filter which fits our PSD and we use the filter's parameters to performs the division by the PSD of the Wiener filter in time domain.

These procedure could help if we know that our noise is not stationary. In that case we could use adaptive whitening filter in time domain, but it is out of the purpose of this lecture to go inside the theory of adaptive filters [6, 5, 11].

2 Parametric modeling for Power Spectral Density: ARMA and AR models

The advantages of parametric modeling with respect to the classical spectral methods are described in an exhaustive way in reference [4]. We focused on the rational function in the field of parametric estimation, because they offer the possibility of building a whitening *stable* filter in the time domain.

What we want to parametrize is the transfer function of our physical system: a linear system can be modeled as an object which transform an input sequence $w[z]$ in the output $x[z]$ by the transfer function $H[z]$ (see fig. 1).

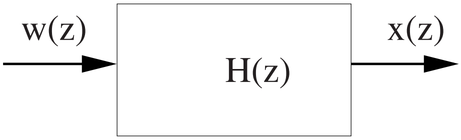


Fig. 1. Linear model

If the transfer function is in the form

$$H(z) = \frac{B(z)}{A(z)} \tag{4}$$

this is a rational transfer function modeled system.

In particular a general process described by a ARMA (p, q) model satisfies the relation:

$$x[n] = -\sum_{k=1}^{p} a_k x[n-k] + \sum_{k=0}^{q} b_k w[n-k] \tag{5}$$

in the time domain, and its transfer function, in z–domain, is given by $H(z) = B(z)/A(z)$, where $A(z) = \sum_{k=0}^{p} a_k z^{-k}$ represent the autoregressive (AR) part and $B(z) = \sum_{k=0}^{q} b_k z^{-k}$ the moving average (MA) part.

An AR model is called an all poles model, while the MA one is called an all zero models. Some physical systems are well described by an ARMA model, other by an AR and other by the MA one.

If we want to model our physical process with a parametric one, we have to choose the appropriate model and then we have to estimate its parameters.

The parameters of an ARMA model are linked to the autocorrelation function of the system $r_{xx}[n]$. So we have to estimate it before determine the a_k or b_k parameters. The relation between these parameters and the autocorrelation function is given by the Yule–Walker equations.

2.1 The Yule–Walker equations

The parameters of the ARMA model are linked to the autocorrelation function of the process by the Yule–Walker equations [4].

One way to derive the Yule–Walker is to write the correlation function $r_{xx}[k]$ in the first term of the equation (5). To do this, we have to simply multiply the equation (5) by $x^*[n-k]$ and take the expectation value on both sides.

We obtain the relation

$$r_{xx}[k] = -\sum_{l=1}^{p} a_l r_{xx}[k-l] + \sum_{l=0}^{q} b_l r_{xw}[k-l], \tag{6}$$

where $r_{xw}[k]$ is the cross correlation between the output $x[n]$ and the driving noise $w[n]$. Let $h[l]$ be the taps of the filter $H(z)$, the filter being causal, we can write the output as $x[n] = \sum\limits_{l=-\infty}^{n} h[n-l]w[l]$. It is evident that $r_{xw}[k] = 0$ for $k > 0$, since the output depends only on the driving input at step $l < n$. Noting that

$$r_{xw}[k] = \overline{x^*[n]w[n+k]} = \overline{(\sum_{l=-\infty}^{n} h^*[n-l]w^*[l]w[n+k])} = \sigma^2 h^*[k],$$

(σ is the amplitude of the driving white noise) we can write the Yule–Walker equations in the following way

$$r_{xx}[k] = \begin{cases} -\sum_{l=1}^{p} a_l r_{xx}[k-l] + \sigma^2 \sum_{l=0}^{q-k} h[l]^* b_{l+k} & \text{for } k = 0, 1, \ldots, q \\ -\sum_{l=1}^{p} a_l r_{xx}[k-l] & \text{for } k \geq q+1\,. \end{cases} \tag{7}$$

In the general case of an ARMA process we must solve a set of non linear equations while, if we specialize to an AR process (that is an all-poles model) the equations to be solved to find the AR parameters become linear.

The relationship between the parameters of the AR model and the autocorrelation function $r_{xx}(n)$ is given by the Yule–Walker equations written in the form

$$r_{xx}[k] = \begin{cases} -\sum_{l=1}^{p} a_l r_{xx}[k-l] & \text{for } k \geq 1 \\ -\sum_{l=1}^{p} a_l r_{xx}[-l] + \sigma^2 & \text{for } k = 0 \,. \end{cases} \tag{8}$$

In the following, we specialize the discussion to the AR estimation, since we are looking the way to build a whitening stable filter in the time domain, and the AR filter give us the solution.

3 AR and whitening process

An AR(P) process is identified by the relation

$$x[n] = \sum_{k=1}^{P} a_k x[n-k] + w[n], \tag{9}$$

$w[n]$ being the driving white noise.

The tight relation between the AR filter and the whitening filter is clear in the figure 2. The figure describes the scheme of an AR filter. The AR filter colors the white process $w[n]$ at the input of the filter (look at the picture from left to right). If you look at the picture from right to left you see a colored process at the input which passes through the AR inverse filter coming out as a white process.

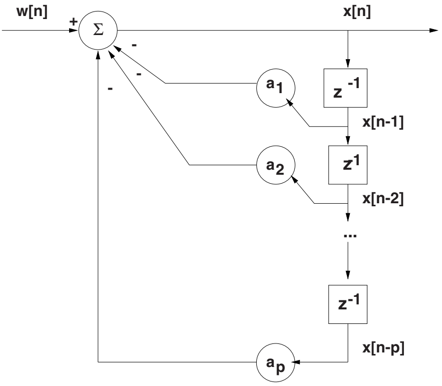


Fig. 2. Link between AR filter and whitening filter

Suppose you have a sequence $x[n]$ of data which is characterized by an autocorrelation $r_{xx}[n]$ which is not a delta function, and that you need to remove all the correlation, making it a white process (see refs [1, 2, 3] for application of whitening in real cases), the idea is to model $x[n]$ as an AR process, find the AR parameters and use them to whiten the process.

Since we want to deal with real physical problem we have to assume the causality of the filter. So when we whiten the data we must assure that the causality has been preserved. Moreover we must have a stable filter to avoid divergences in the application of this filter to the data. In the next section we will show how estimating the AR parameters assures the causality and stability of the whitening filter in the time domain.

3.1 Minimum phase filter and stability

The necessary condition to have a stable and causal filter $H(z)$ is that the all poles of the filter are inside the unit circle of the z–plane [4, 8]. This filter is called minimum-phase filter. A complete anticausal filter will have all its poles outside the unit circle. This filter is called a maximum phase filter.

If a system has poles or zeros outside the unit circle, it can be made minimum phase by moving poles and zeros z_0 inside the unit circle. For example if we want to put inside the unit circle a zero we have to multiply the function by this term

$$\frac{z^{-1} - z_0^*}{1 - z_0 z^{-1}}. \tag{10}$$

This will alter the phase, but not the magnitude of the transfer function. If we want to find the filter $H(z)$ of a linear system for a random process with given PSD $S(z)$ (the complex PSD), which satisfies the minimum phase condition, we must perform a spectral factorization (see [8]) in causal and anti-causal components:

$$S(z) = H(z)H(z^{*-1}). \tag{11}$$

We can perform this operation in alternative way. We can find a rational function fit to the PSD with polynomials that are minimum phase. A minimum-phase polynomial is one that has all of its zeros and poles strictly inside the unit circle. If we consider a rational function $H(z) = B(z)/A(z)$, both $B(z)$ and $A(z)$ must be minimum phase polynomials.

If we restrict to AR fit, we are looking for a polynomial $A(z)$ which is a minimum phase one. The AR estimation algorithm we choose will ensure that this condition is always satisfied.

4 AR parameters estimation

There are different algorithms to estimate the AR parameters of a process which we assumed can be modeled as an autoregressive one, for example the

Levinson, Durbin or Burg ones [4, 8]. We are looking for parameters of a transfer function which models a real physical system.

We can show that problem of determining the AR parameters is the same of that of finding the optimal "weights vector" $\boldsymbol{w} = w_k$, for $k = 1, \dots, P$ for the problem of Linear Prediction [4]. In the Linear Prediction we would predict the sample $x[n]$ using the P previous observed data $\boldsymbol{x}[n] = \{x[n-1], x[n-2], \dots, x[n-P]\}$ building the estimate $\hat{x}[n]$ as a transversal filter:

$$\hat{x}[n] = \sum_{k=1}^{P} w_k x[n-k] \, . \tag{12}$$

We choose the coefficients of the Linear Predictor filter by minimizing a cost function that is the mean squares error $\varepsilon = \mathbb{E}[e[n]^2]$, being

$$e[n] = x[n] - \hat{x}[n] \tag{13}$$

the error we made in this prediction, obtaining the so called Normal or Wiener-Hopf equations

$$\varepsilon_{\min} = r_{xx}[0] - \sum_{k=1}^{P} w_k r_{xx}[-k] \, , \tag{14}$$

which are identical to the Yule–Walker equations with

$$w_k = -a_k \tag{15}$$

$$\varepsilon_{\min} = \sigma^2 . \tag{16}$$

This equivalence relationship between AR model and linear prediction assures us to obtain a filter which is stable and causal [4].

It is possible to show that an equivalent representation for an AR process is based on the value of the autocorrelation function at lag 0 and a set of coefficients called reflection coefficient or parcor (partial correlation coefficients) k_p, $p = 1, \dots, P$, P being the order of our model. The k-th reflection coefficient is the partial correlation coefficient between $x[n]$ and $x[n-k]$, when the dependence of the samples in between has been removed.

We report here after the procedure to estimate the reflection coefficients and the AR parameters using the Levinson-Durbin algorithm.

The algorithm proceeds in the following way:

- Initialize the mean squares error as $\varepsilon_0 = r_{xx}[0]$.
- Introduce the reflection coefficients k_p, linked to the partial correlation between the $x[n]$ and $x[n-p]$ [8]:

$$k_p = \frac{1}{\varepsilon_{p-1}} \left[r_{xx}[p] - \sum_{j=1}^{p-1} a_j^{(p-1)} r_{xx}[p-j] \right] . \tag{17}$$

- At the p stage the parameter of the model is equal to the p-th reflection coefficient

$$a_p^{(p)} = k_p. \tag{18}$$

- The other parameters are updated in the following way:

For $1 \leq j \leq p-1$

$$a_j^{(p)} = a_j^{(p-1)} - k_p a_{p-j}^{(p-1)} \tag{19}$$

$$\varepsilon_p = (1 - k_p^2)\varepsilon_{p-1} \tag{20}$$

- At the end of the p loop, when $p = P$, the final AR parameters are

$$a_j = a_j^{(P)}, \qquad \sigma^2 = \varepsilon_P. \tag{21}$$

5 The whitening filter in the time domain

We can use the reflection coefficients in implementing the whitening filter [2, 3] in a lattice structure. Let us suppose to have a stochastic Gaussian and stationary process $x[n]$ which we modeled as an autoregressive process of order P. Remember that an AR model could be viewed as a linear prediction problem. In this context we can define the *forward* error (FPE) for the filter of order P in the following way

$$e_P^f[n] = x[n] + \sum_{k=1}^{P} a_k^{(P)} x[n-k]\,, \tag{22}$$

where the coefficients a_k are the coefficients for the AR model for the process $x[n]$. The FPE represents the output of our filter. We can write the *zeta* transform for the FPE at each stage p for the filter of order P as

$$FPE(z) = F_p^f[z]X[z] = \left(1 + \sum_{j=1}^{p} a_j^{(p)} z^{-j}\right) X[z]\,. \tag{23}$$

At each stage p of the Durbin algorithm the coefficients a_p are updated as

$$a_j^{(p)} = a_j^{(p-1)} + k_p a_{p-j}^{(p-1)} \qquad 1 \leq j \leq p-1\ . \tag{24}$$

If we use the above relation for the transform $F_p^f[z]$, we obtain

$$F_p^f[z] = F_{p-1}^f[z] + k_p \left[z^{-p} + \sum_{j=1}^{p-1} a_{p-j}^{(p-1)} z^{-j} \right]\ . \tag{25}$$

Now we introduce in a natural way the *backward* error of prediction BPE

$$F^b_{p-1}[z] = z^{-(p-1)} + \sum_{j=1}^{p-1} a^{(p-1)}_{p-j} z^{-(j-1)} \ . \tag{26}$$

In order to understand the meaning of $F^b_p[z]$ let us see its action in the time domain

$$F^b_{p-1}[z]x[n] = e^b_{p-1}[n] = x[n-p+1] + \sum_{j=1}^{p-1} a^{(p-1)}_{p-j} x[n-j+1] \, . \tag{27}$$

So $e^b_{p-1}[n]$ is the error we make, in a backward way, in the prediction of the data $x[n-p+1]$ using $p-1$ successive data $\{x[n], x[n-1], \dots , x[n-p+2]\}$. We can write the eq. (25) using $F^b_{p-1}[z]$. Let us substitute this relation in the z–transform of the filter $F^f_p[z]$

$$F^f_p[z] = F^f_{p-1}[z] + k_p F^b_{p-1}[z]. \tag{28}$$

In order to know the FPE filter at the stage p we must know the BPE filter at the stage $p-1$.

Also for the *backward* error we may write in a similar way the relation

$$F^b_p[z] = z^{-1} F^b_{p-1}[z] + k_p F^f_{p-1}[z] \ . \tag{29}$$

The equations (28) (29) represent our lattice filter that in the time domain could be written

$$e^f_p[n] = e^f_{p-1}[n] + k_p e^b_{p-1}[n-1] \, , \tag{30}$$

$$e^b_p[n] = e^b_{p-1}[n-1] + k_p e^f_{p-1}[n] \, . \tag{31}$$

In figure 3 is showed how the lattice structure is used to estimate the forward and backward errors.

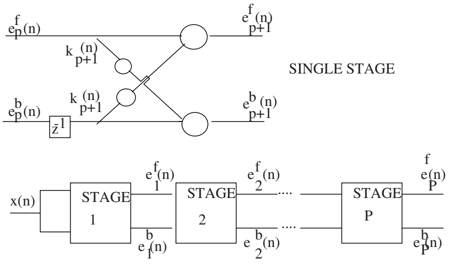


Fig. 3. Lattice structure for Durbin filter.

Using a lattice structure we can implement the whitening filter following these steps:

- estimate the values of the autocorrelation function $\hat{r}_{xx}[k], 0 \leq k \leq P$ of our process $x[n]$;
- use the Durbin algorithm to find the reflection coefficients $k_p, 1 \leq p \leq P$;
- implementation of the lattice filter with these coefficients k_p initiating the filter $e_0^f[n] = e_0^b[n] = x[n]$.

In this way the forward error at the stage P-th is equivalent to the forward error of a transversal filter and represents the output of the whitening filter.

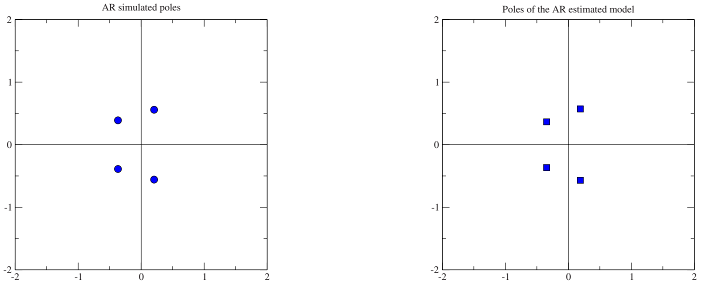
6 An example of whitening

We will show an example of the whitening procedure in time domain. First of all, we generate a stochastic process in time domain, using an AR(4) model with the values for the parameters reported in table 1.

σ	a_1	a_2	a_3	a_4
0.01	0.326526	0.338243	0.143203	0.101489

Table 1. Parameters for the simulated noise

These parameters describe a transfer function which is stable and causal, since all the poles are inside the unit circle (see Fig. 4). We perform an AR(4) fit to this noise, estimating the reflection coefficients for the whitening filter and the AR parameter for the PSD fit. We obtain a stable and causal fitted filter. In fact in figure 4 we reported in the complex plane the poles obtained using the Durbin algorithm: they are all inside the unit circle.


AR simulated poles

Poles of the AR estimated model

Fig. 4. Poles for the simulated AR model and poles for the estimated AR fit

In figure 5 we report the PSD of this noise process and the AR fit. It is evident that the fit reproduces the features of the noise (for realistic examples see [1, 2, 3]).

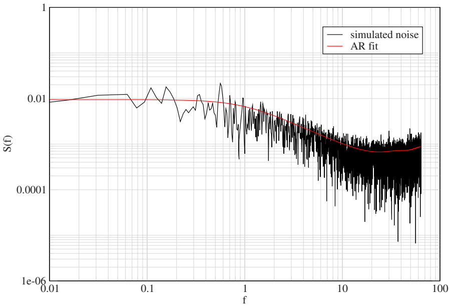


Fig. 5. Simulated power spectral density and AR(4) fit

In figure 6 we show the PSD of the simulated noise process and the PSD of the output of the whitening filter applied in the time-domain, using the estimated reflection coefficients.

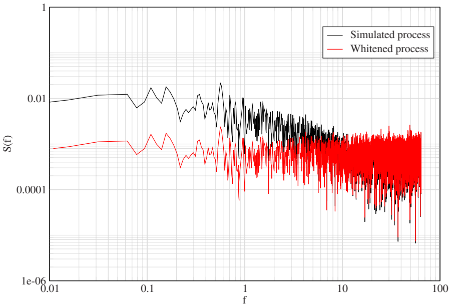


Fig. 6. Simulated power spectral density and PSD of the whitened data

The PSD of the output of the whitening filter, as we expected, is flat.

References

1. M. Beccaria, E. Cuoco, G. Curci, *Adaptive System Identification of VIRGO-like noise spectrum*, *Proc. of 2nd Amaldi Conference*, World Scientific, 1997.
2. E. Cuoco et al, Class.Quant.Grav., 18, 1727-1752, 2001.
3. E. Cuoco et al, Phys. Rev., D 64, 122002, 2001.
4. S. Kay, *Modern spectral estimation:Theory and Application*, Prentice Hall, Englewood Cliffs, 1998.
5. S. Haykin, *Adaptive Filter Theory*, (Upper Saddle River: Prentice Hall), 1996.
6. S.T. Alexander, *Adaptive Signal Processing*, (Berlin: Springer-Verlag), 1986.
7. M.H. Hayes, *Statistical Digital Signal Processing and Modeling*, Wiley, 1996.

8. C.W. Therrien, *Discrete Random Signals and Statistical Signal Processing* (Englewood Cliffs: Prentice Hall), 1992
9. L.A. Zubakov, V.D. Wainstein, *Extraction of signals fromnoise*, (Englewood Cliffs: Prentice Hall), 1962.
10. E. Parzen, *An approach to Time series modeling: determining the order of approximating autoregressive schemes in "Multivariate Analysis"*, North Holland, 1977.
11. B. Widrow, S. D. Stearns, *Adaptive Signal Processing*, (Englewood Cliffs: Prentice Hall), 1985.
12. S.J. Orfanidis, *Introduction to Signal Processing*, (Englewood Cliffs: Prentice-Hall), 1996.

The Laplace Transform in Control Theory

Martine Olivi

INRIA, BP 93,
06902 Sophia-Antipolis Cedex, FRANCE,
olivi@sophia.inria.fr

1 Introduction

The Laplace transform is extensively used in control theory. It appears in the description of linear time-invariant systems, where it changes convolution operators into multiplication operators and allows one to define the transfer function of a system. The properties of systems can be then translated into properties of the transfer function. In particular, causality implies that the transfer function must be analytic in a right half-plane. This will be explained in section 2 and a good reference for these preliminary properties and for a panel of concrete examples is [11].

Via Laplace transform, functional analysis provides a framework to formulate, discuss and solve problems in control theory. This will be sketched in section 3, in which the important notion of stability is introduced. We shall see that several kind of stability, with different physical meaning can be considered in connection with some function spaces, the Hardy spaces of the half-plane. These functions spaces provide with their norms a measure of the distance between transfer functions. This allows one to translate into well-posed mathematical problems some important topics in control theory, as for example the notion of robustness. A design is robust if it works not only for the postulated model, but also for neighboring models. We may interpret closeness of models as closeness of their transfer functions.

In section 4, we review the main properties of finite order linear time-invariant (LTI) causal systems. They are described by state-space equations and their transfer function is rational. We give the definition of the McMillan degree or order of a system, which is a good measure of its complexity, and some useful factorizations of a rational transfer function, closely connected with its pole and zero structure. Then, we consider the past inputs to future outputs map, which provides a nice interpretation of the notions of controllability and observability and we define the Hankel singular values. As claimed by Glover in [6], the Hankel singular values are extremely informative invari-

J.-D. Fournier et al. (Eds.): Harm. Analysis and Ratio. Approx., LNCIS 327, pp. 193–209, 2006.

ants when considering system complexity and gain. For this section we refer the reader to [8] and [6].

Section 5 is concerned with system identification. In many areas of engineering, high-order linear state-space models of dynamic systems can be derived (this can already be a difficult problem). By this way, identification issues are translated into model reduction problems that can be tackle by means of rational approximation. The function spaces introduced in section 3 provide with their norms a measure of the accuracy of a model. The most popular norms are the Hankel-norm and the L^2-norm. In these two cases, the role of the Hardy space H^2 with its Hilbert space structure, is determinant in finding a solution to the model reduction problem. In the case of the Hankel norm, explicit solutions can be found [6] while in the L^2 case, local minima can be numerically computed using gradient flow methods. Note that the approximation in L^2 norm has an interpretation in stochastic identification: it minimizes the variance of the output error when the model is fed by a white noise. These approximation problems are also relevant in the design of controllers which maximize robustness with respect to uncertainty or minimize sensitivity to disturbances of sensors, and other problems from H^∞ control theory. For an introduction to these fields we refer the reader to [4].

In this paper, we are concerned with continuous-time systems for which Laplace transform is a valuable aid. The z-transform performs the same task for discrete-time systems. This is the object of [3] in the framework of stochastic systems. It must be noted that continuous-time and discrete-time systems are related through a Möbius transform which preserves the McMillan degree [6]. For some purposes, it must be easier to deal with discrete-time. In particular, the poles of stable discrete-time systems lay in a bounded domain the unit circle. Laplace transform is also considered among other transforms in [12]. This paper also provides an introduction to [2].

2 Linear time-invariant systems and their transfer functions

Linear time-invariant systems play a fundamental role in signal and system analysis. Many physical processes possess these properties and even for nonlinear systems, linear approximations can be used for the analysis of small derivations from an equilibrium. Laplace transform has a number of properties that makes it useful for analysing LTI systems, thereby providing a set of powerful tools that form the core of signal and system analysis.

A continuous-time system is an "input-output" map

$$u(t) \to y(t),$$

from an input signal $u : \mathbb{R} \to \mathbb{C}^m$ to an output signal $y : \mathbb{R} \to \mathbb{C}^p$. It will be called linear if the map is linear and time-invariant if a time shift in the input signal results in an identical time shift in the output signal.

A linear time-invariant system can be represented by a convolution integral

$$y(t) = \int_{-\infty}^{\infty} h(t-\tau)u(\tau)\mathrm{d}\tau = \int_{-\infty}^{\infty} h(\tau)u(t-\tau)\mathrm{d}\tau,$$

in terms of its response to a unit impulse [11]. The $p \times m$ matrix function h is called the *impulse response* of the system.

The importance of complex exponentials in the study of LTI systems stems from the fact that the response of an LTI system to a complex exponential input is the same complex exponential with a change of amplitude. Indeed, for an input of the form $u(t) = \mathrm{e}^{st}$, the output computed through the convolution integral will be

$$y(t) = \int_{-\infty}^{\infty} h(\tau)\mathrm{e}^{s(t-\tau)}\mathrm{d}\tau = \mathrm{e}^{st}\int_{-\infty}^{\infty} h(\tau)\mathrm{e}^{-s\tau}\mathrm{d}\tau.$$

Assuming that the integral converges, the response to e^{st} is of the form

$$y(t) = H(s)\mathrm{e}^{st},$$

where $H(s)$ is the Laplace transform of the impulse response $h(t)$ defined by

$$H(s) = \int_{-\infty}^{\infty} h(\tau)\mathrm{e}^{-s\tau}\mathrm{d}\tau.$$

In the specific case in which $\mathrm{Re}\{s\} = 0$, the input is a complex integral $\mathrm{e}^{\mathrm{i}\omega t}$ at frequency ω and $H(\mathrm{i}\omega)$, viewed as a function of ω, is known as the *frequency response* of the system and is given by the Fourier transform

$$H(\mathrm{i}\omega) = \int_{-\infty}^{\infty} h(\tau)\mathrm{e}^{-\mathrm{i}\omega\tau}\mathrm{d}\tau.$$

In practice, pointwise measurements of the frequency response are often available and the classical problem of *harmonic identification* consists in finding a model for the system which reproduces these data well enough.

The *Laplace transform* of a scalar function $f(s)$

$$\mathcal{L}f(s) = \int_{-\infty}^{\infty} \mathrm{e}^{-st}f(t)\mathrm{d}t$$

is defined for those $s = x + \mathrm{i}y$ such that

$$\int_{-\infty}^{\infty} |f(\tau)|\mathrm{e}^{-x\tau}\mathrm{d}\tau < \infty.$$

The range of values of s for which the integral converges is called the region of convergence. It consists of strips parallel to the imaginary axis. In particular, if $f \in L^1(\mathbb{R})$, i.e.

$$\int_{-\infty}^{\infty} |f(t)|\mathrm{d}t < \infty,$$

then $\mathcal{L}f$ is defined on the imaginary axis and the Laplace transform can be viewed as a generalization of the Fourier transform.

Another obvious and important property of the Laplace transform is the following. Assume that $f(t)$ is right-sided, i.e. $f(t) = 0, \quad t < T$, and that the Laplace transform of f converges for $\mathrm{Re}\{s\} = \sigma_0$. Then, for all s such that $\mathrm{Re}\{s\} = \sigma > \sigma_0$, we have that

$$\int_{-\infty}^{\infty} |f(\tau)|\mathrm{e}^{-\sigma\tau}\mathrm{d}\tau = \int_{T}^{\infty} |f(\tau)|\mathrm{e}^{-\sigma\tau}\mathrm{d}\tau \leq \mathrm{e}^{-(\sigma-\sigma_0)T} \int_{T}^{\infty} |f(\tau)|\mathrm{e}^{-\sigma_0\tau}\mathrm{d}\tau,$$

and the integral converges so that Laplace transform is well defined in $\mathrm{Re}\{s\} \geq \sigma_0$. If $f \in L^1(\mathbb{R})$, then the Laplace transform is defined on the right half-plane and it can be proved that it is an analytic function there. It is possible that for some right-sided signal, there is no value of s for which the Laplace transform will converge. One example is the signal $h(t) = 0, \quad t < 0$ and $h(t) = \mathrm{e}^{t^2}, \quad t \geq 0$.

The importance of Laplace transform in control theory is mainly due to the fact that it allows to express any LTI system

$$y(t) = \int_{-\infty}^{\infty} h(t-\tau)u(\tau)\mathrm{d}\tau$$

as a multiplication operator

$$Y(s) = H(s)U(s),$$

where

$$Y(s) = \int_{-\infty}^{\infty} y(\tau)\mathrm{e}^{-s\tau}\mathrm{d}\tau, \quad H(s) = \int_{-\infty}^{\infty} h(\tau)\mathrm{e}^{-s\tau}\mathrm{d}\tau, \quad U(s) = \int_{-\infty}^{\infty} u(\tau)\mathrm{e}^{-s\tau}\mathrm{d}\tau,$$

are the Laplace transforms. The $p \times m$ matrix function $H(s)$ is called the *transfer function* of the system.

Causality is a common property for a physical system. A system is causal if the output at any time depends only on the present and past values of the input. A LTI system is causal if its impulse response satisfies

$$h(t) = 0 \quad \text{for} \quad t < 0,$$

and in this case, the output is given by the convolution integral

$$y(t) = \int_{0}^{\infty} h(\tau)u(t-\tau)\mathrm{d}\tau = \int_{0}^{t} h(t-\tau)u(\tau)\mathrm{d}\tau.$$

Then, the transfer function of the system is defined by the *unilateral Laplace transform*

$$H(s) = \int_0^{\infty} h(\tau)\mathrm{e}^{-s\tau}\mathrm{d}\tau, \tag{1}$$

whose region of convergence is, by what precedes, a right half-plane (if it is not empty). In the sequel, we shall restrict ourselves to causal systems.

Of course our signals must satisfy some conditions to ensure the existence of the Laplace transforms. There are many ways to proceed. We shall require our signals to belong to some spaces of integrable functions and this is closely related to the notion of stability of a system. This will be the object of the next section. Via Laplace transform, properties of an LTI system can be expressed in terms of the transfer function and by this way, function theory brings insights in control theory.

3 Function spaces and stability

An undesirable feature of a physical device is instability. In this section, we translate this into a statement about transfer functions. Intuitively, a stable system is one in which small inputs lead to responses that do not diverge. To give a mathematical statement, we need a measure of the size of a signal which will be provided by appropriate function spaces.

We denote by $L^q(X)$ the space of complex valued measurable functions f on X satisfying

$$\|f\|_q^q = \int_X |f(t)|^q \mathrm{d}t < \infty, \quad \text{if } 1 \leq q < \infty,$$

$$\|f\|_\infty = \sup_X |f(t)| < \infty, \quad \text{if } q = \infty.$$

The most natural measure is the L^∞ norm. A signal will be called bounded if there is some $M > 0$ such that

$$\|u\|_\infty = \sup_{t>0} \|u(t)\| < M,$$

where $\|.\|$ denotes the Euclidean norm of a vector. We still denote by $L^\infty(0,\infty)$ the space of bounded signals, omitting to mention the vectorial dimension. A system will be called *BIBO stable* if a bounded input produces a bounded output.

We may also be interested in the energy of a system which is given by the integral

$$\|u\|_2^2 = \int_0^\infty u(t)^* u(t)\mathrm{d}t.$$

We still denote by $L^2(0,\infty)$ the space of signal with bounded energy.

Notions of stability are associated with the requirement that the convolution operator

$$u(t) \to y(t) = h * u(t),$$

is a bounded linear operator, the input and output spaces being endowed with some (maybe different) norms. This implies that the transfer functions of such stable systems belong to some spaces of analytic functions, the Hardy spaces of the right half-plane [7]. We first introduce these spaces.

3.1 Hardy spaces of the half-plane

The Hardy space H^p is defined to be the space of functions $f(s)$ analytic in the right half-plane which satisfy

$$\|f\|_p := \sup_{0<x<\infty} \left\{ \int_{-\infty}^{\infty} |f(x+\mathrm{i}y)|^p \mathrm{d}y \right\}^{1/p} < \infty,$$

when $1 \leq p < \infty$, and, when $p = \infty$,

$$\|f\|_\infty := \sup_{\mathrm{Re}\{s\}>0} |f(s)| < \infty.$$

A theorem of Fatou says that, for any $f \in H^p$, $1 \leq p \leq \infty$,

$$f_0(\mathrm{i}y) = \lim_{x \to 0+} f(x+\mathrm{i}y),$$

exits a.e. on the imaginary axis. We may identify $f \in H^p$ with $f_0 \in L^p(\mathrm{i}\mathbb{R})$ and the identification is isometric, so that we may consider H^p as a subspace of $L^p(\mathrm{i}\mathbb{R})$. The case $p = 2$ is of particular importance since H^2 is an Hilbert space. We denote by H^2_- the left half-plane analog of H^2 : that is $f \in H^2_-$ if and only if the function $s \to f(-s)$ is in H^2. We may also consider H^2_- as a subspace of $L^2(\mathrm{i}\mathbb{R})$. We denote by Π^+ and Π^- the orthogonal projections from $L^2(\mathrm{i}\mathbb{R})$ to H^2 and H^2_- respectively, and we have

$$L^2(\mathrm{i}\mathbb{R}) = H^2 \oplus H^2_-.$$

If $f \in L^1(0,\infty)$, then $\mathcal{L}f$ is defined and analytic on the right half-plane. Moreover, we may extend the definition to functions $f \in L^2(0,\infty)$, since $L^1(0,\infty) \cap L^2(0,\infty)$ is dense in $L^2(0,\infty)$. The Laplace transform of a function $f \in L^2(0,\infty)$ is again defined and analytic on the right half-plane and we have the following theorem [13, Th.1.4.5]

Theorem 1. *The Laplace transform gives the following bijections*

$$\mathcal{L} : L^2(0,\infty) \to H^2,$$

$$\mathcal{L} : L^2(-\infty,0) \to H^2_-,$$

and for $f \in L^2(0,\infty)$ (resp. $L^2(-\infty,0)$)

$$\|\mathcal{L}f\|_2 = \sqrt{2\pi}\|f\|_2.$$

Since we are concerned with multi-input and multi-output systems, vectorial and matricial versions of these spaces are needed. For $p, m \in \mathbb{N}$, $H^\infty_{p\times m}$ and $H^2_{p\times m}$ are the spaces of $p \times m$ matrix functions with entries in H^∞ and H^2 respectively endowed with the norm

$$\|F\|_\infty = \sup_{-\infty<w<\infty} \|F(\mathrm{i}w)\| \tag{2}$$

$$\|F\|_2^2 = \mathrm{Tr} \int_{-\infty}^{\infty} F(\mathrm{i}w)^* F(\mathrm{i}w) \mathrm{d}w, \tag{3}$$

where $\|.\|$ denotes the Euclidean norm for a vector and for a matrix, the operator norm or spectral norm (that is the largest singular value). We shall often write H^∞, H^2 etc. for $H^\infty_{p\times m}$ and $H^2_{p\times m}$, the size of the matrix or vector functions (case $m = 1$) being understood from the context.

3.2 Some notions of stability

We shall study the notions of stability which arises from the following choices of norm on the input and output function spaces:

- **stability $L^\infty \to L^\infty$ (BIBO).** A system is BIBO stable if and only if its impulse response is integrable over $(0, \infty)$. Indeed, if $h(t)$ is integrable and $\|u\|_\infty < M$, then

$$\begin{aligned}\|y(t)\| &\le M \int_0^t \|h(t-\tau)\| \mathrm{d}\tau \\ &= M \int_0^t \|h(\tau)\| \mathrm{d}\tau, \\ &\le M \int_0^\infty \|h(\tau)\| \mathrm{d}\tau,\end{aligned}$$

  and $y(t)$ is bounded. Conversely, if $h(t)$ is not integrable, a bounded input can be constructed which produces an unbounded output (see [13] in the SISO case and [1, Prop.23.1.1] in the MIMO case).
- **stability $L^2 \to L^2$.** By Theorem 1 $\sqrt{2\pi}\,\mathcal{L}$ is a unitary operator from $L^2(0, \infty)$ onto the Hardy space H^2. Thus a system

$$y(t) = h * u(t),$$

  will be $L^2 \to L^2$ stable if its transfer function H is a bounded operator from H^2 to H^2. Now, the transfer function is a multiplication operator

$$M_H : U(s) \to Y(s),$$

  whose operator norm is $\|H\|_\infty$ given by (2) and H must belong to the Hardy space H^∞.

- **stability $L^2 \to L^\infty$.** The interest of this notion of stability comes from the fact that it requires that the transfer function $H(s)$ belongs to the Hardy space H^2 which is an Hilbert space. Indeed, it can be proved that the impulse response of such a stable system must be in $L^2(0,\infty)$ and thus by Theorem 1 its transfer function must be H^2.

4 Finite order LTI systems and their rational transfer functions

Among LTI systems, of particular interest are the systems governed by differential equations

$$\begin{aligned} \dot{x}(t) &= A\,x(t) + B\,u(t) \\ y(t) &= C\,x(t) + D\,u(t), \end{aligned} \tag{4}$$

where A, B, C, D are constant complex matrices matrices of type $n \times n$, $n \times m$, $p \times n$ and $p \times m$, and $x(t) \in \mathbb{C}^n$ is the state of the system. Assuming $x(0) = 0$, the solution is

$$\begin{aligned} x(t) &= \int_0^t \mathrm{e}^{(t-\tau)A} Bu(\tau)\mathrm{d}\tau, \quad t \geq 0 \\ y(t) &= \int_0^t C\mathrm{e}^{(t-\tau)A} Bu(\tau)\mathrm{d}\tau + Du(t), \quad t \geq 0 \end{aligned}$$

and the impulse response given by

$$g(t) = C\mathrm{e}^{At}B + D\delta_0,$$

where δ_0 is the delta function or Dirac measure at 0. Thus g is a generalized function.

As previously, we denote by the capital roman letter the Laplace transform of the function designated by the corresponding small letter. Laplace transform possesses the nice property to convert differentiation into a shift operator

$$\mathcal{L}\dot{x}(s) = sX(s).$$

so that the system (4) takes the form

$$\begin{aligned} sX(s) &= A\,X(s) + B\,U(s) \\ Y(s) &= C\,X(s) + D\,U(s), \end{aligned} \tag{5}$$

and yields

$$Y(s) = [D + C(sI - A)^{-1}B]U(s),$$

where $G(s) = D + C(sI - A)^{-1}B$ is the transfer function of the system. It is remarkable that transfer functions of LTI systems are rational.

Conversely, if the transfer function of a LTI system is rational and proper (its value at infinity is finite), then it can be written in the form (see [1])

$$G(s) = D + C(sI - A)^{-1}B.$$

We call (A, B, C, D) a realization of G and the system then admits a "state-space representation" of the form (4). A rational transfer function has many realizations. If T is a non-singular matrix, then $(TAT^{-1}, TB, T^{-1}C, D)$ is also a realization of $G(s)$. A *minimal realization* of G is a realization in which the size of A is minimal among all the realizations of G. The size n of A in a minimal realization is called the *McMillan degree* of $G(s)$. It represents the minimal number of state variables and is a measure of the complexity of the system.

For finite order systems all the notions of stability agree: a system is stable if and only if all the eigenvalues of A lie in the left half-plane.

To end with this section, we shall answer to some natural questions concerning these rational matrix functions: what is a pole? a zero? their multiplicity ? what could be a fractional representation?

Let $G(s)$ be a rational $p \times m$ matrix function. Then $G(s)$ admits the Smith form

$$G(s) = U(s)D(s)V(s),$$

where $U(s)$ and $V(s)$ are square size polynomial matrices with constant non-zero determinant and $D(s)$ is a diagonal matrix

$$D(s) = \operatorname{diag}\left(\frac{\phi_1}{\psi_1}, \frac{\phi_2}{\psi_2}, \dots, \frac{\phi_r}{\psi_r}, 0, \dots, 0\right)$$

in which for $i = 1, \dots r$, ϕ_i and ψ_i are polynomials satisfying the divisibility conditions

$$\phi_1/\phi_2/\dots/\phi_r,$$
$$\psi_r/\psi_{r-1}/\dots/\psi_1.$$

This representation exhibits the *pole-zeros structure* of a rational matrix. A zero of $G(s)$ is a zero of at least one of the polynomial ϕ_i. The multiplicity of a given zero in each of the ϕ_i is called a partial multiplicity and the sum of the partial multiplicities is the multiplicity of the zero. In the same way, the poles of $G(s)$ are the zeros of the ψ. They are also the eigenvalues of the dynamic matrix A. It must be noticed that a complex number can be a pole and a zero at the same time. For more details on that Smith form, see [8]. It provides a new interpretation of the McMillan degree as the number of poles of the rational function counted with multiplicity, i.e. the degree of $\psi = \psi_1\psi_2 \cdots \psi_r$.

The Smith form also allows one to write a left coprime *polynomial factorization* (see [1, Chap.11] or [8]) of the form

$$G(s) = D(s)^{-1}N(s),$$

where $D(s)$ and $N(s)$ are left coprime polynomial matrices, i.e.

$$D(s)E_1(s) + N(s)E_2(s) = I, \quad s \in \mathbb{C},$$

for some polynomial matrices $E_1(s)$ and $E_2(s)$. In this factorization the matrix $D(s)$ brings the pole structure of $G(s)$ and the matrix $N(s)$ its zero structure.

This representation is very useful in control theory. In our function spaces context another factorization is more natural. It is the *inner-unstable or Douglas-Shapiro-Shields factorization*

$$G(s) = Q(s)P(s),$$

where $Q(s)$ is an inner function in H^∞, i.e. such that

$$Q(\mathrm{i}w)^* Q(\mathrm{i}w) = I, \quad w \in \mathbb{R},$$

and $P(s)$ is unstable (analytic in the left half-plane). We shall also require this factorization to be minimal. It is then unique up to a common left constant unitary matrix and the McMillan degree of Q is the McMillan degree of G. The existence of such a factorization follows from Beurling's theorem on shift invariant subspaces of H^2 [5]. Here again, the inner factor brings the pole structure of the transfer function and the unstable factor the zero structure. In many approximation problems this factorization allows to reduce the number of optimization parameters, since the unstable factor can often be computed from the inner one. This makes the interest of inner function together with the fact that inner functions are the transfer function of conservative systems.

4.1 Controllability, observability and associated gramians

The notions of controllability and observability are central to the state-space description of dynamical systems. Controllability is a measure for the ability to use a system's external inputs to manipulate its internal state. Observability is a measure for how well internal states of a system can be inferred by knowledge of its external outputs.

The following facts are well-known [8]. A system described by a state-space realization (A, B, C, D) is controllable if the pair (A, B) is controllable, i.e. the matrix

$$\begin{bmatrix} B \; AB \; A^2B \cdots \; A^{n-1}B \end{bmatrix}$$

has rank n, and the pair (C, A) observable, i.e. the matrix

$$\begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{n-1} \end{bmatrix}$$

has rank n. A realization is minimal if and only if it is both controllable and observable. Note that the matrix D play no role in this context.

We now give an alternative description of these notions which is more adapted to our functional framework [6, Sect.2]. If the eigenvalues of A are assumed to be strictly in the left half-plane, then we can define the controllability gramian as

$$P = \int_0^\infty \mathrm{e}^{At} B B^* \mathrm{e}^{A^* t} \mathrm{d}t,$$

and the observability gramian as

$$Q = \int_0^\infty \mathrm{e}^{A^* t} C^* C \mathrm{e}^{At} \mathrm{d}t.$$

It is easily verified that P and Q satisfy the following Lyapunov equations

$$\begin{aligned} AP + PA^* + B^*B &= 0, \\ A^*Q + QA + C^*C &= 0. \end{aligned}$$

A standard result is that the pair (A, B) is controllable if and only if P is positive definite and the pair (C, A) observable if and only if Q is positive definite.

These gramians can be illustrated by considering the mapping from the past inputs to the future outputs, $\gamma_g : L^2(-\infty, 0) \to L^2(0, \infty)$, given by

$$(\gamma_g u)(t) = \int_{-\infty}^0 C \mathrm{e}^{A(t-\tau)} B u(\tau) \mathrm{d}\tau = \int_0^\infty C \mathrm{e}^{A(t+\tau)} B v(\tau) \mathrm{d}\tau, \tag{6}$$

where $v(t) = u(-t)$ is in $L^2(0, \infty)$. The mapping γ_g can be view as a composition of two mappings:

$$u(t) \to x(0) = \int_0^\infty \mathrm{e}^{A\tau} B u(-\tau) \mathrm{d}\tau,$$

and

$$x(0) \to y(t) = C \mathrm{e}^{At} x(0),$$

where $x(0)$ is the state at time $t = 0$. Now, consider the following minimum energy problem

$$\min_{u \in L^2(-\infty,0)} \|u\|_2^2 \quad \text{subject to} \quad x(0) = x_0.$$

Since x_0 is a linear function of $u(t)$, the solution $\hat{u}$ exists provided that P is positive definite and is given by the pseudo-inverse

$$\hat{u}(t) = B^* \mathrm{e}^{-A^* t} P^{-1} x_0.$$

It satisfies

$$\|\hat{u}\|_2^2 = x_0^* P^{-1} x_0.$$

If P^{-1} is large, there will be some state that can only be reached if a large input energy is used. If the system is realized from $x(0) = x_0$ with $u(t) = 0, \quad t \geq 0$ then

$$\|y\|_2^2 = x_0^* Q x_0,$$

so that, if the observability gramian Q is nearly singular then some initial conditions will have little effect on the output.

4.2 Hankel singular values and Hankel operator

We now introduce the Hankel singular values which turn out to be fundamental invariants of a linear system related to both gain and complexity [6]. The link with complexity will be further illustrated in section 5.1.

The problem of approximating a matrix by a matrix of lower rank was one of the earliest application of the singular-value decomposition ([10], see [6, Prop.2.2] for a proof).

Proposition 1. *Let $M \in \mathbb{C}^{p\times m}$ have singular value decomposition given by*

$$M = UDV,$$

where U, V are square unitary and $D = \begin{bmatrix} D_r & 0 \\ 0 & 0 \end{bmatrix}$, $D_r = \mathrm{diag}(\alpha_1, \alpha_2, \dots, \alpha_r)$, where $\alpha_1 \geq \alpha_2 \dots \geq \alpha_r > 0$ are the singular values of M. Then,

$$\inf_{\text{rank } \hat{M} \leq k} \|M - \hat{M}\| = \alpha_{k+1},$$

and the bound is achieved by

$$\hat{D}_k = \begin{bmatrix} D_r & 0 \\ 0 & 0 \end{bmatrix}, \quad D_k = \mathrm{diag}(\alpha_1, \alpha_2, \dots, \alpha_k).$$

This result can be generalized to the case of a bounded linear operator $T \in \mathcal{L}(\mathcal{H}, \mathcal{K})$ from an Hilbert space $\mathcal{H}$, to another, $\mathcal{K}$. For $k = 0, 1, 2, \dots$, the kth *singular value* $\sigma_k(T)$ of T is defined by

$$\sigma_k(T) = \inf\{\|T - R\|, R \in \mathcal{L}(\mathcal{H}, \mathcal{K}), \quad \text{rank } R \leq k\}.$$

Thus $\sigma_0(T) = \|T\|$ and

$$\sigma_0(T) \geq \sigma_1(T) \geq \sigma_2(T) \geq \dots \geq 0.$$

When T is compact, it can be proved that $\sigma_k(T)^2$ is an eigenvalue of T^*T [15, Th.16.4]. Any corresponding eigenvector of T^*T is called a *Schmidt vector* of T corresponding to the singular value $\sigma_k(T)$. A *Schmidt pair* is a pair of vectors $x \in \mathcal{H}$ and $y \in \mathcal{K}$ such that

$$Tx = \sigma_k(T)y, \quad T^*y = \sigma_k(T)x.$$

The past inputs to future outputs mapping γ_g associated with a LTI system by (6) is a compact operator from $L^2(-\infty, 0)$ to $L^2(0, \infty)$. The *Hankel singular values* of a LTI system are defined to be the singular values of γ_g. Via the Laplace transform, we may associate with γ_g, the Hankel operator

$$\Gamma_G : H^2_- \to H^2,$$

whose symbol G is the Laplace transform of g. It is defined by

$$\Gamma_G(x) = \Pi_+(Gx), x \in H^2_-.$$

Since γ_g and Γ_G are unitarily equivalent via the Laplace transform, they share the same set of singular values

$$\sigma_0(G) \geq \sigma_1(G) \geq \sigma_2(G) \geq \cdots \geq 0.$$

The Hankel norm is defined to be the operator norm of Γ_G , which turns out to be its largest singular value $\sigma_0(G)$:

$$\|G\|_H = \|\Gamma_G\| = \sigma_0(G).$$

Note that

$$\|G\|_H = \sup_{u \in L^2(-\infty,0)} \frac{\|y\|_{L^2(0,\infty)}}{\|u\|_{L^2(-\infty,0)}},$$

so that the Hankel norm gives the L^2 gain from past inputs to future outputs.

If the LTI system has finite order, then its Hankel singular values correspond to the singular values of the matrix PQ, where P is controllability gramian and Q the observability gramian. Indeed, let σ be a singular value of γ_g with u the corresponding eigenvector of $\gamma_g^*\gamma_g$: $(\gamma_g^*\gamma_g u)(t) = \sigma^2 u(t)$. Then, since the adjoint operator γ_g^* is given by

$$(\gamma_g^* y)(t) = \int_0^\infty B^* \mathrm{e}^{A^*(-t+\tau)} C^* y(\tau) \mathrm{d}\tau,$$

we have that

$$(\gamma_g^*\gamma_g u)(t) = (\gamma_g^* y)(t) = B^* \mathrm{e}^{-A^* t} Q x_0,$$

so that

$$u(t) = \sigma^{-2} B^* \mathrm{e}^{-A^* t} Q x_0. \tag{7}$$

Now,

$$\sigma^2 x_0 = \int_0^\infty \mathrm{e}^{(A\tau)} B \sigma^2 u(-\tau) \mathrm{d}\tau = PQx_0,$$

and σ^2 is an eigenvalue of PQ associated with the eigenvector x_0. Conversely, if σ^2 is an eigenvalue of PQ associated with the eigenvector x_0, then σ is a singular value of γ_g with corresponding eigenvector of $\gamma_g^* \gamma_g$ given by (7). A useful state-space realization in this respect is the balanced realization for which $P = Q = \mathrm{diag}(\sigma_0, \sigma_1, \dots, \sigma_{n-1})$.

Remark 1. The Hankel norm of a finite order LTI system doesn't depend on its 'D matrix'.

5 Identification and approximation

The identification problem is to find an accurate model of an observed system from measured data. This definition covers many different approaches depending on the class of models we choose and on the data we have at hand. We shall pay more attention on harmonic identification. The data are then pointwise values of the frequency response in some bandwidth and the models are finite order linear time-invariant (LTI) systems. A robust way to proceed is to interpolate the data on the bandwidth into a high order transfer function, possibly unstable. A first step consists in approximating the unstable transfer function by a stable one. This can be done by solving bounded extremal problems (see [2]).

For computational reasons, it is desirable if such a high-order model can be replaced by a reduced-order model without incurring to much error. This can be stated as follows:

Model reduction problem: given a $p \times m$ stable rational matrix function $G(z)$ of McMillan degree N, find $\hat{G}$ stable of McMillan degree $n < N$ which minimizes

$$|||G - \hat{G}|||. \tag{8}$$

The choice of the norm $|||.|||$ is influenced by what norms can be minimized with reasonable computational efforts and whether the chosen norm is an appropriate measure of error. The most natural norm from a physical viewpoint is the norm $\|.\|_\infty$. But this is an unresolved problem: there is no known numerical method which is guaranteed to converge. In Banach spaces other than Hilbert spaces, best approximation problems are usually difficult. There are two cases in which the situation is easier since they involve the Hardy space H^2 which is an Hilbert space: the L^2-norm and the Hankel norm, since the Hankel operator acts on H^2. In this last case an explicit solution can be computed.

5.1 Hankel-norm approximation

In the seventies, it was realized that the recent results on L^∞ approximation problems, such as Nehari's theorem and the result of Adamjan, Arov and Krein on the Nehari-Takagi problem, were relevant to the current problems of some engineers in control theory. In the context of LTI systems, they have led to efficient new methods of model reduction.

A first step in solving the model reduction problem in Hankel-norm is provided by Nehari's theorem. Translated in the control theory framework, it states that if one wishes to approximate a causal function $G(s)$ by an anticausal function, then the smallest error norm that can be achieved is precisely the Hankel-norm of $G(s)$.

Theorem 2. *For $G \in H^\infty$*

$$\sigma_0(G) = \|G\|_H = \inf_{F \in H^\infty_-} \|G - F\|_\infty .$$

The model reduction problem, known under the name of Nehari-Takagi, was first solved by Adamjan, Arov and Krein for SISO systems and Kung and Lin for MIMO discrete-time systems. In our continuous-time framework, it can be stated as follow:

Theorem 3. *Given a stable, rational transfer function $G(s)$ then*

$$\sigma_k(G) = \inf_{\hat{G} \in H^\infty} \|G - \hat{G}\|_H, \quad \textit{McMillan degree of } \hat{G} \leq k.$$

The fact that $\|G - \hat{G}\|_H \geq \sigma_k(G)$, for all $\hat{G}(s)$ stable and of McMillan degree $\leq k$, is no more than a continuous-time version of Proposition 1 [6, Lemma 7.1]. This famous paper [6] gives a beautiful solution of the computational problem using state-space methods. An explicit construction of a solution $\hat{G}(s)$ is presented which makes use of a balanced realization of $G(s)$ [6, Th.6.3]. Moreover, in [6] all the optimal Hankel norm approximations are characterized in state-space form.

Since,

$$\|G - \hat{G}\|_H = \inf_{F \in H^\infty_-} \|G - \hat{G} - F\|_\infty ,$$

the Hankel norm approximation $\hat{G}(s)$ can be a rather bad approximant in L^∞ norm. However, the choice of the 'D matrix' for the approximation is arbitrary, since the Hankel-norm doesn't depend on D, while $\|G - \hat{G}\|_\infty$ does depend on D. In [6, Sect.9, Sect.10.2] a particular choice of D is suggested which ensures that

$$\|G - \hat{G}\|_\infty \leq \sigma_k(G) + \sum_{j>k} \sigma_j(G).$$

It is often the case in practical applications that Γ_G has a few sizable singular values and the remaining ones tail away very quickly to zero. In that case the right hand-side can be made very small, and one is assured that an optimal Hankel norm approximant is also good with respect to the L^∞ norm.

5.2 L^2-norm approximation

In the case of the L^2 norm, an explicit solution of the model reduction problem cannot be computed. However, the L^2 norm being differentiable we may think of using a gradient flow method. The main difficulty in this problem is to describe the set of approximants, i.e. of rational stable functions of McMillan degree n. The approaches than can be found in the literature mainly differ from the choice of a parametrization to describe this set of approximants. These parametrizations often arise from realization theory and the parameters are some entries of the matrices (A, B, C, D). To cope with their inherent complexity, some approaches choose to relax a constraint: stability or fixed McMillan degree. They often run into difficulties since smoothness can be lost or an undesirable approximant reached.

Another approach can be proposed. The number of optimization parameters can be reduced using the inner-unstable factorization (see section 4) and the projection property of an Hilbert space. Let $\hat{G}$ be a best L^2 approximant of G, with inner-unstable factorization

$$\hat{G} = QP,$$

where Q is the inner factor and P the unstable one. Then, H^2 being an Hilbert space, $\hat{G}$ must be the projection of G onto the space $H(Q)$ of matrix functions of degree n whose left inner factor is Q. We shall denote this projection by $\hat{G}(Q)$ and the problem consists now in minimizing

$$Q \to \|G - \hat{G}(Q)\|_2,$$

over the set of inner functions of McMillan degree n.

Then, more efficient parametrizations can be used which arise from the manifold structure of this set. It consists to work with an atlas of charts, that is a collection of local coordinate maps (the charts) which cover the manifold and such that changing from one map to another is a smooth operation. Such a parametrization present the advantages to ensure identifiability, stability of the result and the nice behavior of the optimization process. The optimization is run over the set as a whole changing from one chart to another when necessary. Parametrizations of this type are available either from realization theory or from interpolation theory in which the parameters are interpolation values. Their description goes beyond the aim of this paper, and we refer the reader to [9] and the bibliography therein for more informations on this approach.

References

1. J.A. Ball, I. Gohberg, L. Rodman. *Interpolation rational matrix functions*, Birkhäuser, Operator Theory: Advances and Applications, 1990, vol. 45.
2. L. Baratchart. *Identification an Function theory.* This volume, pages 211ff.
3. M. Deistler. *Stationary Processes and Linear Systems.* This volume, pages 159ff.
4. B.A. Francis. *A course in H^∞ control theory*, Springer, 1987.
5. P.A. Fhurmann. *Linear systems and operators in Hilbert Spaces*, McGraw-Hill, 1981.
6. K. Glover. *All optimal Hankel norm approximations of linear multivariable systems and their L^∞ error bounds*, Int. J. Control, 39(6):1115-1193, 1984.
7. K. Hoffman. *Banach spaces of analytic functions*, Dover publications, New York, 1988.
8. T. Kailath. *Linear systems*, Prentice-Hall, 1980.
9. J.-P. Marmorat, M. Olivi, B. Hanzon, R.L.M. Peeters. *Matrix rational H^2 approximation: a state-space approach using Schur parameters*, in Proceedings of the CDC02, Las-Vegas, USA.
10. L. Mirsky. *Symmetric gauge functions and unitarily invariant norms*, Quart. J. Math. Oxford Ser. 2(11):50-59, 1960.
11. A.V. Oppenheim, A.S. Willsky, S.H. Nawab. *Signals and Systems*, Prentice-Hall, 1997.
12. J.R. Partington. *Fourier transforms and complex analysis.* This volume, pages 39ff.
13. J.R. Partington. *Interpolation, identification and sampling*, Oxford University Press, 1997.
14. W. Rudin. *Real and complex analysis*, New York, McGraw-Hill, 1987.
15. N.J. Young. *An introduction to Hilbert space*, Cambridge University Press, 1988.
16. N.J. Young. *The Nehari problem and optimal Hankel norm approximation*, Analysis and optimization of systems: state and frequency domain approach for infinite dimensional systems, Proceedings of the 10th International Conference, Sophia-Antipolis, France, June 9-12, 1992.

Identification and Function Theory

Laurent Baratchart

INRIA, BP 93, 06902 Sophia-Antipolis Cedex, FRANCE
laurent.baratchart@sophia.inria.fr

1 Introduction

We survey in these notes certain constructive aspects of how to recover an analytic function in a plane domain from complete or partial knowledge of its boundary values. This we do with an eye on identification issues for linear dynamical systems, i.e. one-dimensional deconvolution schemes, and for that reason we restrict ourselves either to the unit disk or to the half-plane because these are the domains encountered in this context. To ensure the existence of boundary values, restrictions on the growth of the function must be made, resulting in a short introduction to Hardy spaces in the next section. We hasten to say that, in any case, the problem just mentioned is ill-posed in the sense of Hadamard [32], and actually a prototypical inverse problem: the Cauchy problem for the Laplace equation. We approach it as a constrained optimization issue, which is one of the classical routes when dealing with ill-posedness [51]. There are of course many ways of formulating such issues; those surveyed below make connection with the quantitative spectral theory of Toeplitz and Hankel operators that are deeply linked with meromorphic approximation. Standard regularization, which consists in requiring additional smoothness on the approximate solution, would allow us here to use classical interpolation theory; this is not the path we shall follow, but we warn the reader that linear interpolation schemes are usually not so extremely efficient in the present context. An excellent source on this topic and other matters related to our subject is [39].

2 Hardy spaces

Let $\mathbb{T}$ be the unit circle and $\mathbb{D}$ the unit disk in the complex plane. We let $C(\mathbb{T})$ denote continuous functions and $L^p = L^p(\mathbb{T})$ the familiar Lebesgue spaces. For $1 \leq p \leq \infty$, the Hardy space H^p of the unit disk is the closed subspace of L^p consisting of functions whose Fourier coefficients of strictly negative index do

J.-D. Fournier et al. (Eds.): Harm. Analysis and Ratio. Approx., LNCIS 327, pp. 211–230, 2006.

vanish. These are the nontangential limits of functions g analytic in the unit disk $\mathbb{D}$ having uniformly bounded L^p means over all circles centered at 0 of radius less than 1:

$$\sup_{0<r<1} \|g(r\mathrm{e}^{\mathrm{i}\theta})\|_p < \infty. \tag{1}$$

The correspondence is one-to-one and, using this identification, we alternatively regard members of H^p as holomorphic functions in the variable $z \in \mathbb{D}$. The extension to $\mathbb{D}$ is obtained from the values on $\mathbb{T}$ through a Cauchy as well as a Poisson integral, namely if $g \in H^p$ then:

$$g(z) = \frac{1}{2\mathrm{i}\pi} \int_{\mathbb{T}} \frac{g(\xi)}{\xi - z}\,\mathrm{d}\xi, \quad z \in \mathbb{D}, \tag{2}$$

and also

$$g(z) = \frac{1}{2\pi} \int_{\mathbb{T}} \mathrm{Re}\left\{\frac{\mathrm{e}^{\mathrm{i}\theta} + z}{\mathrm{e}^{\mathrm{i}\theta} - z}\right\} g(\mathrm{e}^{\mathrm{i}\theta})\mathrm{d}\theta, \quad z \in \mathbb{D}. \tag{3}$$

The sup in (1) is precisely $\|g(\mathrm{e}^{\mathrm{i}\theta})\|_p$. The space H^∞ consists of bounded analytic functions in $\mathbb{D}$, and by Parseval's theorem we also get that

$$g(z) \in H^2 \quad \text{iff} \quad f(z) = \sum_{j=0}^{\infty} a_k z^k, \quad \text{with} \quad \sum_{j=0}^{\infty} |a_j|^2 < \infty.$$

If $p \neq 2$ it is not easy to characterize H^p functions from their Fourier-Taylor coefficients. Very good expositions on Hardy spaces are [19, 22, 30], and we recall just a few facts here. Actually, we only work with $p = 2$ and $p = \infty$, but nothing would be gained in this section from such a restriction.

A nonzero $g \in H^p$ can be uniquely factored as $g = jw$ where

$$w(z) = \exp\left\{\frac{1}{2\pi} \int_0^{2\pi} \frac{\mathrm{e}^{\mathrm{i}\theta} + z}{\mathrm{e}^{\mathrm{i}\theta} - z} \log|f(\mathrm{e}^{\mathrm{i}\theta})|\mathrm{d}\theta\right\} \tag{4}$$

belongs to H^p and is called the *outer factor* of g, while $j \in H^\infty$ has modulus 1 a.e. on $\mathbb{T}$ and is called the *inner factor* of g. The latter may be further decomposed as $j = bS_\mu$, where

$$b(z) = cz^k \prod_{z_l \neq 0} \frac{-\bar{z}_l}{|z_l|} \frac{z - z_l}{1 - \bar{z}_l z} \tag{5}$$

is the *Blaschke product*, with order $k \geq 0$ at the origin, associated to a sequence of points $z_l \in \mathbb{D} \setminus \{0\}$ and to the constant $c \in \mathbb{T}$, while

$$S_\mu(z) = \exp\left\{-\frac{1}{2\pi} \int_0^{2\pi} \frac{\mathrm{e}^{\mathrm{i}\theta} + z}{\mathrm{e}^{\mathrm{i}\theta} - z}\,\mathrm{d}\mu(\theta)\right\} \tag{6}$$

is the *singular inner factor* associated with μ, a positive measure on $\mathbb{T}$ which is singular with respect to Lebesgue measure. The z_l are the zeros of g in $\mathbb{D}\setminus\{0\}$, counted with their multiplicities, while k is the order of the zero at 0. If there are infinitely many zeros, the convergence of the product $b(z)$ in $\mathbb{D}$ is ensured by the condition $\sum_l(1-|z_l|)<\infty$ which holds automatically when $g\in H^p\setminus\{0\}$. If there are only finitely many z_l, say n, we say that (5) is a finite Blaschke product of degree n.

That $w(z)$ in (4) is well-defined rests on the fact that $\log|g|\in L^1$ if $f\in H^1\setminus\{0\}$; this also entails that a H^p function cannot vanish on a set of strictly positive Lebesgue measure on $\mathbb{T}$ unless it is identically zero.

Intimately related to Hardy functions is the Nevanlinna class N^+ consisting of holomorphic functions in $\mathbb{D}$ that can be factored as jE, where j is an inner function, and E an outer function of the form

$$E(z)=\exp\left\{\frac{1}{2\pi}\int_0^{2\pi}\frac{\mathrm{e}^{\mathrm{i}\theta}+z}{\mathrm{e}^{\mathrm{i}\theta}-z}\log\varrho(\mathrm{e}^{\mathrm{i}\theta})\,\mathrm{d}\theta\right\}, \tag{7}$$

ϱ being a positive function on $\mathbb{T}$ such that $\log\varrho\in L^1(\mathbb{T})$, although ϱ itself need not be summable. Such functions again have nontangential limits a.e. on $\mathbb{T}$ that serve as definition for their boundary values, and they are often instrumental in that $N^+\cap L^p=H^p$. In fact, (7) defines an H^p-function with modulus ϱ a.e. on $\mathbb{T}$ if, and only if, $\varrho\in L^p$. A useful consequence is that, whenever $g_1\in H^{p_1}$ and $g_2\in H^{p_2}$, we have $g_1g_2\in H^{p_3}$ if, and only if, $g_1g_2\in L^{p_3}$.

We need also introduce the Hardy space $\bar{H}^p$ of the complement of the disk, consisting of L^p functions whose Fourier coefficients of strictly positive index do vanish; these are, a.e. on $\mathbb{T}$, the complex conjugates of H^p-functions, and they can also be viewed as nontangential limits of functions analytic in $\overline{\mathbb{C}}\setminus\overline{\mathbb{D}}$ having uniformly bounded L^p means over all circles centered at 0 of radius bigger than 1. We further single out the subspace $\bar{H}^p_0\subset\bar{H}^p$, consisting of functions vanishing at infinity or, equivalently, having vanishing mean on $\mathbb{T}$. Thus, a function belongs to $\bar{H}^p_0$ if, and only if, it is of the form $\mathrm{e}^{-\mathrm{i}\theta}\overline{g(\mathrm{e}^{\mathrm{i}\theta})}$ for some $g\in H^p$.

We let $\mathcal{R}_{m,n}$ be the set of rational functions of type (m,n) that can be written p/q where p and q are algebraic polynomials of degree at most m and n respectively. Note that a rational function belongs to some H^p if, and only if, its poles lie outside $\overline{\mathbb{D}}$, in which case it belongs to every H^p. Similarly, a rational function belongs to $\bar{H}^p$ if, and only if, it can be written as p/q with $\deg p\leq\deg q$ where q has roots in $\mathbb{D}$ only; in the language of system theory, such a rational function is called stable and proper, and it belongs to $\bar{H}^p_0$ if, and only if, $\deg p<\deg q$ in which case it is called strictly proper. We define H^p_n to be the set of meromorphic functions with at most n poles in $\mathbb{D}$, that may be written g/q where $g\in H^p$ and q is a polynomial of degree at most n with roots in $\mathbb{D}$ only.

We now turn to the Hardy spaces $\mathcal{H}^p$ of the right half-plane. These consist of functions G analytic in $\Pi_+=\{s;\ \operatorname{Re}s>0\}$ such that

$$\sup_{x>0} \int_{-\infty}^{+\infty} |G(x+\mathrm{i}y)|^p \, \mathrm{d}y < \infty,$$

and again they have nontangential limits at almost every point of the imaginary axis, thereby giving rise to a boundary function $G(\mathrm{i}y)$ that lies in $L^p(\mathrm{i}\mathbb{R})$. The space $\mathcal{H}^\infty$ consists of bounded analytic functions in Π_+, and a theorem of Paley-Wiener characterizes $\mathcal{H}^2$ as the space of Fourier transforms of functions in $L^2(\mathbb{R})$ that vanish for negative arguments.

The study of $\mathcal{H}^p$ can be reduced to that of $\bar{H}_0^p$ thanks to the isometry :

$$g \mapsto (s-1)^{-2/p} g\left(\frac{s+1}{s-1}\right) \tag{8}$$

from $\bar{H}_0^p$ onto $\mathcal{H}^p$. The latter preserves rationality and the degree for $p = 2, \infty$.

For applications to system-theory, it is often necessary to consider functions in H^p or $\mathcal{H}^p$ that have the conjugate-symmetry $g(\bar{z}) = \overline{g(z)}$; in the case of H^p this means they have real Fourier coefficients, or in the case of $\mathcal{H}^2$ that they are Fourier transforms of real functions. For rational functions it means that the coefficients of p and q are real in the irreducible form p/q. In the presence of conjugate symmetry, every symbol will be decorated by a subscript or a superscript “$\mathbb{R}$”, like in $H_{\mathbb{R}}^p$ or $\mathcal{R}_{m,n}^{\mathbb{R}}$ etc.

3 Motivations from System Theory

We provide some motivation from control and signal theory for some of the approximation problems that we will consider. The connection between linear control system and function theory has two cornerstones:

- the fact that these systems can be described in the so-called frequency domain as a multiplication operator by the transfer function which belongs to certain Hardy classes if the system has certain stability properties;
- the fact that rational functions are precisely transfer functions of systems having finite-dimensional state-space, namely those that can be designed and handled in practice.

A *discrete control system* is a map $u \to y$ where the input $u = (\ldots, u_{k-1}, u_k, u_{k+1}, \ldots)$ is a real-valued function of the discrete time k, generating an output $y = (\ldots, y_{k-1}, y_k, y_{k+1}, \ldots)$ of the same kind, where y_k depends on u_j for $j \leq k$ only. The system is said to be time-invariant if a shift in time of the input produces a corresponding shift of the output.

Particularly important in applications are the *linear systems*:

$$y_k = \sum_{j=0}^{\infty} f_j u_{k-j},$$

where the output at time k is a linear combination of the past inputs with fixed coefficients $f_j \in \mathbb{R}$. For such systems, function theory enters the picture when signals are encoded by their generating functions:

$$u(z) = \sum_{k\in\mathbb{Z}} u_k z^{-k}, \qquad y(z) = \sum_{k\in\mathbb{Z}} y_k z^{-k}.$$

Indeed, if we define the *transfer function* of the linear control system to be:

$$f(z) = \sum_{k=0}^{\infty} f_k z^{-k},$$

the input-output behavior can be described as $y(z) = f(z)\,u(z)$. In particular:

(i) the system is a bounded operator $l^2 \to l^2$ iff $f \in \bar{H}^\infty_{\mathbb{R}}$, and the operator norm is $\|f\|_\infty$: the system is called (l^2, l^2)-stable;
(ii) the system is a bounded operator $l^2 \to l^\infty$ iff $f \in \bar{H}^2_{\mathbb{R}}$, and the operator norm is $\|f\|_2$: the system is called (l^2, l^∞)-stable.

A linear control system is said to have *finite dimension* n if it can be described as a linear *automata* in terms of a state variable $x_k \in \mathbb{R}^n$ which is updated at each time $k \in \mathbb{Z}$:

$$x_{k+1} = Ax_k + Bu_k, \qquad y_k = Cx_k + Du_k,$$

where A is a real $n \times n$ matrix, B (resp. C) a column (resp. row) vector with n real entries, and D some real number, n being the smallest possible integer for which such an equation holds. The classical result here (see e.g. [29], [42]) is that a linear time-invariant system has dimension n iff its transfer-function is rational of degree n and analytic at infinity. The transfer-function is then

$$f(z) = D + C\,(zI_n - A)^{-1}\,B.$$

For finite-dimensional linear systems the requirement that the poles of f should lie in $\{|z| < 1\}$ amounts to $f \in \bar{H}^p$ for some, and in fact all p, which is equivalent to any reasonable definition of stability. A much broader picture is obtained by letting input and output signals be vector-valued and transfer functions matrix-valued, and indeed many questions to come are significantly enriched by doing so; this, however, is beyond the scope of these notes. We now mention two specific applications of approximation theory in Hardy spaces to identification of linear dynamical systems. Further applications to control can be found in [20, 40].

3.1 Stochastic identification

Consider a discrete time real-valued stationary stochastic process:

$$y = (\dots, y(k-1), y(k), y(k+1), \dots).$$

If it is regular (i.e. purely non-deterministic in a certain sense), we have the Wold decomposition:

$$y(k) = \sum_{j=0}^{\infty} f_j u(k-j)$$

where u is a white noise called the innovation, and where f_j is independent of k by stationarity. We also have, by the Parseval identity, that

$$\sum_{j=0}^{\infty} f_j^2 = \mathbb{E}\left\{y(k)^2\right\}$$

which is independent on k by stationarity. If we set

$$f(z) = \sum_{k=0}^{\infty} f_k z^{-k} \in \bar{H}^2,$$

we see that a regular process is obtained by feeding white noise to an $l^2 \to l^\infty$ stable linear system [45].

In the special case where f is rational, y is called an Auto-Regressive Moving Average process, which is popular because it lends itself to efficient computations. When trying to fit such a model, say of order n to y, a typical interest is in minimizing the variance of the error between the true output and the prediction of the model. In this way one is led to solve for

$$\min_{g \in \mathcal{R}_{n,n}^{\mathbb{R}} \cap \bar{H}^2} \|f-g\|_2.$$

This principle can be used to identify a linear system from observed stochastic inputs, although computing the f_j is difficult because it requires spectral factorization of the function

$$\sum_{j \in \mathbb{Z}} \mathbb{E}\left\{y(j+k)y(k)\right\}$$

whose Fourier coefficients can only be estimated by ergodicity through time averages of the observed sample path of y. In practice, one would rather use time averages already in the optimization *criterion*, but this can be proved asymptotic to the previous problem [26].

To lend perspective to the discussion, let us briefly digress on the more general case where the input is an arbitrary stationary process. Applying the spectral theorem to the shift operator on the Hilbert space of the process allows one to compute the squared variance of the output error as a weighted L_2 integral:

$$\frac{1}{2\pi}\int_0^{2\pi} |f-g|^2 \, \mathrm{d}\mu, \tag{9}$$

where the positive measure μ is the so-called spectral measure of the input process (that reduces to Lebesgue measure when the latter is white noise), and f now has to belong to a weighted Hardy space $\bar{H}^2(\mathrm{d}\mu)$. Though we shall not dwell on this, we want to emphasize that the spectral theorem, as applied to shift operators, stresses deep links between time and frequency representations of a stochastic process, and the isometric character of this theorem (that may be viewed as a far-reaching generalization of Parseval's relation) is a fundamental reason why L^2 approximation problems arise in system theory. The scheme just mentioned is a special instance of *maximum likelihood identification* where the noise model is fixed [26, 33, 49], that aims at a rational extension of the Szegö theory [50] of orthogonal polynomials.

At this point, it must be mentioned that stochastic identification, as applied to linear dynamical systems, is not just concerned with putting up probabilistic interpretations to rational approximation *criteria*. Its main methodological contribution is to provide one with a method of choosing the *degree* of the approximant as the result of a trade-off between the *bias term* (i.e. the approximation error that goes small when the degree goes large) and the *variance term* (i.e. the dispersion of the estimates that goes large when the degree goes large and eventually makes the identification unreliable). We shall not touch on this deeper aspect of the stochastic paradigm, whose deterministic counterpart pertains to the numerical analysis of approximation theory (when should we stop increasing the degree to get a better fit since all we shall approximate further is the error caused by truncation, round off, etc.?). For an introduction to this circle of ideas, the interested reader is referred to the above-quoted textbooks.

3.2 Harmonic Identification

This example deals with continuous-time rather than discrete-time linear control systems, namely with convolution operators $u(t) \to y(t)$ of the form:

$$y(t) = \int_0^t h(t-\tau)u(\tau)\,\mathrm{d}\tau.$$

The function $h : [0, \infty) \to \mathbb{R}$ is called the *impulse response* of the system, as it formally corresponds to the output generated by a delta function. If h and u have exponential growth, so does y and the one-sided Laplace transforms $Y(s)$, $U(s)$ and $H(s)$ are defined on some common half-plane $\{\operatorname{Re} z > \sigma\}$. The system operates in this frequency domain as multiplication by the transfer-function H:

$$Y(s) = H(s)U(s).$$

This time, rational transfer-functions of degree n correspond to linear differential operators of order n forced by the input u.

The Hardy spaces involved are now those of the right half-plane, and their relation to stability is:

(i) the system is bounded $L^2[0,\infty) \to L^2[0,\infty)$ iff $H \in \mathcal{H}^\infty_\mathbb{R}$;
(ii) the system is bounded $L^2[0,\infty) \to L^\infty[0,\infty)$ iff $H \in \mathcal{H}^2_\mathbb{R}$;
(iii) the system is bounded $L^\infty[0,\infty) \to L^\infty[0,\infty)$ iff $H \in \mathcal{W}_\mathbb{R}$, the Wiener algebra of the right half-plane consisting of Laplace transforms of summable functions $[0,\infty) \to \mathbb{R}$.

One of the most effective methods to identify a system which sends bounded inputs to bounded outputs is to plug in a periodic input $u = \mathrm{e}^{\mathrm{i}\omega t}$ and to observe the asymptotic steady-state output which is:

$$y(t) = \lambda \mathrm{e}^{\mathrm{i}\phi}\mathrm{e}^{\mathrm{i}\omega t},$$

where λ and ϕ are respectively the modulus and the argument of $H(\mathrm{i}\omega)$.

In this way, one can estimate the transfer function on the imaginary axis and, for physical as well as computational purposes, one is often led to rationally approximate the experimental data thus obtained. In practice, the situation is considerably more complicated, because experiments cannot be performed on the whole axis and usually the system will no longer behave linearly at high frequencies. In fact, if Ω designates the bandwidth on the imaginary axis where experiments are performed, one can usually get a fairly precise estimate of $H_{|\Omega}$, the restriction of H to Ω, but all one has on $\mathrm{i}\mathbb{R} \setminus \Omega$ are qualitative features of the model induced by the physics of the system. Though it seems natural to seek

$$\min_{G \in \mathcal{R}^\mathbb{R}_{n,n} \cap \mathcal{H}^2} \|H - G\|_{L^2(\Omega)}, \qquad \text{or} \qquad \min_{G \in \mathcal{R}^\mathbb{R}_{n,n} \cap \mathcal{H}^\infty} \|H - G\|_{L^\infty(\Omega)},$$

such a problem is poorly behaved because the optimum may not exist (the best rational may have unstable poles) and even if it exists it may lead to a wild behavior off Ω. One way out, which is taken up in these notes, is to extrapolate a complete model in $\mathcal{H}^p$ from the knowledge of $H_{|\Omega}$ by solving an analytic bounded extremal problems as presented in the forthcoming section. Once the complete model is obtained, one faces a rational approximation problem in $\mathcal{H}^p$ that we will comment upon. The $\mathcal{H}^2$ norm is often better suited, due to the measurement errors, but physical constraints on the global model, like passivity for instance, typically involve the uniform norm. Figure 1 shows a numerical example of this two-steps identification scheme on a hyperfrequency filter (see [5, 44]), and an illustration in the design of transmission lines can be found in [47]. Often, weights are added in the criteria to trade-off between L^2 norms, that tend to oversmooth the data, and L^∞ norms that are put off by irregular samples. We shall not consider this here, and turn to approximation proper.

4 Some approximation problems

We discuss below some approximation problems connected to identification along the previous lines. This will provide us with an opportunity to intro-

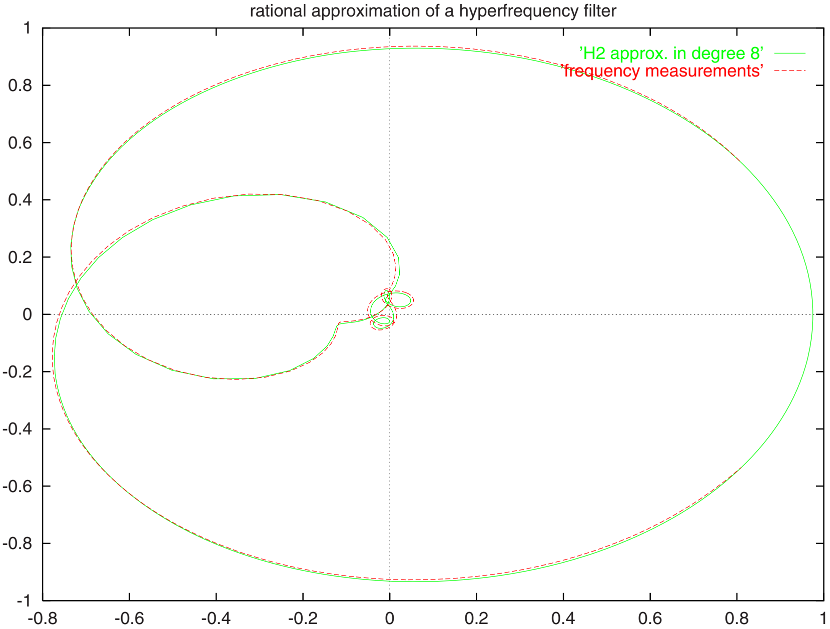


Fig. 1. The dotted line in this diagram is the Nyquist plot (i.e. the image of the bandwidth on the imaginary axis) of the transfer function of the reflexion of a hyperfrequency filter measured by the French CNES (Toulouse). The data were first completed by solving an $\bar{H}^2$ bounded extremal problem and then approximated by a rational function of degree 8 whose Nyquist plot has been superimposed on the figure. The locus is not conjugate-symmetric because a low-pass transformation sending the central frequency to the origin was performed on the data. This illustrates that approximation with complex Fourier coefficients can be useful in system identification, even though the physical system is real.

duce analytic operators of Hankel and Toeplitz type. We begin with analytic extrapolation from partial boundary data.

4.1 Analytic bounded extremal problems

For I is a measurable subset of $\mathbb{T}$ and J the complementary subset, if h_1 is a function defined on E and h_2 a function defined on J, we use the notation $h_1 \vee h_2$ for the concatenated function, defined on the whole of $\mathbb{T}$, which is h_1 on I and h_2 on J. The $L^2(I)/L^2(J)$ analytic bounded extremal problem is:

$[ABEP\big(L^2(I), L^2(J)\big)]$
Given $f \in L^2(I)$, $\psi \in L^2(J)$ and a strictly positive constant M, find $g_0 \in H^2$ such that

$$\|g_0(\mathrm{e}^{\mathrm{i}\theta}) - \psi(\mathrm{e}^{\mathrm{i}\theta})\|_{L^2(J)} \leq M \quad \text{and}$$

$$\|f - g_0\|_{L^2(I)} = \min_{\substack{g \in H^2 \\ \|g-\psi\|_{L^2(J)} \leq M}} \|f - g\|_{L^2(I)} \,. \tag{10}$$

We saw in section 3.2 the relevance of such a problem in identification, although on the line rather than the circle. The isometry (8) transforms the version there to the present one. Moreover, if I is symmetric with respect to the real axis and f has the conjugate symmetry (i.e. $f \vee 0$ has real Fourier coefficients), then g_0 also will have the conjugate symmetry because it is unique by strict convexity.

Problem $[ABEP\big(L^2(I), L^2(J)\big)]$ may be viewed as a Tikhonov-like regularization of a classical ill-posed issue mentioned in the introduction, namely how can one recover an analytic function in a disk from its values on a subset of the boundary circle. In the setting of Hardy spaces, this issue was initially approached using the so-called Carleman interpolation formulas [2]. Apparently the first occurrence of (one instance of) Problem $[ABEP\big(L^2(I), L^2(J)\big)]$ was investigated in [28], that proceeds in the time rather than the frequency domain. Our exposition below is different in that it emphasizes the connection with Harmonic Analysis and Analytic Operator theory. This approach lends itself to various generalizations and algorithmic analyses that we shall discuss briefly.

As in any constrained convex problem, one expects the solution of Problem $[ABEP\big(L^2(I), L^2(J)\big)]$ to depend linearly on the data via some unknown Lagrange parameter. This is best expressed upon introducing the Toeplitz operator :

$$\begin{aligned} \phi_{\chi_J} : H^2 &\to H^2 \\ g &\mapsto P_{H^2}(\chi_J g) \end{aligned} \tag{11}$$

with symbol χ_J, the characteristic function of J.

Theorem 1 ([1, 7]). *Assume that I has positive measure. Then, there is a unique solution g_0 to (10). Moreover, if f is not the restriction to I of a H^2 function whose $L^2(J)$-distance to ψ is less than or equal to M, this unique solution is given by*

$$g_0 = \big(1 + \lambda\phi_{\chi_J}\big)^{-1} P_{H^2}(f \vee (1+\lambda)\psi), \tag{12}$$

where $\lambda \in (-1, +\infty)$ is the unique real number such that the right hand side of (12) has $L^2(J)$-norm equal to M.

Note that (12) indeed makes sense because the spectrum of ϕ_{χ_J} is $[0, 1]$. Theorem 1 provides a constructive means of solving $ABEP\big(L^2(I), L^2(J)\big)$ because, although the correct value for λ is not known a priori, the $L^2(J)$-norm of the right-hand side in (12) is decreasing with λ so that iterating by dichotomy allows one to converge to the solution.

Let us point out that $H^2_{|_I}$ is dense in $L^2(I)$, hence the error in (10) can be made very small, but this is at the cost of making M very big unless

$f \in H^2_{|_I}$, a circumstance that essentially never happens due to modeling and measurement errors. In this connection, it is interesting to ask how fast M goes to $+\infty$ as the error $e = \|f - g_0\|_{L^2(I)}$ goes to 0. Using the constructive diagonalization of Toeplitz operators [41] with multiplicity 1, one can get fairly precise asymptotics when I is an interval. To state a typical result, put $I = (\mathrm{e}^{-\mathrm{i}a}, \mathrm{e}^{\mathrm{i}a})$ with $0 < a < \pi$, and let $\mathcal{W}^{1,1}(I)$ denote the Sobolev space of absolutely continuous functions on I.

Theorem 2 ([6]). *Let I be as above and f satisfies the following two assumptions :*

$$(1 - \mathrm{e}^{-\mathrm{i}\theta}\mathrm{e}^{\mathrm{i}a})^{-1/2}(1 - \mathrm{e}^{-\mathrm{i}\theta}\mathrm{e}^{-\mathrm{i}a})^{-1/2} f(\mathrm{e}^{\mathrm{i}\theta}) \in L^1(I)\,, \tag{13}$$

$$(1 - \mathrm{e}^{-\mathrm{i}\theta}\mathrm{e}^{\mathrm{i}a})^{1/2}(1 - \mathrm{e}^{-\mathrm{i}\theta}\mathrm{e}^{-\mathrm{i}a})^{1/2} f(\mathrm{e}^{\mathrm{i}\theta}) \in \mathcal{W}^{1,1}(I)\,. \tag{14}$$

If we set $e = \|f - g_0\|^2_{L^2(I)}$, where g_0 is the solution to (10), then to each $K_1 > 0$ there is $K_2 = K_2(f) > 0$ such that

$$M^2 \leq K_2\, e^2 \exp\{K_1 e^{-1}\}\,. \tag{15}$$

In the above statement, the factor e^{-1} in the exponent cannot be replaced by $h(e)$ for some function $h : \mathbb{R}^+ \to \mathbb{R}^+$ such that $h(x) = o(1/x)$ as $x \searrow 0$.

Adding finitely many degrees of smoothness would improve the above rate but only polynomially, and the meaning of Theorem 2 is that a good approximation is numerically not feasible if $f \notin C^\infty$, because M goes too large, *unless f is very close to the trace of a Hardy function.* It is striking to compare this with the analogous result when f is the trace on I of a meromorphic function:

Theorem 3. *If f is of the form h/q_N with $h \in H^2$ and q_N a polynomial of degree N whose poles lie at distance $d > 0$ from $\mathbb{T}$. Then*

$$M^2 = O\left(N^2 |\log e|\right)\,, \tag{16}$$

and the Landau symbol O holds uniformly with respect to $\|h\|_2$ and d, the estimate being sharp in the considered class of functions.

Comparing Theorems 2 and 3 suggests that the approximation is much easier if f extends holomorphically in a 2-D neighborhood of I, and that data in rational or meromorphic form should be favored, say as compared to splines. The $L^\infty(I)/L^\infty(J)$ analog can also be addressed:

$[ABEP\big(L^\infty(I), L^\infty(J)\big)]$
Given $f \in L^\infty(I)$, $\psi \in L^\infty(J)$, and a strictly positive constant M, find $g_0 \in H^\infty$ such that

$$\|g_0(\mathrm{e}^{\mathrm{i}\theta}) - \psi(\mathrm{e}^{\mathrm{i}\theta})\|_{L^\infty(J)} \leq M \quad \text{and}$$

$$\|f - g_0\|_{L^\infty(I)} = \min_{\substack{g\in H^\infty \\ \|g-\psi\|_{L^\infty(J)}\leq M}} \|f - g\|_{L^\infty(I)} \,. \tag{17}$$

A more general version is obtained by letting M be a function in $L^\infty(J)$ and the constraint become $|g - \psi| \leq M$ a.e. on J. If $\psi/M \in L^\infty(J)$, this version reduces to the present one because either $\log M \notin L^1(J)$ in which case the inequality $\log|g| \leq \log M + \log(1 + |\psi/M|)$ shows that $g = 0$ is the only candidate approximant, or else $\log M \in L^1(J)$ and we can form the outer function $w_M \in H^\infty$ having modulus 1 on I and M on J; then, upon replacing f by f/w_M and ψ/w_M and observing that g belongs to H^∞ and satisfies $|g| \leq M$ a.e. on J if, and only if, g/w_M lies in H^∞ and satisfies $g/w_M \leq 1$ a.e. on J (because g/w_M lies by construction in the Nevanlinna class whose intersection with $L^\infty(\mathbb{T})$ is H^∞), we are back to $M = 1$. If $\psi/M \notin L^\infty(J)$ the situation is more complicated.

This time we need introduce Hankel rather than Toeplitz operators. Very nice expositions can be found in [38, 40, 36], the first being very readable and the second very comprehensive. Given $\varphi \in L^\infty$, the *Hankel operator of symbol* φ is the operator

$$\Gamma_\varphi : H^2 \to \bar{H}^2_0$$

given by

$$\Gamma_\varphi g = P_{\bar{H}^2_0}(\varphi g)$$

where $P_{\bar{H}^2_0}$ denotes the orthogonal projection of $L^2(\mathbb{T})$ onto $\bar{H}^2_0$. A Hankel operator is clearly bounded, and it is compact whenever it admits a continuous symbol; note that the operator only characterizes the symbol up to the addition of some $\mathcal{H}^\infty$-function. Thus, whenever $\varphi \in H^\infty + C(\mathbb{T})$ (the latter is in fact an algebra), the operator Γ_φ is compact and therefore it has a maximizing vector $v_0 \in H^2$, namely a function of unit norm such that $\|\Gamma_\varphi(v_0)\|_2 = |||\Gamma_\varphi|||$, the norm of Γ_φ. Let us mention also that Hankel operators of finite rank are those admitting a rational symbol.

Theorem 4 ([8]). *Assume that I has positive measure and that ψ extends continuously to $\overline{J}$. Then, there is a solution g_0 to (17). Moreover, if f is not the restriction to I of a H^∞ function whose $L^\infty(J)$-distance to ψ is less than M, so that the value β of the problem is strictly positive, and if moreover $f \vee \psi \in H^\infty + C(\mathbb{T})$, this solution is unique and given by:*

$$g_0 = w^{-1}_{M/\beta}\,\frac{P_{H^2}\left((f\vee\psi)w_{M/\beta}v_0\right)}{v_0}\,, \tag{18}$$

where $w_{M/\beta}$ is the outer function with modulus M/β on I and modulus 1 *on J, and v_0 is a maximizing vector of the Hankel operator $\Gamma_{(f\vee\psi)w_{M/\beta}}$.*

Here again, the solution has conjugate symmetry if the data do. Although the value β of the problem is not known a priori, it is the unique positive real number such that the right hand side of (18) has modulus M a.e. on J, and so the theorem allows for us to constructively solve $[ABEP\big(L^\infty(I), L^\infty(J)\big)]$ if a maximizing vector of $\Gamma_{(f\vee\psi)w_{M/\beta}}$ can be computed for given β. In [9], generically convergent algorithms to this effect are detailed in the case where I is an interval and $f \vee \psi$ is C^1-smooth. They are based on the fact that a smooth H^∞ function may be added to $f\vee\psi$ to make it vanish at the endpoints of I, and in this case $(f \vee \psi)w_{M/\beta}$ may sufficiently well approximated by rationals, say in Hölder norm (uniform convergence is not sufficient here, see [40]). Reference [9] also contains a meromorphic extension.

Analytic bounded extremal problems have been generalized to abstract Hilbert and Banach space settings, with applications to hyperinvariant subspaces [34, 17, 48, 18]; they can be posed with different constraints where bounds are put on the imaginary part rather than the modulus [27]. In another connection, the work [3] investigates the problem of mixed type $[ABEP\big(L^2(I), L^\infty(J)\big)]$, which is important for instance when identifying passive systems whose transfer-function must remain less than 1 in modulus at every frequency. It turns out that the solution can be expressed very much along the same lines as $[ABEP\big(L^2(I), L^2(J)\big)]$, except that this time unbounded Toeplitz operators appear. We shall not go further into such generalizations, and we rather turn to rational approximation part of the two-step identification procedure sketched in section 3.

4.2 Meromorphic and rational approximation

We saw in subsection 3.1, and in the second step of the identification scheme sketched in subsection 3.2, that stable and proper rational approximation of a complete model on the line or the circle is an important problem from the system-theoretic viewpoint. Here again, the isometry (8) makes it enough to consider the case of the circle. We shall start with the Adamjan-Arov-Krein theory (in short: AAK) which deals with a related issue, namely *meromorphic approximation* in the uniform norm.

For $k = 0, 1, 2, \ldots$, recall that the *singular values* of Γ_φ are defined by the formula:

$$s_k(\Gamma_\varphi) := \inf\left\{|||\Gamma_\varphi - A|||, \quad A \text{ an operator of rank } \leq k \text{ on } H^2\right\}.$$

When $\varphi \in H^\infty + C(\mathbb{T})$, the singular values are, by compactness, the square roots of the eigenvalues of $\Gamma_\varphi^*\Gamma_\varphi$ arranged in non-increasing order; a k-th *singular vector* is an eigenvector of unit norm associated to $s_k(\Gamma_\varphi)$.

A celebrated connection between the spectral theory of Hankel operators and best meromorphic approximation on the unit circle is given by AAK theory [38, 40] as follows. Recall from the introduction the notation H_n^∞ for meromorphic functions with at most n poles in L^∞. The main result asserts that:

$$\inf_{g\in H_n^\infty} \|\varphi - g\|_\infty = s_n(\Gamma_\varphi) \tag{19}$$

where the *infimum* is attained; moreover, the *unique* minimizer is given by the formula

$$g_n = \varphi - \frac{\Gamma_\varphi v_n}{v_n} = \frac{P_{H^2}(\varphi v_n)}{v_n}, \tag{20}$$

where v_n is *any* n-th singular vector of Γ_φ. Formula (20) entails in particular that the inner factor of v_n is a Blaschke product of degree at most n. The error function $\varphi - g_n$ has further remarkable properties; for instance it has constant modulus $s_n(\Gamma_\varphi)$ a.e. on $\mathbb{T}$.

From the point of view of constructive approximation, it is remarkable that the infimum in (19) can be computed, and the problem as to whether one can pass from the optimal meromorphic approximant in (20) to a nearly optimal *rational* approximant has attracted much attention. Most notably, it is shown in [23] that $P_{\bar{H}_0^2}(g_n)$, which is rational in $\mathcal{R}_{n-1,n}$, produces an L^∞ error within

$$2\sum_{j=n+1}^{\infty} s_j(\Gamma_\varphi) \tag{21}$$

of the optimal one out of $\mathcal{R}_{n-1,n}$. To estimate how good this bound requires a link between the decay of the singular values of Γ_φ and the smoothness of φ. The summability of the singular values is equivalent to the belonging of $P_{\bar{H}_0^2}(\varphi)$ to the Besov class B_1^1 of the disk [40], but this does not tell how fast the series converges.

For an appraisal of this, we need introduce some basic notions of potential theory. For more on fundamental notions like equilibrium measure, potential, capacity, balayage, as well as the basic theorems concerning them, the reader may want to consult some recent textbook such as [43]. However, for his convenience, we review below the main concepts, starting with logarithmic potentials.

Let $E \subset \mathbb{C}$ be a compact set. To support his intuition, one may view E as a plane conductor and imagine he puts a unit electric charge on it. Then, if a distribution of charge is described as being a Borel measure μ on E, the electrostatic equilibrium has to minimize the internal energy:

$$I(\mu) = \iint \log\frac{1}{|x-t|} \mathrm{d}\mu(x)\mathrm{d}\mu(t)$$

among all probability measures supported on E. This is because on the plane the Coulomb force is proportional to the inverse of the distance between the particles, and therefore the potential is its logarithm. There are sets E (called polar sets or sets of zero logarithmic capacity) which are so thin that the energy $I(\mu)$ is infinite, no matter what the probability measure μ is on E; for those we do not define the equilibrium measure. But if E is such that $I(\mu)$ is finite for *some* probability measure μ on E, then there is a unique minimizer for $I(\mu)$ among all such probability measures. This minimizer is called the *equilibrium measure* (with respect to logarithmic potential) of E, and we denote it by ω_E. For example, the equilibrium measure of a disk or circle is the normalized arc measure on the circumference, while the equilibrium measure of a segment $[a,b]$ is

$$\mathrm{d}\omega_{[a,b]}(x) = \frac{1}{\pi\sqrt{(x-a)(b-x)}}\mathrm{d}x.$$

That the equilibrium measure of a disk is supported on its circumference is no accident: the equilibrium measure of E is always supported on the outer boundary of E.

Associated to a measure μ on a set E is its *logarithmic potential*:

$$U^{\mu}(x) = \int \log\frac{1}{|x-t|}\,\mathrm{d}\mu(t),$$

which is superharmonic on $\mathbb{C}$ with values in $(-\infty,+\infty]$.

From the physical viewpoint, this is simply the electrostatic potential corresponding to the distribution of charge μ. Perhaps the nicest characterization of ω_E among all probability measures on E is that $U^{\omega_E}(x)$ is equal to some constant M on E, except possibly on a polar subset of E where it may be less than M. Of necessity then, we have that $M = I(\omega_E)$ because measures of finite energy like ω_E do not charge polar sets. Points at which U^{ω_E} is discontinuous are called *irregular* (of necessity they lie on the outer boundary of E), and E itself is said to be *regular* if it has no irregular points. All nice compact sets are regular, in particular all whose boundary has no connected component that reduces to a point. The regularity of E is equivalent to the *regularity of the Dirichlet problem* in the unbounded component $\mathcal{V}$ of $\overline{\mathbb{C}}\setminus E$, meaning that for any continuous function f on the outer boundary $\partial_e E$ of E, there is a harmonic function in $\mathcal{V}$ which is continuous on $\overline{\mathcal{V}} = \mathcal{V}\cup\partial_e E$ and coincides with f on $\partial_e E$.

The number $\exp\{-I_{\omega_E}\}$ is called the *logarithmic capacity* of E; conventionally, polar sets have capacity zero. A property that holds in the complement of a polar set is said to hold *quasi-everywhere*.

We now turn to Green potentials: when $E \subset \mathcal{U}$ where $\mathcal{U}$ is a domain (i.e. a connected open set) whose boundary is non-polar, one can introduce similar concepts upon replacing the logarithmic kernel $\log 1/|x-t|$ by the *Green function* of $\mathcal{U}$ with pole at t. This gives rise to the notions of *Green equilibrium measure*, *Green potential*, and *Green capacity*. We use them only when $\mathcal{U}$ is the unit disk, in which case the *Green function with pole at a* is

$$g(z,a) = \log\left|\frac{1-\overline{a}z}{z-a}\right|.$$

To each probability measure σ with support in E, we associate the Green energy:

$$I_G(\sigma) = \iint g(z,a)\mathrm{d}\sigma(z)\mathrm{d}\sigma(a).$$

If E is non polar, then among all probability measures supported on E there is one and only one measure $\varOmega_E$ minimizing the Green energy, which is called the *Green equilibrium measure* of E associated with the unit disk. It is the only probability measure on E whose *Green potential*

$$G^{\varOmega_E}(z) = \int g(z,a)\,\mathrm{d}\varOmega_E(a)$$

is equal to a constant M quasi-everywhere on E and less or equal to that constant everywhere (see e.g. [43]). Of necessity $M = I_G(\varOmega_E)$, and the number $1/M$ is called the *Green capacity of E* (note the discrepancy with the logarithmic case: we do not take exponentials here). It is also referred to as the *capacity of the condenser* $(\mathbb{T}, E)$.

To explain this last piece of terminology, let us mention that from the point of view of electrostatics, the Green equilibrium distribution of E is the distribution of charges at the equilibrium if one puts a positive unit charge on E (the first plate of the condenser) and a negative unit charge on $\mathbb{T}$ (the second plate of the condenser).

With these definitions in mind, we are ready to go deeper analyzing our rational approximation problem. When φ is analytic outside some compact $K \subset \mathbb{D}$, it is shown in [52] that, if e_n is the optimal error in uniform approximation to φ from $\mathcal{R}_{n,n}$ on $\overline{\mathbb{C}} \setminus \mathbb{D}$, then

$$\limsup e_n^{1/n} \leq \mathrm{e}^{-1/(C)} \tag{22}$$

where C is the capacity of the condenser $(K, \mathbb{T})$.

Equation (22) shows that the decay of the singular values is geometric when φ is analytic outside $\mathbb{D}$ and across $\mathbb{T}$, and allows for an appraisal of (21) in this case although this appraisal is pessimistic in that, as was proved in [37], one actually has:

$$\liminf e_m^{1/m} \leq \mathrm{e}^{-1/(2C)}.$$

For functions defined by Cauchy integrals over so-called symmetric arcs, this $\liminf$ is a true limit [24]. Moreover, in the particular case of functions analytic except for two branchpoints in the disk, the probability measure having equal mass at each pole of g_n converges weak-* (in the dual of continuous functions with compact support on $\mathbb{C}$) to the equilibrium distribution on the cut K that minimizes the capacity of $(K, \mathbb{T})$ among all cuts joining the branch points [12, 31], a hyperbolic analog to Stahl's theorem on Padé approximants [46]. Generalizing these results would bring us into current research.

Let us conclude with a few words concerning $\bar{H}^2$ rational approximation of type (n, n). We saw in subsection 3.1 and 3.2 the relevance of this problem in identification, but still it is basically unsolved. For a comparison, observe from the Courant minimax principle that the error in AAK approximation to φ is

$$s_n(\Gamma_\varphi) = \inf_{V \in \mathcal{V}_n} \sup_{\substack{v \in V \\ \|v\|_2 = 1}} \|\Gamma_\varphi(v), \|$$

where $\mathcal{V}_n$ denotes the collection of subspaces of codimension at least n, by whereas it is easy to show that the error in $\bar{H}^2$ rational approximation is

$$E_n(\varphi) = \min_{b \in \mathcal{B}_n} \|\Gamma_\varphi(b)\|$$

where $\mathcal{B}_n$ denotes the set of Blaschke products of degree at most n. The nonlinear character of $\mathcal{B}_n$ makes for a much more difficult problem, and practical algorithms have to rely on numerical searches with the usual burden of local minima. Space prevents us from describing in details what is known on this problem, and we refer the reader to [4] for a survey. Let us simply mention that in the special case of Markov functions, namely Cauchy transforms of positive measures on a segment that also correspond to the transfer functions of so-called relaxation systems [53, 16], a lot is known including sharp error rates [10, 13] asymptotic uniqueness of a critical point for Szegö-smooth measures [14] and uniqueness for all orders and small support [15]. For certain entire functions like the exponential, sharp error rates and asymptotic uniqueness of a critical point have also been derived [11], but for most classes of functions the situation is unclear. Results obtained so far concern functions for which the decay of the error is comparable to the one in AAK approximation and fairly regular. Finally we point out that, despite the lack of a general theory, rather efficient algorithms are available to generate local minima e.g. [21, 25, 35].

References

1. D. Alpay, L. Baratchart, J. Leblond. Some extremal problems linked with identification from partial frequency data. In J.L. Lions, R.F. Curtain, A. Bensoussan, editors, *10th conference on analysis and optimization of systems, Sophia–Antipolis 1992*, volume 185 of *Lect. Notes in Control and Information Sc.*, pages 563–573. Springer-Verlag, 1993.
2. L. Aizenberg. *Carleman's formulas in complex analysis.* Kluwer Academic Publishers, 1993.
3. L. Baratchart, J. Leblond, F. Seyfert. *Constrained analytic approximation of mixed H^2/H^∞ type on subsets of the circle.* In preparation.
4. L. Baratchart. Rational and meromorphic approximation in L^p of the circle : system-theoretic motivations, critical points and error rates. In *Computational Methods and Function Theory*, pages 45–78. World Scientific Publish. Co, 1999. N. Papamichael, St. Ruscheweyh and E.B. Saff *eds.*

5. L. Baratchart, J. Grimm, J. Leblond, M. Olivi, F. Seyfert, F. Wielonsky. Identification d'un filtre hyperfréquence. Rapport Technique INRIA No 219., 1998.
6. L. Baratchart, J. Grimm, J. Leblond, J.R. Partington. Asymptotic estimates for interpolation and constrained approximation in H^2 by diagonalization of toeplitz operators. *Integral equations and operator theory*, 45:269–299, 2003.
7. L. Baratchart, J. Leblond. Hardy approximation to L^p functions on subsets of the circle with $1 \leq p < \infty$. *Constructive Approximation*, 14:41–56, 1998.
8. L. Baratchart, J. Leblond, J.R. Partington. Hardy approximation to L^∞ functions on subsets of the circle. *Constructive Approximation*, 12:423–436, 1996.
9. L. Baratchart, J. Leblond, J.R. Partington. Problems of Adamjan–Arov–Krein type on subsets of the circle and minimal norm extensions. *Constructive Approximation*, 16:333–357, 2000.
10. L. Baratchart, V. Prokhorov, E.B. Saff. Best L^P meromorphic approximation of Markov functions on the unit circle. *Foundations of Constructive Math*, 1(4):385–416, 2001.
11. L. Baratchart, E.B. Saff, F. Wielonsky. A criterion for uniqueness of a critical points in H^2 rational approximation. *J. Analyse Mathématique*, 70:225–266, 1996.
12. L. Baratchart, F. Seyfert. An L^p analog to the AAK theory. *Journal of Functional Analysis*, 191:52–122, 2002.
13. L. Baratchart, H. Stahl, F. Wielonsky. Asymptotic error estimates for L^2 best rational approximants to Markov functions on the unit circle. *Journal of Approximation Theory*, (108):53–96, 2001.
14. L. Baratchart, H. Stahl, F. Wielonsky. Asymptotic uniqueness of best rational approximants of given degree to Markov functions in L^2 of the circle. 2001.
15. L. Baratchart, F. Wielonsky. Rational approximation in the real Hardy space H^2 and Stieltjes integrals: a uniqueness theorem. *Constructive Approximation*, 9:1–21, 1993.
16. R.W. Brockett, P.A. Fuhrmann. Normal symmetric dynamical systems. *SIAM J. Control and Optimization*, 14(1):107–119, 1976.
17. I. Chalendar, J.R. Partington. Constrained approximation and invariant subspaces. J. Math. Anal. Appl. 280 (2003), no. 1, 176–187.
18. I. Chalendar, J.R. Partington, M.P. Smith. Approximation in reflexive Banach spaces and applications to the invariant subspace problem. Proc. Amer. Math. Soc. 132 (2004), no. 4, 1133–1142.
19. P.L. Duren. *Theory of H^p-spaces.* Academic Press, 1970.
20. B. Francis. *A course in H^∞ control theory.* Lecture notes in control and information sciences. Springer–Verlag, 1987.
21. P. Fulcheri, M. Olivi. Matrix rational H^2–approximation: a gradient algorithm based on Schur analysis. 36(6):2103–2127, 1998. SIAM Journal on Control and Optimization.
22. J.B. Garnett. *Bounded Analytic Functions.* Academic Press, 1981.
23. K. Glover. All optimal Hankel–norm approximations of linear multivariable systems and their L^∞–error bounds. *Int. J. Control*, 39(6):1115–1193, 1984.
24. A.A. Gonchar, E.A. Rakhmanov. Equilibrium distributions and the degree of rational approximation of analytic functions. *Math. USSR Sbornik*, 176:306–352, 1989.

25. J. Grimm. Rational approximation of transfer functions in the hyperion software. Rapport de recherche 4002, INRIA, September 2000.
26. E.J. Hannan, M. Deistler. *The statistical theory of linear systems.* Wiley, New York, 1988.
27. B. Jacob, J. Leblond, J.-P. Marmorat, J.R. Partington. A constrained approximation problem arising in parameter identification. Linear Algebra and its Applications, 351-352:487-500, 2002.
28. M.G. Krein, P. Ya Nudel'man. *Approximation of $L^2(\omega_1, \omega_2)$ functions by minimum–energy transfer functions of linear systems.* Problemy Peredachi Informatsii, 11(2):37–60, English transl., 1975.
29. R.E. Kalman, P.L. Falb, M.A. Arbib. *Topics in mathematical system theory.* Mc Graw Hill, 1969.
30. P. Koosis. *Introduction to H_p-spaces.* Cambridge University Press, 1980.
31. R. Küstner. Distribution asymptotique des zéros de polynômes orthogonaux par rapport à des mesures complexes ayant un argument à variation bornée. Ph.D. thesis, University of Nice, 2003.
32. M.M. Lavrentiev. *Some Improperly Posed Problems of Mathematical Physics.* Springer, 1967.
33. L. Ljung. *System identification: Theory for the user.* Prentice–Hall, 1987.
34. J. Leblond, J.R. Partington. Constrained approximation and interpolation in Hilbert function spaces. *J. Math. Anal. Appl.*, 234(2):500–513, 1999.
35. J.P. Marmorat, M. Olivi, B. Hanzon, R.L.M. Peeters. Matrix rational H^2 approximation: a state-space approach using schur parameters. In *Proceedings of the C.D.C.*, 2002.
36. N.K. Nikolskii. *Treatise on the shift operator.* Grundlehren der Math. Wissenschaften 273. Springer, 1986.
37. O.G. Parfenov. Estimates of the singular numbers of a Carleson operator. *Math USSR Sbornik*, 59(2):497–514, 1988.
38. J.R. Partington. *An Introduction to Hankel Operators.* Cambridge University Press, 1988.
39. J.R. Partington. Robust identification and interpolation in H_∞. *Int. J. of Control*, 54:1281–1290, 1991.
40. V.V. Peller. *Hankel Operators and their Applications.* Springer, 2003.
41. M. Rosenblum, J. Rovnyak. *Hardy classes and operator theory.* Oxford University Press, 1985.
42. H.H. Rosenbrock. *State Space and Multivariable Theory.* Wiley, New York, 1970.
43. E.B. Saff, V. Totik. *Logarithmic Potentials with External Fields*, volume 316 of *Grundlehren der Math. Wiss.* Springer-Verlag, 1997.
44. F. Seyfert, J.P. Marmorat, L. Baratchart, S. Bila, J. Sombrin. Extraction of coupling parameters for microwave filters: Determination of a stable rational model from scattering data. *Proceedings of the International Microwave Symposium, Philadelphia*, 2003.
45. A.N. Shiryaev. *Probability.* Springer, 1984.
46. H. Stahl. The convergence of Padé approximants to functions with branch points. *J. of Approximation Theory*, 91:139–204, 1997.
47. J. Skaar. *A numerical algorithm for extrapolation of transfer functions.* Signal Processing, 83:1213–1221, 2003.
48. M.P. Smith. *Constrained approximation in Banach spaces.* Constructive Approximation, 19(3):465-476, 2003.

49. T. Söderström, P. Stoica. *System Identification.* Prentice–Hall, 1987.
50. G. Szegö. *Orthogonal Polynomials.* Colloquium Publications. AMS, 1939.
51. A. Tikhonov, N. Arsenine. *Méthodes de résolution des problèmes mal posés.* MIR, 1976.
52. J.L. Walsh. *Interpolation and approximation by rational functions in the complex domain.* A.M.S. Publications, 1962.
53. J.C. Willems. Dissipative dynamical systems, Part I: general theory, Part II: linear systems with quadratic supply rates. *Arch. Rat. Mech. and Anal.*, 45:321–351, 352–392, 1972.

Perturbative Series Expansions: Theoretical Aspects and Numerical Investigations

Luca Biasco and Alessandra Celletti

Dipartimento di Matematica,
Università di Roma Tre,
Largo S. L. Murialdo 1,
I-00146 Roma (Italy)
biasco@mat.uniroma3.it

Dipartimento di Matematica,
Università di Roma Tor Vergata,
Via della Ricerca Scientifica 1,
I-00133 Roma (Italy)
celletti@mat.uniroma2.it

Abstract

Perturbation theory is introduced by means of models borrowed from Celestial Mechanics, namely the two–body and three–body problems. Such models allow one to introduce in a simple way the concepts of integrable and nearly–integrable systems, which can be conveniently investigated using Hamiltonian formalism. After discussing the problem of the convergence of perturbative series expansions, we introduce the basic notions of KAM theory, which allows (under quite general assumptions) to state the persistence of invariant tori. The value at which such surfaces break–down can be determined by means of numerical algorithms. Among the others, we review three methods to which we refer as Greene, Padé and Lyapunov. We present some concrete applications to discrete models of the three different techniques, in order to provide complementary information about the break–down of invariant tori.

1 Introduction

The dynamics of the planets and satellites is ruled by Newton's law, according to which the gravitational force is proportional to the product of the masses of the interacting bodies and it is inversely proportional to the square of their distance. The description of the trajectories spanned by the celestial bodies starts with the simplest model in which one considers only the attraction exerted by the Sun, neglecting all contributions due to other planets or satellites. Such model is known as the *two–body problem* and it is fully described by Kepler's laws, according to which the motion is represented by a conic. Consider, for example, the trajectory of an asteroid moving on an elliptic orbit around the Sun. In the two–body approximation the semimajor axis and the eccentricity of the ellipse are fixed in time. However, such example represents

J.-D. Fournier et al. (Eds.): Harm. Analysis and Ratio. Approx., LNCIS 327, pp. 233–261, 2006.

only the first approximation of the asteroid's motion, which is actually subject also to the attraction of the other planets and satellites of the solar system. The most important contribution comes from the gravitational influence of Jupiter, which is the largest planet of the solar system, its mass being equal to 10^{-3} times the mass of the Sun. Therefore, next step is to consider the *three–body problem* formed by the Sun, the asteroid and Jupiter. A complete mathematical solution of such problem was hunted for since the last three centuries. A conclusive answer was given by H. Poincaré [18], who proved that the three–body problem does not admit a mathematical solution, in the sense that it is not possible to find explicit formulae which describe the motion of the asteroid under the simultaneous attraction of the Sun and Jupiter. For this reason, the three–body problem belongs to the class of non–integrable systems. However, since the mass of Jupiter is much smaller than the mass of the Sun, the trajectory of the asteroid will be in general weakly perturbed with respect to the Keplerian solution. In this sense, one can speak of the three–body problem as a nearly–integrable system (see section 3).

Let us consider a trajectory of the three–body problem with preassigned initial conditions. Though an explicit solution of such motion cannot be found, one can look for an approximate solution of the equations of motion by means of mathematical techniques [2] known as *perturbation theory* (see section 4), which can be conveniently introduced in terms of the Hamiltonian formalism reviewed in section 2. More precisely, one can construct a transformation of coordinates such that the system expressed in the new variables provides a better approximation of the true solution. The coordinate's transformations are builded up by constructing suitable series expansions, usually referred to[1] as Poincaré–Lindstedt series (see [2]), and a basic question (obviously) concerns their domain of convergence. In the context of perturbation theory, a definite breakthrough is provided by Kolmogorov's theorem, later extended by Arnold and Moser, henceforth known as KAM theory by the acronym of the authors [15], [1], [17]. Let ε denote the perturbing parameter, such that for $\varepsilon = 0$ one recovers the integrable case (in the three–body problem the perturbing parameter is readily seen to represent the Jupiter–Sun mass–ratio). The novelty of KAM theory relies in fixing the frequency, rather than the initial conditions, and in using a quadratically convergent procedure of solution (rather than a linear one, like in classical perturbation theory). The basic assumption consists in assuming a strongly irrational (or diophantine) condition on the frequency ω. In conclusion, having fixed a diophantine frequency ω for the unperturbed system ($\varepsilon = 0$), KAM theory provides an explicit algorithm to prove the persistence, for a sufficiently small $\varepsilon \neq 0$, of an invariant torus on which a quasi–periodic motion with frequency ω takes place; moreover, Kolmogorov's theory proves that the set of such invariant tori has positive measure in the phase space.

[1] Quoting [2] the Lindstedt technique "is one of the earliest methods for eliminating fast phases. We owe its contemporary form to Poincaré".

KAM theorem provides a lower bound on the break–down threshold of the invariant torus with frequency ω, say $\varepsilon = \varepsilon_c(\omega)$. Nowadays there are several numerical techniques which allow to evaluate, with accurate precision, the transition value $\varepsilon_c(\omega)$. One of the most widely accepted methods was developed by Greene [13] and it is based on the conjecture that the break–down of the invariant torus with rotation number ω is related to the transition from stability to instability of the periodic orbits with frequency equal to the rational approximants to ω. We remark that such conjecture was partially justified in [10], [16].

Being the invariant torus described in terms of the Poincaré–Lindstedt series expansions, it is definitely important to analyze the domain of convergence of such series in the complex parameter plane; we denote by $\varrho_c(\omega)$ the corresponding radius of convergence. Numerical experiments suggest that Greene's threshold coincides with the intersection of such domain with the positive real axis. Whenever the domain of convergence deviates from a circle, the two thresholds $\varepsilon_c(\omega)$ and $\varrho_c(\omega)$ may be markedly different. The domain of analyticity can be obtained by implementing Padé's method [3], which allows to approximate the perturbative series by the ratio of two polynomials. The zeros of the denominators provide the poles, which contribute to the determination of the analyticity domain. In order to evaluate the radius of convergence one can apply an alternative technique developed in [5], to which we refer as Lyapunov's method, based on the computation of a quantity related to the numerical definition of Lyapunov's exponents.

Concrete applications of Greene's, Padé's and Lyapunov's techniques are presented in sections 5 and 6 for a discrete model, known as the *standard map*. We consider also variants of such mapping (adding suitable Fourier components) and we analyze the behavior of invariant curves with three different frequencies. Notice that the results (Tables and Figures) provided in sections 5 and 6 are taken from [5].

2 Hamiltonian formalism

Consider a smooth function $H := H(y, x)$ with $(y, x) \in M := \mathbb{R}^n \times \mathbb{R}^n$, $n = 1, 2, 3, \ldots$ and the following systems of O.D.E.'s:

$$\begin{aligned} \dot{y}(t) &= -H_x(y(t), x(t)), \\ \dot{x}(t) &= H_y(y(t), x(t)) \end{aligned} \tag{1}$$

(here and in the following $H_x(y, x) \equiv \frac{\partial H}{\partial x}(y, x)$, $H_y(y, x) \equiv \frac{\partial H}{\partial y}(y, x)$). Define

$$\Phi_H^t(y_0, x_0) := (y(t), x(t))$$

as the solution at time t with initial data $y(0) := y_0$ and $x(0) := x_0$. We remark that the value of H along the solution of (1) is constant, i.e.

$$H(y(t),x(t)) \equiv H(y_0,x_0) =: E \tag{2}$$

for a suitable $E \in \mathbb{R}$. The function H is called *Hamiltonian*, M is the *phase space*, (1) are *Hamilton's equations* associated to H, Φ^t_H is the *Hamiltonian flow* of H at time t starting from (y_0,x_0) and (2) expresses the preservation of the energy.

An elementary example of the derivation of the Hamiltonian function associated to a mechanical model is provided by a pendulum. Consider a point of mass m, which keeps constant distance d from the origin of a reference frame. Suppose that the system is embedded in a weak gravity field of strength ε. Normalizing the units of measure so that $m = 1$ and $d = 1$, the Hamiltonian describing the motion is given by

$$H(r,\theta) = \frac{1}{2}r^2 + \varepsilon\cos\theta, \tag{3}$$

where $\theta \in \mathbb{T} \equiv \mathbb{R}/(2\pi\mathbb{Z})$ is the angle described by the particle with the vertical axis and $r \in \mathbb{R}$ is the velocity, i.e. $r = \dot{\theta}$. Being one–dimensional, the system can be easily integrated by quadratures.

Another classical example of a physical model which can be conveniently studied through Hamiltonian formalism, is provided by the harmonic oscillator. The equation governing the small oscillations (described by the coordinate $x \in \mathbb{R}$) of a body attached to the end of an elastic spring is given by Hooke's law, which can be expressed as

$$\ddot{x} = -\nu^2 x,$$

for a suitable $\nu > 0$ representing the strength of the spring. The corresponding Hamiltonian is given by

$$H(y,x) = \frac{1}{2}y^2 + \frac{1}{2}\nu^2 x^2,$$

whose associated Hamilton's equations are

$$\dot{y} = -\nu^2 x, \qquad \dot{x} = y. \tag{4}$$

Equations (4) can be trivially solved as

$$\begin{aligned} y(t) &= -\alpha\nu\sin(\nu t + \beta), \\ x(t) &= \alpha\cos(\nu t + \beta), \end{aligned} \tag{5}$$

for suitable constants α, β related to the initial conditions $x(0)$, $y(0)$. It is rather instructive to use this example in order to introduce suitable coordinates, known as *action–angle variables*, which will play a key role in the context of perturbation theory. Indeed, we proceed to construct a change of coordinates $\Psi(I,\varphi) = (y,x)$, defined through the relation

$$\Phi^t_{H\circ\Psi} = \Phi^t_H \circ \Psi. \tag{6}$$

For $(y, x) \neq (0, 0)$, we can define the change of variables

$$\begin{aligned} y &= \sqrt{2\nu I}\cos\varphi, \\ x &= \sqrt{2I/\nu}\sin\varphi, \end{aligned} \tag{7}$$

where $I > 0$ and $\varphi \in \mathbb{T} := \mathbb{R}/(2\pi\mathbb{Z})$. Notice that the coordinate I has the dimension of an action[2], while φ is an angle. A trivial computation shows that the previous change of coordinates satisfies (6). Moreover, we remark that the new Hamiltonian

$$K := K(I) := H \circ \Psi(I, \varphi) = \nu I$$

does not depend on φ. Denoting by $\omega := \nabla K(I) \equiv \frac{\partial K(I)}{\partial I}$, equations (1) become

$$\dot{I} = 0, \dot{\varphi} = \nu, \tag{8}$$

whose solution is given by

$$\begin{aligned} I(t) &\equiv I_0 \\ \varphi(t) &\equiv \nu t + \varphi_0. \end{aligned} \tag{9}$$

Notice that inserting (9) in (7), one recovers the solution (5). In view of this example, we are led to the following

Definition: A change of coordinates Ψ verifying (6) is said to be *canonical.* The coordinates $(I, \varphi) \in M$ (where the phase space is $M := \mathbb{R}^n \times \mathbb{T}^n$ with $\mathbb{T}^n := \mathbb{R}^n/(2\pi\mathbb{Z})^n$) are called *action–angle variables.*

Notice that if the Hamiltonian H is expressed in terms of action–angle variables, namely $H = H(I, \varphi)$, then equations (1) become

$$\dot{I}(t) = -H_\varphi(I(t), \varphi(t)), \qquad \dot{\varphi}(t) = H_I(I(t), \varphi(t)). \tag{10}$$

3 Integrable and nearly–integrable systems

Let us consider a Hamiltonian function expressed in action–angle variables, namely $H = H(I, \varphi)$, where $(I, \varphi) \in M := \mathbb{R}^n \times \mathbb{T}^n$. A system described by a Hamiltonian H, which does not depend on the angles, namely

$$H(I, \varphi) = h(I)$$

[2] Namely energy×time.

for a suitable function h, is said to be (completely) integrable. For these systems, Hamilton's equations can be written as $\dot{I} = 0$, $\dot{\varphi} = \nabla h(I)$. Correspondingly, we introduce the *invariant tori*[3]

$$\mathbb{T}_{\omega_0} := \{(I, \varphi) \,|\, I \equiv I_0, \quad \varphi \in \mathbb{T}^n\},$$

run by the linear flow $\varphi(t) = \omega_0 t + \varphi_0$, with $\omega_0 := \nabla h(I_0) \in \mathbb{R}^n$. If the frequency ω_0 is rationally independent (i.e., $\omega_0 \cdot k \neq 0$ for all $k \in \mathbb{Z}^n \setminus \{0\}$), the torus $\mathbb{T}_{\omega_0}$ is called non–resonant and it is densely filled by the flow $t \to \omega_0 t + \varphi_0$. In this case, the flow is said to be quasi–periodic. If the frequency ω_0 is rationally dependent, the torus is called resonant (and it is foliated by lower dimensional invariant tori). We remark that this case is highly degenerate, since the probability to have a rationally dependent frequency is zero.

In the previous statements we have fully classified all motions associated to integrable Hamiltonian systems. A more difficult task concerns the study of systems which are close to integrable ones. More precisely, consider a weak perturbation of an integrable Hamiltonian $h(I)$: denoting by $\varepsilon \in \mathbb{R}$ the size of the perturbation, we can write a perturbed system as

$$H(I, \varphi; \varepsilon) := h(I) + \varepsilon f(I, \varphi; \varepsilon), \tag{11}$$

where $(I, \varphi) \in M := \mathbb{R}^n \times \mathbb{T}^n$, f is a smooth function and the real parameter ε is small, i.e. $0 < \varepsilon < 1$. The equations of motion (10) become

$$\begin{aligned} \dot{I}(t) &= -\varepsilon \frac{\partial f}{\partial \varphi}(I(t), \varphi(t); \varepsilon), \\ \dot{\varphi}(t) &= \nabla h(I(t)) + \varepsilon \frac{\partial f}{\partial I}(I(t), \varphi(t); \varepsilon). \end{aligned} \tag{12}$$

Mechanical systems described by Hamiltonian functions of the form (11) are called *nearly–integrable*, since for $\varepsilon = 0$ equations (12) can be trivially integrated (compare with (8), (9)).

In order to provide explicit examples of integrable and nearly–integrable Hamiltonian systems, we recall in the following subsections the celebrated two–body and three–body problems.

3.1 The two–body problem

Consider the motion of an asteroid A orbiting around the Sun S, which is assumed to coincide with the origin of a fixed reference frame. Suppose to neglect the gravitational interaction of the asteroid with the other objects of the solar system. The two–body asteroid–Sun motion is described by Kepler's laws, which ensure that for negative energy the orbit of the asteroid around the

[3] Namely $\Phi^t_H(\mathbb{T}_{\omega_0}) \subseteq \mathbb{T}_{\omega_0}$.

Sun is an ellipse with the Sun located at one of the two foci. The Hamiltonian formulation of the two–body problem is described as follows. Choose the units of length, mass and time so that the gravitational constant and the mass of the Sun are normalized to one. In order to investigate the asteroid–Sun problem, it is convenient to introduce suitable coordinates, known as Delaunay variables:

$$(I_1, I_2) \equiv (L, G) \in \mathbb{R}^2, \qquad (\varphi_1, \varphi_2) \equiv (l, g) \in \mathbb{T}^2,$$

whose definition is the following. Denoting by a and e, respectively, the semimajor axis and eccentricity of the orbit of the asteroid, the Delaunay actions are:

$$L := \sqrt{a}, \qquad G := \sqrt{a(1-e^2)}. \tag{13}$$

The conjugated angle variables are defined as follows: l is the *mean anomaly*, which provides the position of the asteroid along its orbit and g is the *longitude of perihelion*, namely the angle between the perihelion line and a fixed reference direction (see [20]).

The Hamiltonian function in Delaunay variables can be written as

$$H(L, G, l, g) := h(L) := -\frac{1}{2L^2}, \tag{14}$$

which shows that the system is integrable, since $H = h(L)$ depends only on the actions. From the equations of motion $\dot{L} = 0$, $\dot{G} = 0$, we immediately recognize that L and G are constants, $L = L_0$, $G = G_0$, which in view of (13) is equivalent to say that the orbital elements (semimajor axis and eccentricity) do not vary along the motion. Being also g constant ($\dot{g} = 0$), we obtain that the orbit is a fixed ellipse with one of the foci coinciding with the Sun. Finally, the mean anomaly is obtained from $\dot{l} = \frac{\partial H(L)}{\partial L} := \omega(L)$ as $l(t) = \omega(L_0)t + l(0)$. Let us remark that the Hamiltonian (14) does not depend on all the actions, being independent on G: such kind of systems are called *degenerate* and often arise in Celestial Mechanics.

3.2 The three–body problem

The two–body problem describes only a rough approximation of the motion of the asteroid around the Sun; indeed, the most important contribution we neglected while considering Kepler's model comes from the gravitational influence of Jupiter. We are thus led to consider the motion of the three bodies: the Sun (S), Jupiter (J) and a minor body of the asteroidal belt (A). We restrict our attention to the special case of the *planar, circular, restricted three–body problem*. More precisely, we assume that the Sun and Jupiter revolve around their common center of mass, describing circular orbits (circular case). Choose the units of length, mass and time so that that the gravitational constant, the orbital angular velocity and the sum of the masses of the primaries (Sun and

Jupiter) are identically equal to one. Consider the motion of an asteroid A moving in the same orbital plane of the primaries (planar case). Assume that the mass of A is negligible with respect to the masses of S and J; this hypotheses implies that the motion of the primaries is not affected by the gravitational attraction of the asteroid (restricted case). Finally, let us identify the mass of J with a suitable small parameter ε. Though being the simplest (non trivial) three–body model, as shown by Poincaré [18] such problem cannot be explicitly integrated (like the two–body problem through Kepler's solution). In order to introduce the Hamiltonian function associated to such problem, it is convenient to write the equations of motion in a barycentric coordinate frame (with origin at the center of mass of the Sun–Jupiter system), which rotates uniformly at the same angular velocity of the primaries. The resulting system is described by a nearly–integrable Hamiltonian function with two degrees of freedom (see [20]) with the perturbing parameter ε representing the Jupiter–Sun mass–ratio.

We immediately recognize that for $\varepsilon = 0$ (i.e., neglecting Jupiter), the system reduces to the unperturbed two–body problem. As described in the previous section, we can identify the Delaunay elements with action–angle variables; the only difference is that in the rotating reference frame the variable g represents the longitude of the pericenter, evaluated from the axis coinciding with the direction of the primaries. If $H = h^{(AS)}(L)$ denotes the Hamiltonian function associated to the asteroid–Sun problem, it can be shown [20] that the Hamiltonian describing the three–body problem has the form

$$H(L, G, l, g; \varepsilon) := h^{(AS)}(L) - G + \varepsilon f(L, G, l, g; \varepsilon), \qquad (15)$$

for a suitable analytic function f, which represents the interaction between the asteroid and Jupiter. The Hamiltonian (15) is a prototype of a nearly–integrable system, since the integrable two–body problem is recovered as soon as the perturbing parameter ε is set equal to zero.

The action–angle formalism provides a standard tool to solve explicitly the equations of motion associated to integrable systems; on the other side, one could expect that for small perturbations the behavior of nearly–integrable systems is similar to the integrable ones. By (12) this remark is obviously true as long as the time is less than $1/\varepsilon$, though the question remains open for longer time scales. In order to face this problem, one could naively try to perform a change of variables, which transforms the nearly–integrable system (11) into a trivially integrable one (at least for ε small enough). However, a natural question concerns the existence of a canonical transformation $\Psi := (J, \psi) \to (I, \varphi)$ such that on a given time scale the nearly–integrable Hamiltonian system (11) is transformed into a new Hamiltonian system, $K := H \circ \Psi$, which does not depend on the new angle coordinates on a given time scale; perturbation theory will provide a tool to investigate such strategy of approaching nearly–integrable systems.

4 Perturbation theory

Perturbation theory flowered during the last two centuries through the works of Laplace, Leverrier, Delaunay, Poincaré, Tisserand, etc.; it provides constructive methods to investigate the behavior of nearly–integrable systems. The importance of studying the effects of small Hamiltonian perturbations on an integrable system was pointed out by Poincaré, who referred to it as the *fundamental problem of dynamics*. We introduce perturbation theory in section 4.1 and we present in section 4.2 the celebrated Kolmogorov–Arnold–Moser theorem.

4.1 Classical perturbation theory

Consider an analytic Hamiltonian of the form

$$H(I,\varphi;\varepsilon) := h(I) + \varepsilon f(I,\varphi;\varepsilon), \tag{16}$$

where $I \in B_R := \{I \in \mathbb{R}^n, \quad |I| \leq R\}$, $\varphi \in \mathbb{T}^n$ and $|\varepsilon| \leq \bar{\varepsilon}$ for suitable real constants $R > 0$, $\bar{\varepsilon} > 0$. Let us expand f in Taylor series of ε as

$$f(I,\varphi;\varepsilon) =: \sum_{j\geq 0} \varepsilon^j f^{(j)}(I,\varphi),$$

for suitable functions $f^{(j)}(I,\varphi)$, which can be expanded in Fourier series as $f^{(j)}(I,\varphi) = \sum_{k\in\mathbb{Z}^n} f_k^{(j)}(I)\mathrm{e}^{\mathrm{i}k\cdot\varphi}$. We implement an integrating transformation

$$\Psi : (J,\psi) \longmapsto (I,\varphi) \ ,$$

defined by the implicit equations

$$I = J + \varepsilon \partial_\psi S(J,\psi;\varepsilon), \qquad \varphi = \psi + \varepsilon \partial_J S(J,\psi;\varepsilon), \tag{17}$$

where the generating function S can be expanded as a Poincaré–Lindstedt series of the form

$$S(J,\psi;\varepsilon) = \sum_{j\geq 0} \varepsilon^j S^{(j)}(J,\psi;\varepsilon)$$

for suitable analytic functions $S^{(j)}$ to be determined as follows.

By the implicit function theorem it is simple to prove that, for ε small enough, (17) defines a good diffeomorphism, which is also canonical. We want to determine recursively $S^{(j)}$ so that the function $K := H \circ \Psi$ does not depend on ψ:

$$\Psi(J,\psi) = (J + \varepsilon \frac{\partial S^{(0)}(J,\psi;\varepsilon)}{\partial \varphi} + O(\varepsilon^2),\ \psi + O(\varepsilon)).$$

To this end, we start by expanding $H \circ \Psi$ as

$$(H \circ \Psi)(J,\psi) = h(J) + \varepsilon\Big[\nabla h(J)\cdot\partial_\psi S^{(0)}(J,\psi;\varepsilon) + f^{(0)}(J,\psi)\Big] + O(\varepsilon^2). \tag{18}$$

We look for $S^{(0)}$ so that the term of order ε does not depend on the angles; we are thus led to the equation

$$\nabla h(J)\cdot\partial_\psi S^{(0)}(J,\psi;\varepsilon) + f^{(0)}(J,\psi) = h^{(1)}(J;\varepsilon), \tag{19}$$

where $h^{(1)}(J;\varepsilon) \equiv f_0^{(0)}(J)$ and

$$S^{(0)}(J,\varphi) := \mathrm{i}\sum_{k\neq 0}\frac{f_k^{(0)}(J)}{\nabla h(J)\cdot k}\mathrm{e}^{\mathrm{i}k\cdot\varphi}\ . \tag{20}$$

The above expression contains a quantity at the denominator to which we refer as the *small divisor*, since it can become arbitrarily small:

$$\nabla h(J)\cdot k. \tag{21}$$

Indeed, the function $S^{(0)}$ can be defined only for values of the actions such that the small divisors (21) are different from zero, namely only for J belonging to the set of rational independent frequencies

$$\aleph := \{J \in \mathbb{R}^n,\ \text{such that } \nabla h(J)\cdot k \neq 0, \forall k \in \mathbb{Z}^n\setminus\{0\}\}.$$

Since $\aleph$ has empty interior, if we want to define $S^{(0)}$ in an open neighborhood of a fixed $J_0 \in \aleph$, we must truncate the series in (20) up to a suitable order, say $|k| \leq d_0$ for a given $d_0 \in \mathbb{Z}_+$. In particular, let us write the Fourier expansion of $f^{(0)}$ as

$$f^{(0)}(J,\psi) = \sum_{|k|\leq d_0} f_k^{(0)}(J)\mathrm{e}^{\mathrm{i}k\cdot\psi} + \sum_{|k|>d_0} f_k^{(0)}(J)\mathrm{e}^{\mathrm{i}k\cdot\psi}\ ; \tag{22}$$

choosing a sufficiently large value of the truncation index $d_0 := d_0(\varepsilon)$, we can make the second sum in (22) of order ε, so that it will finally contribute to the $O(\varepsilon^2)$–term in the development (18). In summary, we have eliminated the angles in the expression (18) up to $O(\varepsilon)$ by using the transformation associated to $S^{(0)}$, which is defined for $J \in B_{\varrho_0}(J_0)$, for a sufficiently small $\varrho_0 := \varrho_0(\varepsilon)$ such that no zero–divisors will occur, i.e.

$$\nabla h(J)\cdot k \neq 0, \qquad \forall 0 < |k| \leq d_0(\varepsilon), \quad J \in B_{\varrho_0(\varepsilon)}(J_0).$$

In order to eliminate the angle dependence in (18) for the terms of order $\varepsilon^2, \varepsilon^3, \ldots, \varepsilon^{m+1}, \ldots$, we determine $S^{(1)}$, $S^{(2)}$, …, $S^{(m)}$, …, by solving equations similar to (19). Again, we need to truncate the series associated to $S^{(m)}$ at the orders $|k| \leq d_m$ for sufficiently large indexes

$$d_m := d_m(\varepsilon) \ \longrightarrow \infty \quad \text{for} \quad m \to \infty.$$

Therefore, $S^{(m)}$ will be defined only for $J \in B_{\varrho_m}(J_0)$ for sufficiently small values of the radii, such that

$$\varrho_m := \varrho_m(\varepsilon) \longrightarrow 0 \quad \text{for} \quad m \to \infty.$$

As a consequence, we remark that the radii of the domains of definition of the functions $S^{(0)}$, $S^{(1)}$, ..., $S^{(m)}$, ... will drastically shrink to 0 as m increases, so that we cannot expect that the overall procedure will converge on an open set.

Finally, iterating infinitely many times the above procedure we can formally write the resulting Hamiltonian function as

$$K(J;\varepsilon) := H \circ \Psi(J,\psi;\varepsilon) = h(J) + \varepsilon h_\infty(J;\varepsilon),$$

for a suitable function h_∞. However, we are immediately faced with the problem of the convergence of such procedure, namely with the question of the existence of $J_0 \in \aleph$ and $\varepsilon_0 > 0$, such that $K(J_0;\varepsilon)$ is well–defined for $|\varepsilon| \leq \varepsilon_0$. In general, as shown by Poincaré, the answer to this question is negative. We report here an example quoted in [2] of a diverging Poincaré–Lindstedt series.

Consider the Hamiltonian

$$H(I,\varphi;\varepsilon) := \omega \cdot I + \varepsilon \left[I_1 + \sum_{k\in\mathbb{N}^2\setminus\{0\}} a_k \sin(k\cdot\varphi) \right], \tag{23}$$

$(I,\varphi) \in \mathbb{R}^2 \times \mathbb{T}^2$, where $a_k := \exp(-|k|)$ and $\omega = (\omega_1,\omega_2)$, $\omega_1 < 0$, $\omega_2 > 0$ with $\omega_1/\omega_2 \in \mathbb{R}\setminus\mathbb{Q}$. For $\varepsilon = 0$ the phase space is foliated by non–resonant invariant tori, wrapped by the quasi–periodic flow $t \to \omega t + \varphi_0$. In this example, it is very simple to evaluate the Poincaré–Lindstedt series at any order, though it is not necessary for our purposes. Indeed, we just notice that if the Poincaré–Lindstedt series converges, we can define the canonical transformation $\Psi_\varepsilon : (J,\psi) \to (I,\varphi)$, which integrates the system, i.e. $H \circ \Psi_\varepsilon(J,\psi) =: K(J;\varepsilon)$. Therefore, from (6) the values of I_1 and I_2 remain bounded in time, since J is a constant vector. On the other hand, we can readily integrate (23) and in particular we have $\varphi_1(t) = (\omega_1+\varepsilon)t + \varphi_1(0)$ and $\varphi_2(t) = \omega_2 t + \varphi_2(0)$, so that the solution for I_1 (with initial condition $\varphi(0) = 0$) is given by

$$I_1(t) = I_1(0) - \varepsilon \sum_{k\in\mathbb{N}/\{0\}} a_k\; k_1 \int_0^t \cos\Big(\big((\omega_1+\varepsilon)k_1 + \omega_2 k_2\big)t\Big)\,\mathrm{d}t, \tag{24}$$

which can be solved by quadratures. Moreover, whenever

$$\frac{\omega_1+\varepsilon}{\omega_2} = \frac{p}{q} \qquad \text{for some} \quad p \in \mathbb{Z},\; q \in \mathbb{N},$$

the sum in (24) over all the terms with $(k_1,k_2) \neq (-q,p)$ gives a uniformly bounded contribution (for any $t \in \mathbb{R}$), while the term with $(k_1,k_2) = (-q,p)$

gives a contribution which goes to $\pm\infty$ when $t \to \pm\infty$, since we identically have

$$(\omega_1 + \varepsilon)k_1 + \omega_2 k_2 = 0.$$

This computation allows to conclude that the Poincaré–Lindstedt series diverges.

As a final remark, we note that if the Poincaré–Lindstedt series converges for some $J_0 \in \aleph$ and $|\varepsilon| \leq \varepsilon_0$, then for any $|\varepsilon| \leq \varepsilon_0$ the perturbed system admits the invariant torus

$$\mathcal{T}_{\omega_0} := \Big\{ \Psi(J_0, \psi; \varepsilon), \quad \psi \in \mathbb{T}^n \Big\}$$

with frequency $\omega_0 := \nabla h(J_0) + \varepsilon \nabla h^{(1)}(J_0; \varepsilon) + ...$ In general, the rational dependence of the vector ω_0 will vary according to the values of ε. Therefore the torus $\mathcal{T}_{\omega_0}$ will be non–resonant or resonant according to the values of ε. As we have seen before, a resonant torus is foliated into lower dimensional tori; actually, this situation is highly degenerate and it determines the intrinsic reason for the divergence of the Poincaré–Lindstedt series.

4.2 KAM theory

The Kolmogorov–Arnold–Moser (KAM) theory provides a constructive method to investigate nearly–integrable Hamiltonian systems [15], [1], [17]. In analogy to the Poincaré–Lindstedt theory, the basic idea relies on the elimination of the angle variables through suitable changes of coordinates, with the further requirement that the sequence of transformations will be *quadratically* convergent. Ignoring the size of the contribution of the small divisors, after performing the m–th change of coordinates the part of the remainder which depends on the angle variables will be of order ε^{2^m}. The superconvergent estimate of the remainder terms counteracts the influence of the small divisors ensuring the convergence of the KAM procedure on a suitable nonresonant set. Indeed, the second ingredient of KAM theory is to focus on a given strongly rationally independent frequency, rather than on a given action.

To be more precise, consider an analytic Hamiltonian of the form (16). Fix a rationally independent frequency vector ω_0, such that $\omega_0 := \nabla h(J_0)$ for a suitable $J_0 \in \mathbb{R}^n$. We perform a canonical change of variables Ψ_1, implicitly defined as in (17) with $S = S^{(0)}$, where $S^{(0)}$ can be determined as in section 4.1. Due to the strong rational independence of ω_0, for $J \approx J_0$ the following estimate on the small divisors holds

$$|\nabla h(J) \cdot k| \approx |\omega_0 \cdot k| \gg 0. \tag{25}$$

The transformed Hamiltonian becomes

$$\mathcal{K}_1 := H \circ \Psi_1 = h(J) + \varepsilon h^{(1)}(J;\varepsilon) + \varepsilon^2 \mathcal{R}_2(J,\psi;\varepsilon)$$

for a suitable remainder function $\mathcal{R}_2$.

Next step is substantially different from the corresponding case of section 4.1; let $h^*(J;\varepsilon) := h(J) + \varepsilon h^{(1)}(J;\varepsilon)$ be the integrable part and let $\varepsilon^2\mathcal{R}_2(J,\psi;\varepsilon)$ be the perturbation. The crucial advantage is that after one transformation the new perturbative parameter is $\delta := \varepsilon^2$ with Hamiltonian $\mathcal{K}_2 = h^* + \delta\mathcal{R}_2$. Then we introduce a second canonical change of variables Ψ_2, defined in a neighborhood of a suitable $J_1(\varepsilon)$ which is chosen to satisfy $\nabla h^*(J_1(\varepsilon);\varepsilon) = \omega_0$; this relation will ensure a good estimate on the small divisors as in (25). For ε small enough, we can find $J_1(\varepsilon)$ by the implicit function theorem, if we assume that

$$\det h''(J_0) \neq 0. \tag{26}$$

It is easy to see that after the last change of variables the new remainder term will be of order $\delta^2 = \varepsilon^{2^2}$.

Iterating this procedure, at the m–step the angle–depending remainder will be of order ε^{2^m}. Finally, applying infinitely many times the KAM scheme, we end up with an integrable Hamiltonian of the form

$$\mathcal{K}(\mathcal{J};\varepsilon) := H \circ \Psi_1 \circ \Psi_2 \circ \ldots =: H \circ \Psi_\infty(\mathcal{J},\phi;\varepsilon), \tag{27}$$

which is defined for $\mathcal{J} = J_\infty(\varepsilon)$ and it satisfies

$$\nabla\mathcal{K}(J_\infty(\varepsilon);\varepsilon) = \omega_0.$$

We have thus proved the existence of the invariant torus

$$\mathcal{T}_{\omega_0} := \Big\{\Psi_\infty(J_\infty(\varepsilon),\phi;\varepsilon), \quad \phi \in \mathbb{T}^n\Big\},$$

provided that the overall procedure converges on a non trivial domain. Actually such domain results to be a Cantor set; the equality (27) holds precisely on such set and can be differentiated infinitely many times on it, see [19], [7].

Remark 1. Before stating the KAM theorem, let us summarize the main differences between Poincaré–Lindstedt series and KAM procedures.

In the first case:

(1) after the m–th coordinate's transformation the remainder term (depending upon angles) is of order ε^{m+1},
(2) the initial datum J_0 is kept fixed, while the final frequency $\nabla h(J_0;\varepsilon)$ (respectively the invariant surface $\mathbb{T}_{\nabla h(J_0;\varepsilon)}$) varies with ε, eventually becoming rationally dependent (respectively, a resonant torus).

Concerning the KAM procedure we recall that:

(1) after the m–th coordinate's transformation the remainder term (depending upon angles) is of order ε^{2^m} (ignoring the contribution of the small divisors[4]),

(2) the frequency vector ω_0 is fixed and it is supposed to be strongly rational independent, so that the corresponding torus $\mathbb{T}_{\omega_0}$ is strongly non resonant.

We now proceed to state the KAM theorem as follows. Consider a Hamiltonian system as in (11). As we discussed previously, for $\varepsilon = 0$ the phase space of the unperturbed (integrable) system h is foliated by n–dimensional invariant tori labeled by $I = I_0$. Such tori are resonant or nonresonant according to the fact that the frequency $\omega_0 := \nabla h(I_0)$ is rationally dependent or not. KAM theorem describes the fate of nonresonant tori under perturbation. We recall that the three assumptions necessary to prove the theorem are the following: the smallness of ε, the strong rational independence of ω_0 and the nondegeneracy of the unperturbed Hamiltonian h as given in (26).

Theorem 1 (KAM Theorem). *If the unperturbed system is nondegenerate, then, for a sufficiently small perturbation, most nonresonant invariant tori do not break–down, though being deformed and displaced with respect to the integrable situation. Therefore, in the phase space of the perturbed system there exist invariant tori densely filled with quasi–periodic phase trajectories winding around them. These invariant tori form a majority, in the sense that the measure of the complement of their union is small when the perturbation is small.*

The proof of the KAM theorem is based on the superconvergent procedure described above. We remark that the strong rational independence of ω_0 (see (25)) plays a central role. In particular, the KAM scheme works if ω_0 satisfies the so–called *Γ–τ Diophantine condition*

$$|\omega_0 \cdot k| \geq \frac{\Gamma}{|k|^\tau}, \quad \forall\, k \in \mathbb{Z}^n \setminus \{0\}, \tag{28}$$

for a suitable $\Gamma = \Gamma(\varepsilon) > 0$ and $\tau > n - 1$.

More precisely, it results that all tori having Γ–τ Diophantine frequency vector with

$$\Gamma > const. \times \sqrt{\varepsilon}$$

are preserved under the perturbation.

We remark that the KAM result is global, in the sense that for all $\varepsilon \leq \varepsilon_0$ (ε_0 fixed) and for all Γ–τ Diophantine frequency vectors ω_0 satisfying (28), the Hamiltonian system (11) admits an invariant perturbed torus with frequency vector ω_0. We refer to [9], [11], [6], [12] for different methods to construct

[4] Taking into account the contribution of the small divisors, one can nevertheless obtain a superexponential decay even if it is not strictly necessary for the convergence of the KAM scheme.

invariant tori through a classical proof of the convergence of the perturbative series involved.

It is worth stressing that KAM theorem can be also stated as follows. Consider an unperturbed ($\varepsilon = 0$) surface with a fixed diophantine frequency ω_0 and look at its fate when the perturbation is switched on ($\varepsilon > 0$). Then, for ε small enough, the torus persists, being deformed and displaced with respect to the integrable limit, until the perturbing parameter ε reaches a threshold $\varepsilon_c := \varepsilon_c(\omega_0)$ at which the surface is destroyed (i.e., it looses regularity). For low–dimensional systems, KAM theory allows to prove a strong stability result concerning the confinement of the action variables in phase space.

Theorem 2. *In a non–degenerate system with two degrees of freedom, for $\varepsilon \leq \varepsilon_0$ (for a suitable sufficiently small $\varepsilon_0 > 0$) and for all initial conditions, the values of the actions remain forever near their initial values.*

The above statement is based on the following remark. For a two degrees of freedom Hamiltonian system, the phase space is four–dimensional, the energy level sets (on which the motion takes place) are three–dimensional, while the KAM tori are two–dimensional, filling a large part of each energy level. Any orbit starting in the region between two invariant tori remains forever trapped in it.

As an immediate corollary of the previous theorem, the stability of the actions in the planar, circular, restricted three–body problem follows. Using the original versions of Arnold's theorem, M. Hénon [14] proved the existence of invariant tori provided that the perturbing parameter is less than 10^{-333}; this value can be improved by implementing Moser's theorem, which yields an estimate of the order of 10^{-50}. We recall that the perturbing parameter represents the ratio of the masses of Jupiter and the Sun; its astronomical value amounts to about 10^{-3}.

Accurate analytical estimates based on a computer–assisted implementation allow to drastically improve such results by showing the existence of invariant tori on a preassigned energy level for parameter values less or equal than 10^{-3}, in agreement with the physical measurements (see [4] for further details).

5 A discrete model: the standard map

In the previous sections we focused our attention on *continuous* systems; now we want to introduce *discrete* systems, which can be viewed as surfaces of section of Hamiltonian flows. Such models are definitely simpler than continuous systems, since their evolution can be known without introducing any numerical error due to integration algorithms, though roundoff errors cannot be avoided. Let us start by considering the familiar pendulum model described by equation (3). In order to compute numerically the trajectory associated to Hamilton's equations for (3), i.e.

$$\begin{aligned} \dot{r} &= \varepsilon \sin\theta \\ \dot{\theta} &= r, \end{aligned} \tag{29}$$

we can use a leap-frog method, such that if T is the time–step and (r_n, θ_n) denotes the solution at time nT, then (29) can be integrated as

$$\begin{aligned} r_{n+1} &= r_n + T\varepsilon \sin\theta_n \\ \theta_{n+1} &= \theta_n + Tr_{n+1}. \end{aligned}$$

Normalizing the time–step to one, we are reduced to the study of the so–called *standard map* introduced by Chirikov in [8]:

$$\begin{aligned} r_{n+1} &= r_n + \varepsilon \sin\theta_n \\ \theta_{n+1} &= \theta_n + r_{n+1}, \end{aligned} \tag{30}$$

with $r_n \in \mathbb{R}$ and $\theta_n \in \mathbb{T} \equiv \mathbb{R}/(2\pi\mathbb{Z})$. We will also consider the generalized standard map, which is obtained replacing the sine term in (30) by any periodic, continuous function $f(\theta)$:

$$\begin{aligned} r_{n+1} &= r_n + \varepsilon f(\theta_n) \\ \theta_{n+1} &= \theta_n + r_{n+1} = \theta_n + r_n + \varepsilon f(\theta_n), \end{aligned} \tag{31}$$

where $r_n \in \mathbb{R}$ and $\theta_n \in \mathbb{T} \equiv \mathbb{R}/(2\pi\mathbb{Z})$. We notice that the Jacobian associated to (31) is identically one, being the map area–preserving. We remark also that for $\varepsilon = 0$ the mapping reduces to a simple rotation, i.e.

$$\begin{aligned} r_{n+1} &= r_n = r_0 \\ \theta_{n+1} &= \theta_n + r_0. \end{aligned}$$

We refer to $\omega \equiv r_0$ as the *frequency* or *rotation number* of the unperturbed mapping, which is generally defined as

$$\omega \equiv \lim_{n\to\infty} \frac{\theta_n}{n}.$$

In the unperturbed case the motion takes always place on an invariant circle. If $\omega \in \mathbb{Q}$, the trajectory described by the iteration of the mapping with initial data (r_0, θ_0) is a periodic orbit; if $\omega \in \mathbb{R}\backslash\mathbb{Q}$, the motion fills densely an invariant curve with frequency ω, say $\mathcal{C}_{0,\omega}$. When $\varepsilon \neq 0$ the system becomes non–integrable and we proceed to describe the conclusions which can be drawn by applying perturbation theory. As we described in the previous sections, KAM theorem [15], [1], [17] ensures that if ε is sufficiently small, there still exists for the perturbed system an invariant curve $\mathcal{C}_{\varepsilon,\omega}$ with frequency ω.

The KAM theorem can be applied provided that the frequency ω satisfies a strong irrationality assumption, namely the diophantine condition (28), which can now be rephrased as

$$|\frac{\omega}{2\pi} - \frac{p}{q}| \geq \frac{1}{Cq^2}, \qquad \forall p, q \in \mathbb{Z} \setminus \{0\},$$

for some positive constant C. As the perturbing parameter is increased, the invariant curve becomes more and more distorted and displaced, until one reaches a critical value at which $\mathcal{C}_{\varepsilon,\omega}$ breaks–down. Let us define the critical break–down threshold $\varepsilon_c(\omega)$ as the supremum of the positive values of ε, such that there still exists an invariant curve with rotation number ω.

We provide a mathematical definition of $\mathcal{C}_{\varepsilon,\omega}$ as the invariant curve described by the parametric equations

$$\begin{aligned} r_n &= \omega + U(\varphi_n) - U(\varphi_n - \omega), \\ \theta_n &= \varphi_n + U(\varphi_n), \end{aligned} \tag{32}$$

where $\varphi_n \in \mathbb{T}$ and the conjugating function $U(\varphi_n)$ is 2π–periodic and analytic in ε; moreover, the parameterization is defined so that the flow is linear, i.e. $\varphi_{n+1} = \varphi_n + \omega$.

The function U can be expanded as a Taylor series in ε and a Fourier series in φ (the so–called Poincaré–Lindstedt series) as

$$U(\varphi_n) \equiv \sum_{j=1}^{\infty} \varepsilon^j U_j(\varphi_n) = \sum_{j=1}^{\infty} \varepsilon^j \sum_{l \in \mathbb{Z}} \hat{U}_{lj} \mathrm{e}^{\mathrm{i} l \varphi_n}$$

for suitable Taylor coefficients U_j and Fourier–Taylor coefficients $\hat{U}_{lj}$. We define the *radius of convergence* of the Poincaré–Lindstedt series as

$$\varrho_c(\omega) = \inf_{\varphi \in \mathbb{T}} (\limsup_{j \to \infty} \ |U_j(\varphi)|^{1/j})^{-1} .$$

We intend to study the relation between the two quantities $\varepsilon_c(\omega)$ and $\varrho_c(\omega)$ in the complex parameter space (i.e., taking $\varepsilon \in \mathbb{C}$). In particular, we want to investigate the shape of the domain of convergence of the Poincaré–Lindstedt series and the location of the complex singularities. This task will be performed by looking at the behavior of the periodic orbits, which approach more closely the invariant curve with frequency ω. We stress that if the domain of analyticity is not a circle, then the analyticity radius and the critical break–down threshold might be very different.

The tools adopted to investigate the shape of the analyticity domains are based on different methods: Greene's technique [13], the computation of Padé approximants and a recent method developed in [5].

5.1 Link between periodic orbits and invariant curves

In order to familiarize with the concepts of periodic orbits and invariant curves, let us consider a specific example provided by the mapping (30) and by the invariant curve with frequency proportional to the golden ratio: $\omega = 2\pi\frac{\sqrt{5}-1}{2}$. As it is well known, the golden ratio is approximated by the sequence of Fibonacci's numbers $\{F_k/F_{k+1}\} \in \mathbb{Q}$, defined as

$$F_{k+1} = F_k + F_{k-1} \quad (k \geq 1) \qquad \text{with} \qquad F_0 = 0, \quad F_1 = 1.$$

Figure 1 shows in the (θ, r)–plane (with θ normalized by a factor 2π), how the different periodic orbits with frequency $2\pi F_k/F_{k+1}$ approach the golden–mean invariant curve.

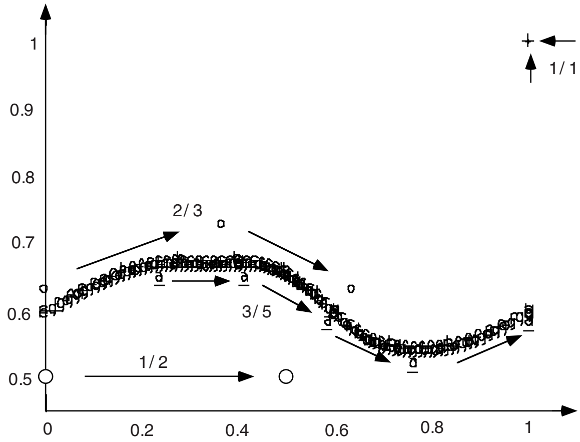


Fig. 1. The invariant curve with rotation number $\omega = 2\pi\frac{\sqrt{5}-1}{2}$ and its approximating periodic orbits with frequencies (proportional by a factor 2π) 1, 1/2, 2/3, 3/5 for the mapping (30) in the (θ, r)–plane.

In order to explore the link between periodic orbits and invariant curves, it is useful to recall some simple notions of number theory and in particular about rational approximants.

For any $\omega \in \mathbb{R}$, let the *continued fraction representation* of ω be defined as the sequence of integer numbers a_j, such that

$$\omega = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \ldots}} \equiv [a_0; a_1, a_2, \ldots] \, .$$

If $\omega \in \mathbb{R} \setminus \mathbb{Q}$, then the sequence of the $\{a_j\}$'s is infinite; if $\omega \in \mathbb{Q}$, the sequence is finite: $\{a_j\} = \{a_1, ..., a_N\}$. For example, in the case of the golden ratio, one has:

$$\frac{\sqrt{5}-1}{2} = [0; 1, 1, 1, 1...] \equiv [0; 1^\infty] \,.$$

Those numbers whose continued fraction expansion is eventually 1, i.e.

$$\alpha = [a_0; a_1, ..., a_N, 1, 1, 1, 1....]$$

for some integer N, are called *noble* numbers. Let $\{p_j/q_j\}$ be the sequence of rational approximants to ω, whose terms are computed as the truncations of the continued fraction expansion, i.e.

$$\begin{aligned}
\frac{p_0}{q_0} &= a_0 \\
\frac{p_1}{q_1} &= a_0 + \frac{1}{a_1} \\
\frac{p_2}{q_2} &= a_0 + \frac{1}{a_1 + \frac{1}{a_2}} \\
& \qquad\qquad\qquad\qquad ...
\end{aligned}$$

The rational numbers $\{p_j/q_j\}$ are the *best rational approximants* to the irrational number α.

5.2 Perturbative series expansions

Let us consider the standard map defined in (30); by the relations $\theta_{n+1} - \theta_n = r_{n+1}$ and $\theta_n - \theta_{n-1} = r_n$, one obtains that θ must satisfy the equation

$$\theta_{n+1} - 2\theta_n + \theta_{n-1} = \varepsilon \sin\theta_n. \tag{33}$$

Let us look for a periodic solution with frequency $\omega = 2\pi p/q$, satisfying the periodicity conditions

$$\begin{aligned}
\theta_{n+q} &= \theta_n + 2\pi p \\
r_{n+q} &= r_n.
\end{aligned} \tag{34}$$

In analogy to (32), we parameterize the solution as

$$\theta_n = \varphi_n + u(\varphi_n),$$

where $\varphi_n \in \mathbb{T}$ and $u(\varphi_n)$ is $2\pi p$–periodic; moreover, the flow is linear with frequency $2\pi p/q$: $\varphi_{n+1} = \varphi_n + 2\pi p/q$ (notice that the conditions (34) are trivially satisfied). Next, we expand u in Fourier–Taylor series as

$$u(\varphi_n) \equiv \sum_{j=1}^{\infty} u_j(\varphi_n)\varepsilon^j = \sum_{j=1}^{\infty} \varepsilon^j \sum_{l=1}^{\min(j,q)} a_{lj} \sin(l\varphi_n), \tag{35}$$

where the real coefficients a_{lj} will be recursively determined by means of (33).

Notice that the reason for which the summation ends at $\min(j,q)$ is due to the fact that one needs to avoid zero divisors, as an explicit computation shows.

From (33), one finds that u must satisfy the relation

$$u(\varphi_n + 2\pi\frac{p}{q}) - 2u(\varphi_n) + u(\varphi_n - 2\pi\frac{p}{q}) = \varepsilon\sin(\varphi_n + u(\varphi_n)). \qquad (36)$$

The coefficients a_{lj} can be obtained by inserting the series expansion (35) in (36) and equating same orders of ε. More precisely, defining $\beta_{l,j}$ such that

$$\sin(\varphi_n + u(\varphi_n)) \equiv \sum_{j=0}^{\infty}\Big(\sum_{l=1}^{j+1}\beta_{l,j+1}\sin(l\varphi_n)\Big)\varepsilon^j ,$$

one obtains (see [5]) that the coefficients a_{lj} in (35) are given by

$$a_{lj} = \frac{\beta_{lj}}{2[\cos(2\pi lp/q) - 1]} .$$

At the order ε^q, one meets a singularity whenever $l = q$. In order to bypass this problem, one is forced to select φ_n so that the sine term in (35) is zero, i.e.

$$\sin(q\varphi_n) = 0 ; \qquad (37)$$

such choice compensates the zero term (for $l = q$) occurring in

$$\cos(2\pi q\ \frac{p}{q}) - 1.$$

Equation (37) provides a value for φ_n as well as for φ_0, since $\varphi_n = \varphi_0 + 2\pi np/q$. Correspondingly, one has two solutions (modulus 2π), given by

$$\varphi_0 = \frac{\pi}{q} \qquad \text{and} \qquad \varphi_0 = \frac{2\pi}{q} .$$

The two solutions are stable or unstable according to the parity of q. In particular π/q is stable for q odd, unstable for q even, while the opposite situation occurs for $2\pi/q$. Due to (37) the coefficient a_{qq} is not determined by the recursive formulae; however, this term does not contribute to the general solution since $\sin(q\varphi_0) = 0$.

Remark 2. For an invariant curve with rotation number ω, the conjugating function must satisfy the equation

$$U(\varphi_n + \omega) - 2U(\varphi_n) + U(\varphi_n - \omega) = \varepsilon\sin(\varphi_n + U(\varphi)).$$

Notice that in this case the function $U = U(\varphi_n;\varepsilon)$ depends explicitly on the parametric coordinate φ_n.

6 Numerical investigation of the break–down threshold

6.1 Padé approximants

A numerical investigation of the analyticity domains of periodic orbits is performed using the Padé approximants of order $[N/N]$, i.e. $P_N(\varepsilon)$, $Q_N(\varepsilon)$, such that $\sum_{j=1}^{\infty} u_j\varepsilon^j \equiv \frac{P_N(\varepsilon)}{Q_N(\varepsilon)} + O(\varepsilon^{2N+1})$. The shape of the analyticity domain will be provided by the zero of the denominators. In particular, having fixed a value for φ_0, we consider the series

$$u(\varphi_0) = \sum_{j=1}^{\infty} u_j(\varphi_0)\varepsilon^j, \qquad \varepsilon \in \mathbb{C}. \tag{38}$$

We compute the Padé approximants of order [200/200], where for consistency the coefficients $u_j(\varphi_0)$ must be calculated with a precision of 400 decimal digits. False poles have been discarded by comparison with the zeros; indeed, we recognize a pole as spurious whenever its coordinates are close to a pole within a suitable tolerance parameter.

We study the periodic orbits generated by the best approximants to the rotation numbers defined as $\gamma = 2\pi[0;1^\infty]$, $\omega_1 = 2\pi[0;3,12,1^\infty]$, $\omega_2 = 2\pi[0;2,10,1^\infty]$; the corresponding sequences of rational approximants are

$$\{\frac{1}{2},\frac{2}{3},\frac{3}{5},\frac{5}{8},\frac{8}{13},\frac{13}{21},\frac{21}{34},\frac{34}{55},\frac{55}{89},\frac{89}{144},...\} \to \frac{\gamma}{2\pi}$$

$$\{\frac{1}{3},\frac{12}{37},\frac{13}{40},\frac{25}{77},\frac{38}{117},...\} \to \frac{\omega_1}{2\pi}$$

$$\{\frac{10}{21},\frac{11}{23},\frac{21}{44},\frac{32}{67},\frac{53}{111},\frac{85}{178},...\} \to \frac{\omega_2}{2\pi}.$$

Figure 2 (see [5]) shows the Padé approximants of the periodic orbits 3/5, 13/21, 34/55, 89/144 (times 2π), associated to the mapping (30); the inner black region denotes the analyticity domain of the invariant curve with frequency γ. We remark that, as the order q of the periodic orbit grows, the singularities associated to the periodic orbits approach more and more the analyticity domain of the golden–mean invariant curve.

Similarly, the Padé approximants corresponding to ω_1 and ω_2 and to some of their rational approximants are shown, respectively, in Figures 3*a* and 3*b* (see [5]).

Let u satisfy (36) and let the *radius of convergence* of the series $u(\varphi_0) = \sum_{j=1}^{\infty} u_j(\varphi_0)\varepsilon^j$ be defined as

$$\varrho_c(\frac{p}{q}) = \Big(\limsup_{j\to\infty}\ |u_j(\varphi_0)|^{1/j}\Big)^{-1},$$

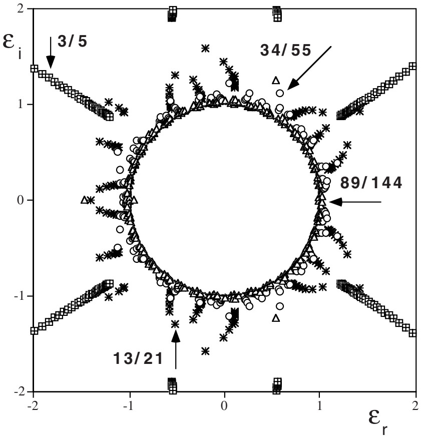


Fig. 2. Padé approximants of order [200/200] of the golden mean curve and of the periodic orbits with frequencies 3/5, 13/21, 34/55, 89/144 (times 2π) for the map (30) (after [5]).

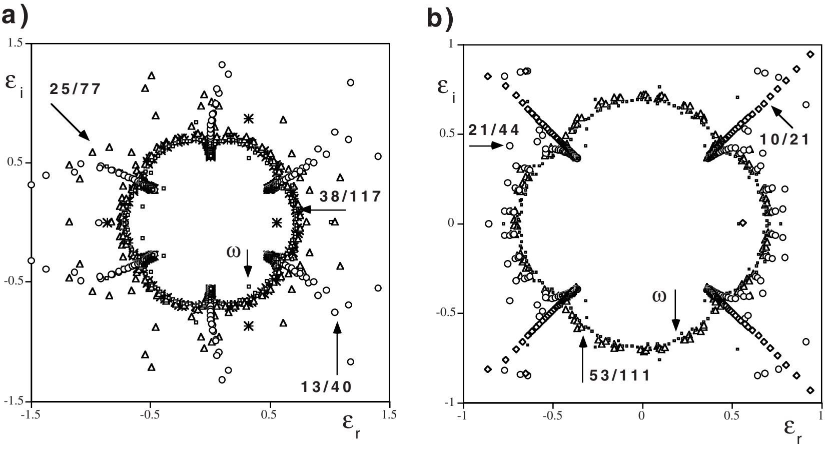


Fig. 3. Padé approximants of order [200/200] for the invariant curve with rotation number ω_1 (a) and ω_2 (b) and for some of their rational approximants in the framework of the mapping (30) (after [5]).

while, for ω irrational, we have already defined

$$\varrho_c(\omega) = \inf_{\varphi \in \mathbb{T}} \left(\limsup_{j \to \infty} \ |U_j(\varphi)|^{1/j} \right)^{-1} .$$

As a byproduct of our numerical analysis, we conjecture that

$$\lim_{k \to \infty} \varrho_c(\frac{p_k}{q_k}) = \varrho_c(\omega); ,$$

where $\{p_k/q_k\}$ are the rational approximants to ω.

Remark 3. The evidence that the domains of the approximating periodic orbits tend to a limiting domain seems to suggest that the position of the poles of the invariant curve does not depend on the specific value of the coordinate φ, which must be fixed while computing the Padé approximants.

In order to make our analysis more exhaustive, we investigate also different standard–map like systems; in particular, we consider the following examples:

$$f_{12}: \qquad r_{n+1} = r_n + \varepsilon(\sin\theta_n + \frac{1}{20}\sin 2\theta_n) \tag{39}$$

$$\theta_{n+1} = \theta_n + r_{n+1}, \tag{40}$$

$$f_{13}: \qquad r_{n+1} = r_n + \varepsilon(\sin\theta_n + \frac{1}{30}\sin 3\theta_n) \tag{41}$$

$$\theta_{n+1} = \theta_n + r_{n+1} \ . \tag{42}$$

Figure 4 (see [5]) provides the singularities of f_{12} (Figure 4a) and f_{13} (Figure 4b) for $\gamma = 2\pi\frac{\sqrt{5}-1}{2}$.

6.2 Lyapunov's method

In order to estimate the radius of convergence of the Poincaré–Lindstedt series (38), we review the algorithm proposed in [5], to which we refer as the *Lyapunov's method*. This technique consists in applying the following procedure:

1) consider discrete values of the small parameter ε from an initial ε_{in} to a final ε_{fin} with a relative increment $(1+h)\varepsilon$;

2) for any of these ε–values, compute the distance d_k between the truncated series at order k calculated at ε with that at $(1+h)\varepsilon$; more precisely, denoting by

$$u^{(k)}(\varphi_0;\varepsilon) \equiv \sum_{j=1}^{k} u_j(\varphi_0)\varepsilon^j ,$$

we define the quantity $d_k(\varepsilon)$ as

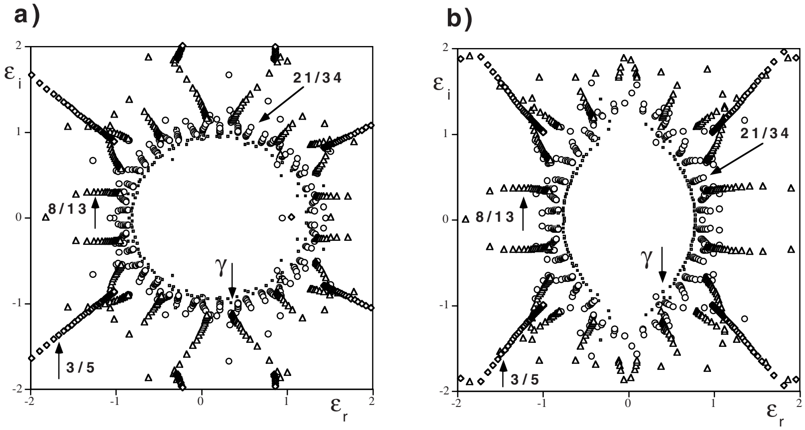


Fig. 4. Padé approximants of order [200/200] for the golden mean curve and some rational approximants associated to the mapping f_{12} (a) and f_{13} (b), respectively (after [5])

$$d_k = d_k(\varepsilon) \equiv |u^{(k)}(\varphi_0; (1+h)\varepsilon) - u^{(k)}(\varphi_0; \varepsilon)| \ ;$$

3) for $N \in \mathbb{Z}_+$ large enough, compute the sum

$$s_1 = s_1(\varepsilon) \equiv \frac{1}{N-1} \sum_{k=2}^{N} \log \frac{d_k}{d_1} \ ;$$

4) plot s_1 versus ε (see Figure 5a). Experimentally one notices that all graphs show an initial almost constant value of s_1 as ε is increased, followed by a small well, and then by a sharp increase with almost linear behavior;

5) estimate the analyticity radius as follows (compare with Figure 5b): having fixed the order N (see step 3), at which the series is explicitly computed, and the increment h (see step 1), we interpolate the points before the well with a straight line. The critical value, say ε_L, is determined as the intersection of such line with the portion of the curve after the minimum is reached.
Figure 5 (see [5]) shows an implementation of such algorithm for the standard mapping (30) and the frequency $\omega = 2\pi 34/55$; the parameters has been set as $N = 800$ and $h = 0.001$.

6.3 Greene's method

Let $\mathcal{C}_{\varepsilon,\omega}$ denote the invariant curve with irrational frequency ω. We define the critical threshold at which $\mathcal{C}_{\varepsilon,\omega}$ breaks down as

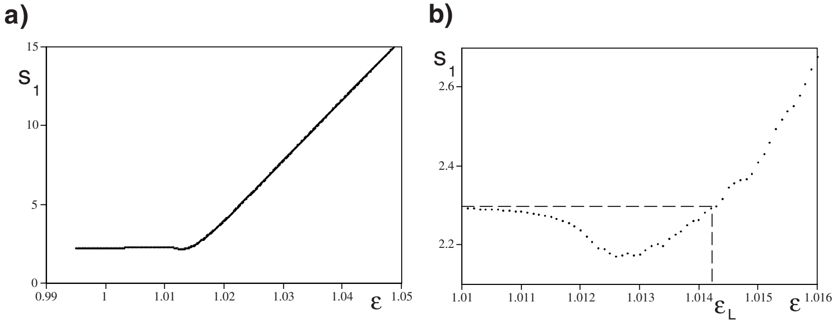


Fig. 5. (a) Plot of s_1 versus ε for $\omega = 2\pi 34/55$, $N = 800$, $h = 0.001$; (b) zoom of (a) for $1.01 \le \varepsilon \le 1.016$ and computation of ε_L (after [5]).

$$\varepsilon_c(\omega) = \sup\{\varepsilon' \ge 0 : \text{ for any } \varepsilon'' < \varepsilon', \text{ there exists an analytic invariant curve } \mathcal{C}_{\varepsilon'',\omega}\}.$$

The most widely accepted numerical technique to compute $\varepsilon_c(\omega)$ is Greene's method [13], which is based on the analysis of the stability of the periodic orbits approaching $\mathcal{C}_{\varepsilon,\omega}$. In order to investigate the stability character of a periodic orbit with frequency $2\pi p/q$ (for some p, $q \in \mathbb{Z}$), we look at the eigenvalues of the monodromy matrix

$$M = \prod_{i=1}^{q} \begin{pmatrix} 1 + \varepsilon\cos(\theta_i) & 1 \\ \varepsilon\cos(\theta_i) & 1 \end{pmatrix},$$

where $(\theta_1, ..., \theta_q)$ are successive points on the periodic orbit associated to the mapping (30). From the area–preserving property, one has that $\det M = 1$. Let T be the trace of M, whose eigenvalues are solutions of the equation: $\lambda^2 - T\lambda + 1 = 0$. Then, if $|T| < 2$, the eigenvalues of M are complex conjugates on the unit circle and the periodic orbit is stable. On the contrary, if $|T| > 2$ the eigenvalues are real and the periodic orbit is unstable.

To be concrete, let us fix a periodic orbit with frequency $2\pi p/q$, such that it is elliptic for small values of ε (as well as for $\varepsilon = 0$). As ε increases, the trace of the matrix M exceeds eventually 2 in modulus and the periodic orbit becomes unstable.

Figure 6a (see [5]) shows the quantity $\varrho_{Gr}(p_k/q_k)$, which corresponds to the value of ε marking the transition from stability to instability of the periodic orbit with frequency $2\pi p_k/q_k$. We selected the sequence of rational approximants to the golden ratio γ and we represented with a dotted line the estimated break–down value of $\mathcal{C}_{\varepsilon,\gamma}$. Figure 6$b$ provides a comparison between Greene's and Lyapunov's methods, by showing the plot of the relative error of the quantities $\varrho_c(p_k/q_k)$ and $\varrho_{Gr}(p_k/q_k)$.

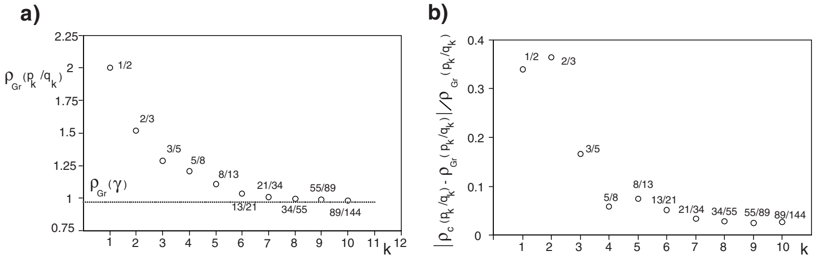


Fig. 6. (a) $\varrho_{Gr}(\frac{p_k}{q_k})$ versus the order k pertaining the rational approximants to the golden ratio; the dotted line represents Greene's threshold about equal to 0.971635. (b) The relative error associated to $\varrho_c(\frac{p_k}{q_k})$ and $\varrho_{Gr}(\frac{p_k}{q_k})$ for some rational approximants to the golden ratio (after [5]).

6.4 Results

In the framework of the standard map (30), we show some results on the behavior of the invariant curves with frequencies γ, ω_1, ω_2, where $\gamma = 2\pi[0; 1^\infty]$, $\omega_1 = 2\pi[0; 3, 12, 1^\infty]$, $\omega_2 = 2\pi[0; 2, 10, 1^\infty]$. We have implemented the three techniques presented in the previous sections, i.e. Padé, Greene and Lyapunov. Such methods depend on the choice of some parameters and of tolerance errors. In particular, Lyapunov's method depends on the order N of the truncation and on the increment h of the perturbing parameter ε. For a fixed N the agreement of the three methods is almost optimal taking a suitable value of the increment, typically $h = 0.001$.

Table 1 shows a comparison of the three methods for the golden ratio approximants of the standard map (30). The series have been computed up to $N = 800$ and the increment of Lyapunov's method has been set to $h = 10^{-3}$ (compare with [5]).

Table 2 (see [5]) provides the results for the invariant curve associated to the standard map (30) with frequency equal to ω_1; we underline the good agreement between Padé's and Lyapunov's methods, providing estimates of the analyticity radius. Due to the length of the calculations, it was possible to compute only the first few approximants while applying Padé's method (up to 38/117 in Table 2 and up to 85/178 in Table 3).

Similarly, we report in Table 3 the results for the invariant curve associated to the standard map (30) with frequency equal to ω_2 (compare with [5]).

In the cases of the mappings f_{12} and f_{13}, we found the results reported in Figure 7 (see [5]), where the abscissa k refers to the order of the approximant p_k/q_k of the golden ratio: for example, $k = 1$ corresponds to 2/3, while $k = 9$ corresponds to 89/144. With reference to Figure 7, the break–down threshold, as computed by Greene's method (*squares*), amounts to $\varepsilon_c(\gamma) = 1.2166$ for f_{12} (see Figure 7a) and to $\varepsilon(\gamma) = 0.7545$ for f_{13} (see Figure 7b). The Lyapunov's method has been applied with a series truncated at $N = 800$ and for $h = 0.01$

p/q	Greene	Padé	Lyap.:$h = 10^{-3}$
2/3	1.5176	2.0501	2.0584
3/5	1.2856	1.4873	1.4913
5/8	1.1485	1.2495	1.2561
8/13	1.0790	1.1440	1.1492
13/21	1.0353	1.0753	1.0766
21/34	1.0106	1.0366	1.0397
34/55	0.9953	1.0115	1.0142
55/89	0.9862	0.9974	0.9960
89/144	0.9832	0.9909	0.9897

Table 1. Comparison of Greene's, Padé's and Lyapunov's methods for the rational approximants to the golden ratio in the framework of the mapping (30) (after [5])

p/q	Greene	Padé	Lyap.:$h = 10^{-3}$
12/37	0.7486	0.5730	0.57513
13/40	0.7322	0.5447	0.54658
25/77	0.7232	0.5579	0.56063
38/117	0.7145	0.5539	0.55580
63/194	0.7134		0.55755
101/311	0.7085		0.55688
164/505	0.7073		0.55715
265/816	0.7066		0.55706
429/1321	0.7056		0.55705

Table 2. Same as Table 1 for the invariant curve with frequency ω_1 (after [5]).

p/q	Greene	Padé	Lyap.:$h = 10^{-3}$
10/21	0.7487	0.5395	0.54165
11/23	0.7198	0.5008	0.49557
21/44	0.7060	0.5158	0.51795
32/67	0.6919	0.5084	0.51058
53/111	0.6869	0.5116	0.51354
85/178	0.6842	0.5105	0.51239
138/289	0.6801		0.51281
223/467	0.6785		0.51265
361/756	0.6781		0.51271
584/1223	0.6771		0.51268

Table 3. Same as Table 1 for the invariant curve with frequency ω_2 (after [5])

(*crosses*) and $h = 0.001$ (*triangles*). Padé approximants of order [200/200] have been also computed (*diamonds*). Due to the length of the calculations, it was possible to compute only the first few approximants (up to 21/34) by applying Padé's method.

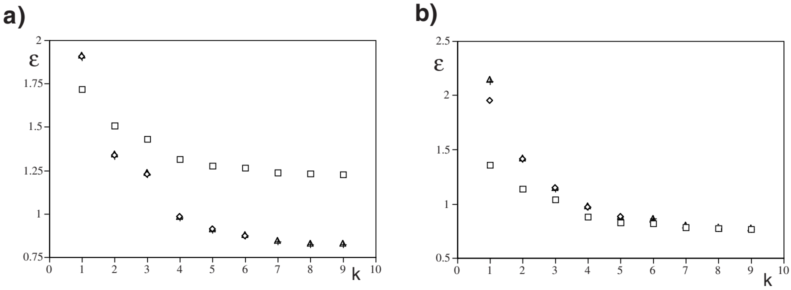


Fig. 7. Comparison of the results for the mappings f_{12} (a) and f_{13} (b) and for some rational approximants labeled by the index k; square: Greene's value, diamond: Padé's results, cross: Lyapunov's indicator s_2 (after [5]).

The results indicate that there is a good agreement between all methods as far as the approximants to the golden mean curve associated to (30) are analyzed. In such case the analyticity domain is close to a circle, so that the intersection with the positive real axis (providing an estimate of Greene's threshold) almost coincides with the analyticity radius, which is obtained implementing Padé's and Lyapunov's methods.

Such situation does not hold whenever the analyticity domain is not close to a circular shape. This happens, for example, if the invariant curves with rotation numbers $\omega_1 = 2\pi\ [0; 3, 12, 1^\infty]$ and $\omega_2 = 2\pi\ [0; 2, 10, 1^\infty]$ are considered; in these cases the two thresholds (break–down value and analyticity radius) are markedly different.

References

1. V.I. Arnold. *Proof of a Theorem by A.N. Kolmogorov on the invariance of quasi–periodic motions under small perturbations of the Hamiltonian*, Russ. Math. Surveys 18(9), 1963.
2. V.I. Arnold (editor). *Encyclopedia of Mathematical Sciences*, Dynamical Systems III, Springer–Verlag 3, 1988.
3. G.A. Baker Jr., P. Graves-Morris. *Padé Approximants*, Cambridge University Press, New York, 1996.
4. A. Celletti, L. Chierchia. *KAM Stability and Celestial Mechanics*, Preprint (2003), *http://www.mat.uniroma3.it/users/chierchia/PREPRINTS/SJV_03.pdf*
5. A. Celletti, C. Falcolini. *Singularities of periodic orbits near invariant curves*, Physica D 170(2):87, 2002.

6. L. Chierchia, C. Falcolini. *A direct proof of a theorem by Kolmogorov in Hamiltonian systems.* Ann. Scuola Norm. Sup. Pisa Cl. Sci. 4(21):541–593, 1994.
7. L. Chierchia, G. Gallavotti. *Smooth prime integrals for quasi-integrable Hamiltonian systems*, Nuovo Cimento B (11) 67(2):277–295, 1982.
8. B.V. Chirikov. *A universal instability of many dimensional oscillator systems*, Physics Reports 52:264–379, 1979.
9. L.H. Eliasson. *Absolutely convergent series expansions for quasi periodic motions*, Math. Phys. Electron. J. 2:33+ 1996.
10. C. Falcolini, R. de la Llave. *A rigorous partial justification of the Greene's criterion*, J. Stat. Phys., 67:609, 1992.
11. G. Gallavotti. *Twistless KAM tori*, Comm. Math. Phys. 164(1):145–156, 1994.
12. A. Giorgilli, U. Locatelli. *Kolmogorov theorem and classical perturbation theory*, Z. Angew. Math. Phys. 48(2):220–261, 1997.
13. J.M. Greene. *A method for determining a stochastic transition*, J. Math. Phys. 20, 1979.
14. M. Hénon. *Explorationes numérique du problème restreint IV: Masses egales, orbites non periodique*, Bullettin Astronomique 3(1)(fasc. 2):49–66, 1966.
15. A.N. Kolmogorov. *On the conservation of conditionally periodic motions under small perturbation of the Hamiltonian*, Dokl. Akad. Nauk. SSR 98:469, 1954.
16. R.S. MacKay. *Greene's residue criterion*, Nonlinearity, 5:161–187, 1992.
17. J. Moser. *On invariant curves of area-preserving mappings of an annulus*, Nach. Akad. Wiss. Göttingen, Math. Phys. Kl. II 1(1), 1962.
18. H. Poincaré. *Les Methodes Nouvelles de la Mechanique Celeste*, Gauthier Villars, Paris, 1892.
19. J. Pöschel. *Integrability of Hamiltonian systems on Cantor sets*, Comm. Pure Appl. Math. 35(5):653–696, 1982.
20. V. Szebehely. *Theory of orbits*, Academic Press, New York and London, 1967.

Resonances in Hyperbolic and Hamiltonian Systems

Viviane Baladi

CNRS UMR 7586,
Institut Mathématique de Jussieu,
75251 Paris, (France) baladi@math.jussieu.fr

Abstract

This text is a brief introduction to "Ruelle resonances," i.e. the spectra of transfer operators and their relation with poles and zeroes of dynamical zeta functions, and with poles of the Fourier transform of correlation functions.

1 Two elementary key examples – Basic concepts

1.1 Finite transition matrix and dynamical zeta function

Let A be a finite, say $N \times N$, complex matrix, with $N \geq 2$. Then, denoting by I the $N \times N$ identity matrix, we have (recall the Taylor series for $\log(1-t)$ and check, first in the diagonal case, that $\log\det B = \operatorname{Tr}\log B$ for any finite matrix B):

$$\det(I - zA) = \exp\left(- \sum_{m=1}^{\infty} \frac{z^m}{m} \operatorname{Tr} A^m \right) .$$

The left hand side of the above expression is a polynomial in z of degree at most N. Its zeroes are the inverses of the nonzero eigenvalues of A (the order of the zero coincides with the algebraic multiplicity of the eigenvalue). Let us show that the right hand side can be viewed as the inverse of a dynamical zeta function

$$\zeta_f(z) = \exp \sum_{m=1}^{\infty} \frac{z^m}{m} \# \operatorname{Fix} f^m ,$$

for a discrete-time dynamical system, i.e. the iterates $f^m = \overbrace{f \circ \cdots \circ f}^{m \text{ times}}$ of a transformation f, and their fixed points $\operatorname{Fix} f^m = \{x \mid f^m(x) = x\}$ (note that $\operatorname{Fix} f^m$ contains $\operatorname{Fix} f^k$ for each k which is a divisor of m).

J.-D. Fournier et al. (Eds.): Harm. Analysis and Ratio. Approx., LNCIS 327, pp. 263–274, 2006.

Indeed, if A is an $N \times N$ matrix with coefficients 0 and 1, it can be seen as a transition matrix, and one can associate to it a *subshift of finite type* on the alphabet $\mathcal{S} = \{1, \dots, N\}$. This subshift is the shift to the left $(\sigma_A(x))_i = x_{i+1}$ on the space of unilateral admissible sequences

$$\Sigma_A^+ = \{(x_i) \in \mathcal{S}^{\mathbb{N}} \mid A_{x_i x_{i+1}} = 1 \, , \forall i \in \mathbb{N}\} \, .$$

It is then easy to see that $\operatorname{Tr} A^m = \# \operatorname{Fix} \sigma_A^m$ (consider first $m = 1$ and note that x is fixed by σ_A if and only if $x = aaaaaaaaaaaaa \cdots$ with $A_{aa} = 1$). We thus have $\det(I - zA) = 1/\zeta_{\sigma_A}(z)$ (cf [32]).

In this example the matrix A can be seen as the (transposed) matrix of the restriction of the unweighted transfer operator

$$\mathcal{L}\varphi(x) = \sum_{\sigma(y)=x} \varphi(y) \, ,$$

to functions $\varphi : \Sigma_A^+ \to \mathbb{C}$ which only depend on the coefficient x_0 (this is an N-dimensional vector space).

To finish, note that since the coefficients of A are non-negative, the classical Perron-Frobenius theorem holds (cf e.g. [3]). For example, if A satisfies an aperiodicity assumption, i.e. if there exists m_0 such that all coefficients of A^{m_0} are (strictly) positive, then the matrix A (and thus the operator $\mathcal{L}$) has a simple eigenvalue $\lambda > 0$ equal to its spectral radius, with strictly positive right $Au = \lambda u$ and left $vA = \lambda v$ eigenvectors, while the rest of the spectrum is contained in a disc of radius strictly smaller than λ. In fact, λ is the exponential of the topological entropy of σ_A and the vectors u et v can be used to construct a σ_A-invariant measure which maximises entropy (cf e.g. [3]).

This situation gives a good introductory example, but it is far too simple: in general, the transfer operator must be considered as acting on an infinite-dimensional space on which it is often not trace-class. Nevertheless, one can still sometimes interpret the zeroes of a dynamical zeta function as the inverses of some subset of the eigenvalues of this operator.

1.2 Correlation functions and spectrum of the transfer operator

Let us consider a circle mapping f which is a small C^2 perturbation of the map $x \mapsto 2x$ (modulo 1). This transformation is not invertible (it has two branches) and it is locally uniformly expanding ("hyperbolicity"). Let us associate to f a weighted transfer operator

$$\mathcal{L}\varphi(x) = \sum_{f(y)=x} \frac{\varphi(y)}{|f'(y)|} \, ,$$

which acts boundedly (but not compactly) on each of the infinite-dimensional Banach spaces $L^1(Leb)$, $C^0(S^1)$, or $C^1(S^1)$. Our choice $1/|f'|$ for the weight, i.e. the jacobian of the inverse branch, implies that the dual of $\mathcal{L}$ acting (e.g.) on Radon measures preserves Lebesgue measure: $\int \mathcal{L}(\varphi)\mathrm{d}x = \int \varphi\,\mathrm{d}x$.

Recall that if $\mathcal{M}$ is a bounded operator on a Banach space $\mathcal{B}$, the *essential spectral radius* of $\mathcal{M}$ is the smallest $\varrho \geq 0$ so that the spectrum of $\mathcal{M}$ outside of the disc of radius ϱ contains only isolated eigenvalues of finite multiplicity (cf [13, 3]).

In this situation, we can prove *quasi-compactness*: the spectral radius of $\mathcal{L}$ on $C^1(S^1)$ is 1 while its essential spectral radius ϱ_{ess} is < 1 (cf. e.g. [3]). (Note that the spectrum on $L^1(Leb)$ or $C^0(S^1)$ is too "big": on these two Banach spaces, each point of the open unit disc is an eigenvalue of infinite multiplicity [41].) In fact, for the operator acting on $C^1(S^1)$, we even have a Perron-Frobenius-type picture: 1 is a simple eigenvalue for a positive eigenvector φ_0 (up to normalisation, one can assume that the integral of φ_0 is 1), while the rest of the spectrum is contained in a disc of radius τ with $\varrho_{\mathrm{ess}} \leq \tau < 1$. There is thus a *spectral gap.* The eigenvalues in the annulus $\varrho_{\mathrm{ess}} < |z| \leq \tau$, if there are any, are called *resonances.* To motivate this terminology, let us describe ergodic-theoretical consequences of these spectral properties. Before this, note that one can show that the dynamical zeta function

$$\zeta_{1/|f'|} = \exp \sum_{m=1}^{\infty} \frac{z^m}{m} \sum_{x \in \mathrm{Fix}\, f^m} |(f^m)'(x)|^{-1}$$

is meromorphic in the disc of radius $1/\varrho_{\mathrm{ess}}$, where its poles are the inverses of the eigenvalues of $\mathcal{L}$ acting on $C^1(S^1)$, i.e. the resonances (together with the simple pole at 1).

Let us first observe that the absolutely continuous probability measure μ_0 with density φ_0 (with respect to Lebesgue) is f-invariant: if $\varphi \in L^1(Leb)$

$$\int \varphi \circ f \cdot \varphi_0 \,\mathrm{d}x = \int \mathcal{L}((\varphi \circ f) \cdot \varphi_0)\,\mathrm{d}x = \int \varphi \mathcal{L}(\varphi_0)\,\mathrm{d}x = \int \varphi\varphi_0\,\mathrm{d}x\,.$$

One can show that μ_0 is ergodic, therefore the Birkhoff ergodic theorem says that for all φ in $L^1(Leb)$ and μ_0-almost every x (i.e., Lebesgue almost every x!), the temporal averages $(1/m)\sum_{k=0}^{m-1} \varphi(f^k(x))$ converge to the spatial average $\int \varphi\,\mathrm{d}\mu_0$.

We shall next see that this measure μ_0 is exponentially mixing for test functions φ_1, φ_2 in $C^1(S^1)$. Since the spectral projector corresponding to the eigenvalue 1 is $\varphi \mapsto \varphi_0 \cdot \int \varphi\,\mathrm{d}x$, we have the spectral decomposition

$$\mathcal{L}\varphi = \varphi_0 \int \varphi\,\mathrm{d}x + \mathcal{P}\mathcal{L}\varphi\,,$$

with $\mathcal{P}$ the spectral projector associated to the complement of 1 in the spectrum. This projector satisfies $\|\mathcal{P}\mathcal{L}^m\| \leq C\tilde{\tau}^m$ for any $\tau < \tilde{\tau} < 1$ and for the operator-norm acting on C^1. Therefore,

$$\begin{aligned}\int \varphi_1 \circ f^m \cdot \varphi_2 \cdot \varphi_0 \,\mathrm{d}x = \int \mathcal{L}^m(\varphi_1 \circ f^m \cdot \varphi_2 \cdot \varphi_0)\,\mathrm{d}x &= \int \varphi_1 \mathcal{L}^m(\varphi_2 \cdot \varphi_0)\,\mathrm{d}x \\ &= \int \varphi_1 \left[\varphi_0 \Big(\int \varphi_2 \cdot \varphi_0 \,\mathrm{d}x\Big) + \mathcal{P}\mathcal{L}^m(\varphi_2 \cdot \varphi_0)\right]\,\mathrm{d}x \\ &= \int \varphi_1 \,\mathrm{d}\mu_0 \cdot \int \varphi_2 \,\mathrm{d}\mu_0 + \int \varphi_1 \mathcal{P}\mathcal{L}^m(\varphi_2 \cdot \varphi_0)\,\mathrm{d}x\,,\end{aligned}$$

and since

$$\begin{aligned}\left|\int \varphi_1 \mathcal{P}\mathcal{L}^m(\varphi_2 \cdot \varphi_0)\mathrm{d}x\right| &\le \left(\int |\varphi_1|\mathrm{d}x\right) \cdot \|\mathcal{P}\mathcal{L}^m(\varphi_2\varphi_0)\|_{C^1} \\ &\le C\left(\int |\varphi_1|\mathrm{d}x\right)\|\varphi_2\|_{C^1} \cdot \tilde{\tau}^m,\end{aligned}$$

we obtain the claimed exponential decay.

Finally, note that for φ_1 and φ_2 in $C^1(S^1)$, the Fourier transform

$$\widehat{C}_{\varphi_1\varphi_2}(\omega) = \sum_{m \in \mathbb{Z}} \mathrm{e}^{\mathrm{i}m\omega} C_{\varphi_1,\varphi_2}(m)\,,$$

of their correlation function

$$C_{\varphi_1,\varphi_2}(m) = \begin{cases} \int \varphi_1 \circ f^m \cdot \varphi_2 \,\mathrm{d}\mu_0 - \int \varphi_1 \,\mathrm{d}\mu_0 \int \varphi_2 \,\mathrm{d}\mu_0 & m \ge 0\,, \\ C_{\varphi_2,\varphi_1}(-m) & m \le 0\,, \end{cases}$$

is meromorphic in the strip $|\operatorname{Im} \omega| \le \log(\tau^{-1})$ where its poles are those ω such that $\mathrm{e}^{\pm \mathrm{i}\omega}$ is a *resonance*:

$$\begin{aligned}&\sum_{m \in \mathbb{N}} \mathrm{e}^{\mathrm{i}m\omega}\Big(\int \varphi_1 \circ f^m \varphi_2 \,\mathrm{d}\mu_0 - \int \varphi_1 \,\mathrm{d}\mu_0 \int \varphi_2 \,\mathrm{d}\mu_0\Big) \\ &\qquad = \sum_{m \in \mathbb{N}} \int \varphi_1 \mathcal{P}(\mathrm{e}^{\mathrm{i}\omega}\mathcal{L})^m(\varphi_2\varphi_0)\,\mathrm{d}x = \int \varphi_1 \sum_{m \in \mathbb{N}} \mathcal{P}(\mathrm{e}^{\mathrm{i}\omega}\mathcal{L})^m(\varphi_2\varphi_0)\,\mathrm{d}x \\ &\qquad = \int \varphi_1 \Big((1 - \mathrm{e}^{\mathrm{i}\omega}\mathcal{P}\mathcal{L})^{-1}(\varphi_2\varphi_0)\Big)\,\mathrm{d}x\,.\end{aligned}$$

1.3 Basic concepts

Let $f : M \to M$ be a map and let $g : M \to \mathbb{C}$ be a weight. We assume (these assumptions can in fact be weakened) that $\operatorname{Fix} f^m$ is a finite set for each fixed $m \ge 1$ and that the set $\{y \mid f(y) = x\}$ is finite for each x.

Definition 1. *(Ruelle) transfer operator – Resonances*

The transfer operator is the linear operator

$$\mathcal{L}_{f,g}\varphi(x) = \sum_{y: f(y)=x} g(y)\varphi(y)\,,$$

acting on an appropriate Banach (sometimes Hilbert) space of functions or distributions φ on M. In general $\mathcal{L}$ is bounded but not compact. If the essential spectral radius ϱ_{ess} of $\mathcal{L}$ is strictly smaller than its spectral radius, one says that $\mathcal{L}$ is *quasi-compact*. The spectrum of $\mathcal{L}$ outside of the disc of radius ϱ_{ess} is called the set of resonances of (f, g).

Definition 2. *Weighted zeta function*

A weighted zeta function is a power series

$$\zeta_{f,g}(z) = \exp \sum_{m=1}^{\infty} \frac{z^m}{m} \sum_{x \in \operatorname{Fix} f^m} \prod_{k=0}^{m-1} g(f^k(x))\,.$$

One can sometimes show that it is meromorphic in a disc where its poles are in bijection with the resonances.

Definition 3. *Dynamical (Ruelle-Fedholm) determinant*

Assume moreover that f is (at least) C^1 and set $\operatorname{Fix}_h f^m = \{x \in \operatorname{Fix} f^m \mid \det(I - Df^{-m}(x)) \neq 0\}$. The dynamical determinant is the power series

$$d_{f,g}(z) = \exp\left(- \sum_{m=1}^{\infty} \frac{z^m}{m} \sum_{x \in \operatorname{Fix}_h f^m} \frac{\prod_{k=0}^{m-1} g(f^k(x))}{\det(I - Df^{-m}(x))} \right).$$

One can sometimes show that this series is holomorphic in a disc which is larger than the disc associated to $\zeta_{f,g}$, and that in this larger disc, its zeroes are in bijection with the resonances.

Definition 4. *Correlation function*

Let μ be an f-invariant probability measure (for example, a measure absolutely continuous with respect to Lebesgue or an equilibrium state for $\log |g|$). The correlation function for (f, μ) and a class of functions $\varphi : M \to \mathbb{C}$, is the function $C_{\varphi_1,\varphi_2} : \mathbb{N} \to \mathbb{C}$ defined for φ_1, φ_2 in this class by

$$C_{\varphi_1,\varphi_2}(m) = \int \varphi_1 \circ f^m \cdot \varphi_2 \, \mathrm{d}\mu - \int \varphi_1 \, \mathrm{d}\mu \int \varphi_2 \, \mathrm{d}\mu\,.$$

Analogous concepts exist for continuous-time dynamics (flows, in particular geodesic flows in not necessarily constant negative curvature - an example of an intersection between hyperbolic and hamiltonian dynamics). The corresponding zeta function $\zeta(s)$ is then often holomorphic in a half-plane $\mathrm{Re}(s) > s_0$ and it admits a meromorphic extension in a larger half-plane. We refer to the various surveys mentioned in the bibliography, which contain references to the fundamental articles of Smale, Artin-Mazur, Ruelle, etc.

2 Theorems of Ruelle, Keller, Pollicott, Dolgopyat...

The authors mentioned in the title are by far not the only ones to have made important contributions to the theory of dynamical zeta functions and transfer operators. One should also mention (see the bibliography) Fried, Mayer, Hofbauer, Haydn, Sharp, Rugh, Kitaev, Liverani, Buzzi, and many others (in particular Cvitanović for a more physical approach). Let us discuss a selection of themes.

1. *Ruelle* [29] observed that the transfer operator associated to a discrete-time dynamical system given by an *expanding and analytic* map, together with an analytic weight g is nuclear ("trace-class") on an appropriate Banach space of holomorphic functions. In this case, the dynamical zeta function is an alternated product of Fredholm-Grothendieck determinants and the dynamical determinant is a determinant. (In the case when contraction and expansion coexist, an assumption of regularity of foliations was needed until the work of Rugh and Fried [30, 24].) This fact is the key to proving that the dynamical zeta function and the Fourier transform of the correlation function admit a meromorphic extension to the whole complex plane, in some cases. Let us also mention a recent "Hilbert space" version of this theory by Guillopé, Lin et Zworski [84], who are able to estimate the density of resonances (in the "classical" sense, which coincides here with the sense of Ruelle) of certain Schottky groups (another example of application to hamiltonian dynamics).

2. *Symbolic dynamics* allows us to model a hyperbolic dynamical system by a subshift of finite type, via Markov partitions (Sinai, Ratner, cf. [7]). The unstable Jacobian is Lipschitz (for a suitable metric) in symbolic coordinates. So we are led to study transfer operators $\mathcal{L}_{\sigma_A,g}$ associated to the unilateral subshift σ_A and a *Lipschitz (or Hölder)* weight g: they are bounded but not compact, on the Banach space of Lipschitz (or Hölder) functions. *Ruelle* [7, 4] proved the first Perron-Frobenius-type theorem in this kind of setting: there is a spectral gap, and thus exponential decay of correlations in good cases. By combining the results of Ruelle [19], Pollicott [37] and Haydn [33], we obtain a meromorphic extension of the Fourier transform of the correlation function to a strip, which in fact is the largest possible that can be obtained in this setting.

3. Several years after the pioneering work of Lasota and Yorke on existence of absolutely continuous invariant measures, Hofbauer and Keller [50, 51, 52] obtained quasi-compactness of the transfer operator associated to piecewise expanding (not necessarily Markov) interval maps acting on functions of bounded variation. This operator is not compact, but the dynamical zeta function $\zeta_{f,g}$ has a nontrivial meromorphic extension to a disc where its poles are in bijection with the resonances (eigenvalues) of the operator [46]. The higher-dimensional case is much more recent [49, 54] and only partial results have been obtained [BuKe]. There are stronger results for (Markov) *differentiable* locally *expanding maps* [41, 43] for which one may also study the dynamical determinant $d_{f,g}(z)$ [45, 42] (see also [44, 39] in the hyperbolic case). Let us also mention the recent results of Collet et Eckmann [40] who show that in general the "essential" rate of decay of correlations is slower than the smallest Lyapunov exponent, contrary to a widespread misconception.

4. The case of *continuous-time dynamical systems* is much more delicate. A meromorphic extension of the zeta function of a hyerbolic flow to a half-plane larger than the half-plane of convergence was obtained in the eighties by Ruelle, Pollicott [76, 74]. Parry–Pollicott [73] obtained a striking analogue of the prime number theorem for hyperbolic flows. This result was followed by many other "counting" results. Ikawa [86] proved a modified Lax-Phillips conjecture (see also [98]). However, in order to get exponential decay of correlations, a vertical strip without poles is required, and this is not always possible: Ruelle [75] constructed examples of uniformly hyperbolic flows which do not mix exponentially fast. Only recently could Dolgopyat [70, 71] prove (among other things) exponential decay of correlations for certain Anosov flows, by using oscillatory integrals. This result has consequences for billiards [97, 94, 91], yet another hyperbolic/hamiltonian system. Liverani very recently introduced a new method to prove exponential decay of correlations instead of representing the flow as a (local) suspension of hyperbolic diffeomorphisms under return times (using the Poincaré map associated to Makov sections), he studies directly the semi-group of operators associated to the flow [14, 72].

Despite its length, the bibliography is not complete. We hope that the decomposition in items, although rather arbitrary, will make it more useful. We do not mention at all the vast existing literature on sub-exponential decay of correlations.

References

Surveys and books

1. V. Baladi. *Dynamical zeta functions*, Proceedings of the NATO ASI "Real and Complex Dynamical Systems" (1993), B. Branner et P. Hjorth, Kluwer Academic Publishers, Dordrecht, pages 1–26, 1995. See `www.math.jussieu.fr/~baladi/zeta.ps`.
2. V. Baladi. *Periodic orbits and dynamical spectra*, Ergodic Theory Dynamical Systems, 18:255–292, 1998. See `www.math.jussieu.fr/~baladi/etds.ps`.
3. V. Baladi. *Positive Transfer Operators and Decay of Correlations*, World Scientific, Singapore, 2000. Erratum available on `www.math.jussieu.fr/~baladi/erratum.ps`.
4. V. Baladi. *The Magnet and the Butterfly: Thermodynamic formalism and the ergodic theory of chaotic dynamics*, Développement des mathématiques au cours de la seconde moitié du XXe siècle, Birkhäuser, Basel, 2000. Available on `www.math.jussieu.fr/~baladi/thermo.ps`

5. V. Baladi. *Spectrum and Statistical Properties of Chaotic Dynamics*, Proceedings Third European Congress of Mathematics Barcelona 2000, pages 203–224, Birkhäuser, 2001. Available on `www.math.jussieu.fr/∼baladi/barbal.ps`
6. V. Baladi. *Decay of correlations*, AMS Summer Institute on Smooth ergodic theory and applications, (Seattle, 1999), Proc. Symposia in Pure Math. AMS, 69:297–325, 2001. See `www.math.jussieu.fr/∼baladi/seattle.ps`
7. R. Bowen. *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, Springer (Lecture Notes in Math., Vol. 470), Berlin, 1975.
8. P. Cvitanović, R. Artuso, R. Mainieri, G. Tanner, G. Vattay. *Chaos: Classical and Quantum*, Niels Bohr Institute, Copenhagen, 2005.
9. D. Dolgopyat, M. Pollicott. *Addendum to: "Periodic orbits and dynamical spectra"*, Ergodic Theory Dynam. Systems, 18:293–301, 1998.
10. N. Dunford, J.T. Schwartz. *Linear Operators, Part I, General Theory*, Wiley-Interscience (Wiley Classics Library), New York, 1988.
11. J.-P. Eckmann. *Resonances in dynamical systems*, IXth International Congress on Mathematical Physics, (Swansea, 1988), Hilger, Bristol, pages 192–207, 1989.
12. I. Gohberg, S. Goldberg, N. Krupnik. *Traces and Determinants of Linear Operators*, Birkhäuser, Basel, 2000.
13. T. Kato. *Perturbation Theory for Linear Operators*, Springer-Verlag, Berlin, 1984. Second Corrected Printing of the Second Edition.
14. C. Liverani. *Invariant measures and their properties. A functional analytic point of view*, Dynamical systems. Part II, pages 185–237, Pubbl. Cent. Ric. Mat. Ennio Giorgi, Scuola Norm. Sup., Pisa, 2003.
15. D.H. Mayer. *The Ruelle-Araki transfer operator in classical statistical mechanics*, Lecture Notes in Physics, Vol 123, Springer-Verlag, Berlin-New York, 1980.
16. D. Mayer. *Continued fractions and related transformations*, Ergodic Theory, Symbolic Dynamics and Hyperbolic Spaces, T. Bedford et al., Oxford University Press, 1991.
17. W. Parry, M. Pollicott. *Zeta functions and the periodic orbit structure of hyperbolic dynamics*, Société Mathématique de France (Astérisque, vol. 187-188), Paris, 1990.
18. D. Ruelle. *Resonances of chaotic dynamical systems*, Phys. Rev. Lett, 56:405–407, 1986.
19. D. Ruelle. *Dynamical Zeta Functions for Piecewise Monotone Maps of the Interval*, CRM Monograph Series, Vol. 4, Amer. Math. Soc., Providence, NJ, 1994.
20. D. Ruelle. *Dynamical zeta functions and transfer operators*, Notices Amer. Math. Soc, 49:887–895, 2002.
21. M. Zinsmeister. *Formalisme thermodynamique et systèmes dynamiques holomorphes*, Panoramas et Synthèses, 4. Société Mathématique de France, Paris, 1996.

Analytical framework

22. V. Baladi, H.H. Rugh. *Floquet spectrum of weakly coupled map lattices*, Comm. Math. Phys, 220:561–582, 2001.
23. D. Fried. *The zeta functions of Ruelle and Selberg I*, Ann. Sci. École Norm. Sup. (4), 19:491–517, 1986.

24. D. Fried. *Meromorphic zeta functions for analytic flows*, Comm. Math. Phys., 174:161–190, 1995.
25. A. Grothendieck. *Produits tensoriels topologiques et espaces nucléaires*, (Mem. Amer. Math. Soc. 16), Amer. Math. Soc., 1955.
26. A. Grothendieck. *La théorie de Fredholm*, Bull. Soc. Math. France, 84:319–384, 1956.
27. G. Levin, M. Sodin, P. Yuditskii,. *A Ruelle operator for a real Julia set* Comm. Math. Phys., 141:119–131, 1991.
28. D. Mayer. *On the thermodynamic formalism for the Gauss map*, Comm. Math. Phys., 130:311–333, 1990.
29. D. Ruelle. *Zeta functions for expanding maps and Anosov flows*, Inv. Math., 34:231–242, 1976.
30. H.H. Rugh. *Generalized Fredholm determinants and Selberg zeta functions for Axiom A dynamical systems*, Ergodic Theory Dynam. Systems, 16:805–819, 1996.
31. H.H. Rugh. *Intermittency and regularized Fredholm determinants*, Invent. Math., pages 1–24, 1999.

Symbolic dynamics framework (Hölder-Lipschitz)

32. R. Bowen, O.E. Lanford III. *Zeta functions of restrictions of the shift transformation*, Proc. Sympos. Pure Math., 14:43–50, 1970.
33. N.T.A. Haydn. *Gibbs functionals on subshifts*, Comm. Math. Phys., 134:217–236, 1990.
34. N.T.A. Haydn. *Meromorphic extension of the zeta function for Axiom A flows*, Ergodic Theory Dynamical Systems, 10:347–360, 1990.
35. A. Manning. *Axiom A diffeomorphisms have rational zeta functions*, Bull. London Math. Soc., 3:215–220, 1971.
36. M. Pollicott. *A complex Ruelle operator theorem and two counter examples*, Ergodic Theory Dynamical Systems, 4:135–146, 1984.
37. M. Pollicott. *Meromorphic extensions of generalised zeta functions*, Invent. Math., 85:147– 164, 1986.
38. D. Ruelle. *One-dimensional Gibbs states and Axiom A diffeomorphisms*, J. Differential Geom., 25:117–137, 1987.

Differentiable framework

39. M. Blank, G. Keller, C. Liverani. *Ruelle-Perron-Frobenius spectrum for Anosov maps*, Nonlinearity, 15:1905–1973, 2002.
40. P. Collet, J.-P. Eckmann. *Liapunov Multipliers and Decay of Correlations in Dynamical Systems*, J. Statist. Phys., 115:217–254, 2004.
41. P. Collet, S. Isola. *On the essential spectrum of the transfer operator for expanding Markov maps*, Comm. Math. Phys., 139:551–557, 1991.
42. D. Fried. *The flat-trace asymptotics of a uniform system of contractions*, Ergodic Theory Dynamical Systems, 15:1061–1073, 1995.
43. V.M. Gundlach, Y. Latushkin. *A sharp formula for the essential spectral radius of the Ruelle transfer operator on smooth and Holder spaces*, Ergodic Theory Dynam. Systems, 23:175–191, 2003.
44. A. Kitaev. *Fredholm determinants for hyperbolic diffeomorphisms of finite smoothness*, Nonlinearity, 12:141–179, 1999. See also Corrigendum, 1717–1719.
45. D. Ruelle. *An extension of the theory of Fredholm determinants*, Inst. Hautes Etudes Sci. Publ. Math., 72:175–193, 1991.

Non Markov settings (BV, logistic maps, Hénon...)

46. V. Baladi, G. Keller. *Zeta functions and transfer operators for piecewise monotone transformations*, Comm. Math. Phys., 127:459–479, 1990.
47. M. Benedicks, L.-S. Young. *Markov extensions and decay of correlations for certain Hénon maps*, In Géométrie complexe et systèmes dynamiques (Orsay, 1995), Astérisque (261):13–56, 2000.
48. J. Buzzi, G. Keller. *Zeta functions and transfer operators for multidimensional piecewise affine and expanding maps*, Ergodic Theory Dynam. Systems, 21:689–716, 2001.
49. J. Buzzi, V. Maume-Deschamps. *Decay of correlations for piecewise invertible maps in higher dimensions*, Israel J. Math, 131:203–220, 2002.
50. F. Hofbauer, G. Keller. *Ergodic properties of invariant measures for piecewise monotonic transformations*, Math. Z., 180:119–140, 1982.
51. F. Hofbauer, G. Keller. *Zeta-functions and transfer-operators for piecewise linear transformations*, J. reine angew. Math., 352:100–113, 1984.
52. G. Keller. *On the rate of convergence to equilibrium in one-dimensional systems*, Comm. Math. Phys, 96:181–193, 1984.
53. G. Keller, T. Nowicki. *Spectral theory, zeta functions and the distribution of periodic points for Collet–Eckmann maps*, Comm. Math. Phys., 149:31–69, 1992.
54. B. Saussol. *Absolutely continuous invariant measures for multidimensional expanding maps*, Israel J. Math., 116:223–248, 2000.
55. L.-S. Young. *Statistical properties of dynamical systems with some hyperbolicity*, Ann. of Math. (2), 147:585–650, 1998.

Birkhoff cone methods

56. P. Ferrero, B. Schmitt. *Produits aléatoires d'opérateurs matrices de transfert*, Probab. Theory Related Fields, 79:227–248, 1988.
57. C. Liverani. *Decay of correlations*, Ann. of Math. (2), 142:239–301, 1995.
58. C. Liverani. *Decay of correlations for piecewise expanding maps*, J. Stat. Phys., 78:1111–1129, 1995.

Milnor-Thurston "kneading" methods

59. M. Baillif. *Kneading operators, sharp determinants, and weighted Lefschetz zeta functions in higher dimensions*, Duke Math. J., 124:145–175, 2004.
60. M. Baillif, V. Baladi. *Kneading determinants and spectrum in higher dimensions: the isotropic case*. Preprint (2003).
61. V. Baladi, A. Kitaev, D. Ruelle, S. Semmes. *Sharp determinants and kneading operators for holomorphic maps*, Proc. Steklov Inst. Math., 216:186–228, 1997.
62. V. Baladi, D. Ruelle. *Sharp determinants*, Invent. Math., 123:553–574, 1996.
63. S. Gouëzel. *Spectre de l'opérateur de transfert en dimension 1*, Manuscripta Math, 106:365–403, 2001.
64. J. Milnor, W. Thurston. *Iterated maps of the interval*, Dynamical Systems (Maryland 1986-87), Lecture Notes in Math. Vol. 1342, J.C. Alexander, Springer-Verlag, Berlin Heidelberg New York, 1988.
65. D. Ruelle. *Sharp zeta functions for smooth interval maps*, Proceedings Conference on Dynamical Systems (Montevideo, 1995), pages 188–206, Pitman Res. Notes Math. Ser. 362, Longman, Harlow, 1996.

Random systems and spectral stability

66. V. Baladi, M. Viana. *Strong stochastic stability and rate of mixing for unimodal maps*, Annales scient. Ecole normale sup. (4), 29:483–517, 1996.
67. V. Baladi, L.-S. Young. *On the spectra of randomly perturbed expanding maps*, Comm. Math. Phys., 156:355–385, 1993. Erratum, Comm. Math. Phys., **166**, 219–220 (1994).
68. J. Buzzi. *Some remarks on random zeta functions*, Ergodic Theory Dynam. Systems, 22:1031–1040, 2002.
69. G. Keller, C. Liverani. *Stability of the spectrum for transfer operators*, Annali Scuola Normale Sup. Pisa (4), XXVIII:141–152, 1999.

Flows

70. D. Dolgopyat. *On decay of correlations in Anosov flows*, Ann of Math., 147:357–390, 1998.
71. D. Dolgopyat. *Prevalence of rapid mixing for hyperbolic flows*, Ergodic Theory Dynam. Systems, 18:1097–1114, 1998.
72. C. Liverani. *On contact Anosov flows*, Preprint (2002). To apear Ann. of Math.
73. W. Parry, M. Pollicott. *An analogue of the prime number theorem for closed orbits of Axiom A flows*, Ann. of Math. (2), 118:573–591, 1983.
74. M. Pollicott. *On the rate of mixing of Axiom A flows*, Invent. Math., 81:413–426, 1985.
75. D. Ruelle. *Flots qui ne mélangent pas exponentiellement*, C. R. Acad. Sci. Paris Sér. I Math., 296:191–193, 1983.
76. D. Ruelle. *Resonances for Axiom A flows*, J. Differential Geom., 25:99–116, 1987.

Numerics and aplications

77. R. Artuso, E. Aurell, P. Cvitanović. *Recycling of strange sets: I. Cycle expansions, II. Applications*, Nonlinearity, 3':325–359+361–386, 1990.
78. M. Babillot, M. Peigné, *Homologie des géodésiques fermées sur des variétés hyperboliques avec bouts cuspidaux*, Ann. Sci. École Norm. Sup. (4), vol 33, pages 81–120, 2000.
79. V. Baladi, J.-P. Eckmann, D. Ruelle. *Resonances for intermittent systems*, Nonlinearity, 2:119–135, 1989.
80. C.-H. Chang, D.H. Mayer. *Eigenfunctions of the transfer operators and the period functions for modular groups*, Dynamical, spectral, and arithmetic zeta functions (San Antonio, TX, 1999), Contemp. Math., 290:1–40, Amer. Math. Soc., Providence, RI, 2001.
81. F. Christiansen, P. Cvitanović, H.H. Rugh. *The spectrum of the period-doubling operator in terms of cycles*, J. Phys. A, 23:L713–L717, 1990.
82. P. Cvitanović, P.E. Rosenqvist, G. Vattay, H.H. Rugh. *A Fredholm determinant for semiclassical quantization*, Chaos, 3:619–636, 1993.
83. M. Dellnitz, G. Froyland, S. Sertl. *On the isolated spectrum of the Perron-Frobenius operator*, Nonlinearity, 13:1171–1188, 2000.
84. L. Guillopé, K.K. Lin, and M. Zworski. *The Selberg zeta function for convex co-compact Schottky groups*, Comm. Math. Phys., 245:149–176, 2004.

85. J. Hilgert, D. Mayer. *Transfer operators and dynamical zeta functions for a class of lattice spin models*, Comm. Math. Phys, 232:19–58, 2002.
86. M. Ikawa. *Singular perturbation of symbolic flows and poles of the zeta functions*, Osaka J. Math., 27:281–300, 1990.
87. S. Isola. *Resonances in chaotic dynamics*, Comm. Math. Phys, 116:343–352, 1988.
88. O. Jenkinson, M. Pollicott. *Calculating Hausdorff dimensions of Julia sets and Kleinian limit sets*, Amer. J. Math., 124:495–545, 2002.
89. D. Mayer. *The thermodynamic formalism approach to Selberg's zeta function for $PSL(2,\mathbb{Z})$*, Bull. Amer. Math. Soc., 25:55–60, 1991.
90. T. Morita, *Markov systems and transfer operators associated with cofinite Fuchsian groups*, Ergodic Theory Dynam. Systems, 17:1147–1181, 1997.
91. F. Naud. *Analytic continuation of a dynamical zeta function under a Diophantine condition*, Nonlinearity, 14:995–1009, 2001.
92. F. Naud, *Expanding maps on Cantor sets, analytic continuation of zeta functions with applications to convex co-compact surfaces.* Preprint, 2003.
93. S.J. Patterson, P.A. Perry. *The divisor of Selberg's zeta function for Kleinian groups*, Duke Math. J., 106:321–390, 2001.
94. V. Petkov. *Analytic singularities of the dynamical zeta function*, Nonlinearity, 12:1663–1681, 1999.
95. M. Pollicott and A.C. Rocha. *A remarkable formula for the determinant of the Laplacian*, Invent. Math, 130:399–414, 1997.
96. M. Pollicott, R. Sharp. *Exponential error terms for growth functions on negatively curved surfaces*, Amer. J. Math., 120:1019–1042, 1998.
97. L. Stoyanov. *Spectrum of the Ruelle operator and exponential decay of correlations for open billiard flows*, Amer. J. Math., 123:715–759, 2001.
98. L. Stoyanov. *Scattering resonances for several small convex bodies and the Lax-Phillips conjecture*, Preprint, 2003.

Signal Processing Methods Related to Models of Turbulence

Pierre Borgnat

Laboratoire de Physique (UMR-CNRS 5672)
ÉNS Lyon 46 allée d'Italie
69364 Lyon Cedex 07 (France)
Pierre.Borgnat@ens-lyon.fr

1 An overview of the main properties of Turbulence

Turbulence deals with the complex motions in fluid at high velocity and/or involving a large range of length-scales. Understanding turbulence is challenging and involves many questions from modeling this complexity to measuring it. In this text, we aim at describing some tools of signal processing that have been used to study signals measured in turbulence experiments. Before that, another objective is the survey of some properties relevant for turbulent flows (experiments and/or models): scaling laws, self-similarity, multifractality and non-stationarity, that will explain why those techniques are useful.

1.1 Qualitative Analysis of Turbulence

Introduction.

Turbulence is first a problem of mechanics applied to fluids. The fundamental relation of dynamics may be written for a fluid element and this rules its evolution. The velocity field $\boldsymbol{u}(\boldsymbol{r}(0);t)$, giving the velocity at time t of a fluid element that is in $\boldsymbol{r}(0)$ at initial time, is called the Lagrangian velocity in the fluid. This follows the point of view of Lagrange: one tracks the behavior of each part of the fluid along its trajectory $\boldsymbol{r}(t)$. The velocity is driven by a balance between the inertial effects and the force s in the fluid: friction, pressure, gravity. If the fluid is incompressible, the pressure derives from the whole velocity field and the resulting problem is not local. Added to that, it is experimentally hard to track the movement of one fluid element: nothing distinguishes it from all the fluid. Nevertheless, we will see later how some measurements of Lagrangian velocity were made possible.

But usually, instead of this Lagrangian velocity, the problem is studied through the point of view of Euler: the velocity $\boldsymbol{v}(\boldsymbol{r},t)$ at the fixed position $\boldsymbol{r}$ and at time t characterizes all the motions in the fluid. It is called

J.-D. Fournier et al. (Eds.): Harm. Analysis and Ratio. Approx., LNCIS 327, pp. 277–301, 2006.

the Eulerian velocity. Both velocities are related via the change of variable $\boldsymbol{u}(\boldsymbol{r}(0);t) = \boldsymbol{v}(\boldsymbol{r}(t);t)$. The partial derivative equation for this Eulerian velocity is called the Navier-Stokes (NS) equation and reads:

$$\underbrace{\partial_t \boldsymbol{v}}_{\text{local derivative}} + \underbrace{(\boldsymbol{v}\cdot\nabla)\boldsymbol{v}}_{\text{convective derivative}} = \underbrace{-(1/\varrho)\nabla p}_{\text{pressure}} + \underbrace{\nu\Delta\boldsymbol{v}}_{\text{viscous friction}} + \sum \boldsymbol{f}_v . \quad (1)$$

Here ϱ is the density of the fluid. The term $\boldsymbol{f}_v$ stands for volumic forces in the fluid (electric forces, gravity,...) Internal friction in the fluid (supposed Newtonian) is proportional to the viscosity ν. Due to this friction, the boundary conditions are taken so that the fluid has zero velocity relatively to the boundaries. The friction will impose also that the motion of the fluid will decay if there is no forcing external to the fluid. For an incompressible flow, the continuity equation completes the problem: $\nabla\cdot\boldsymbol{v} = 0$. Remark that the pressure term is non-local because of a Poisson equation that relates p to $\boldsymbol{v}$: $\Delta p = -\partial^2(v_i v_j)/\partial x_i \partial x_j$.

The NS equation could be analyzed from its inner symmetries but, because the boundaries and the forcing will usually not satisfy the same symmetries, a simple approach adopted by physicists is to study turbulence in open systems far from the boundaries, in order to find a possible generic behavior of an incompressible turbulent fluid, disregarding the specific geometry of the boundaries. The purpose of this part is to provide, first an overview of the properties of this situation, called homogeneous turbulence, second some elements of its statistical modeling.

Dimensional analysis of turbulence.

A difficulty of the NS equation is the non-linearity of the convective term that is part of the inertial behavior of the fluid. On the one hand, one may expect solutions with irregular shapes but, on the other hand, the friction term works to impose some regularity on the solutions. The balance between the two effects is evaluated by engineering dimensional analysis [6]. Let U be a typical velocity, and L a typical length scale of the full flow (for instance the size of an experiment). Let us use the symbol $\sim$ for equality of typical values. Then: $(\boldsymbol{v}\cdot\nabla)\boldsymbol{v} \sim U^2/L$, and $\nu\Delta\boldsymbol{v} \sim \nu U/L^2$. The ratio is the Reynolds number Re and equals UL/ν. This is the only quantity left if one takes out the dimensions from the variables. When Re is large, the non-linear term is dominant and the flow becomes irregular, with motions at many difference length scales. Typical turbulent flows seem far from having symmetries: the flow is disordered spatially; it is unpredictable temporally with strong variations from one time to another; neither is the velocity clearly stationary, displaying excursions far from its mean during long period of times. Those events at long time-scales are mixed with unceasing short-time variations of the velocity.

A turbulent flow is very different from a flow with zero viscosity, even for the situation of fully developed turbulence, when Re $\to \infty$. Indeed, the energy dissipated in the flow is never zero because the irregularity of the solutions increases correspondingly, creating stronger gradients in the flow. The local dissipation is defined as $\varepsilon(\boldsymbol{r}) = \nu(\partial v_j \partial x_i + \partial v_i \partial x_j)^2/2$. If the flow is stationary, its mean along the time is constant. If further the flow is homogeneous, this mean equals the spatial mean: $\overline{\varepsilon} = \langle\, \varepsilon(\boldsymbol{r})\, \rangle_{\boldsymbol{r}}$. For a simple dimensional analysis, we keep only $\overline{\varepsilon}$ and ν as relevant parameters. From them, one can build a dissipative length scale $\eta = (\nu^3/\overline{\varepsilon})^{1/4}$ where the solution should become smooth because of the fluid friction. As a consequence, the estimated number of modes $(L/\eta)^3$ (needed for computer simulations) is, for the three-dimensional velocity, proportional to $\mathrm{Re}^{9/4}$. This number is too large to conveniently use methods from non-linear dynamical systems. Characterizing the flow directly from the NS equation is hard task because of all those properties.

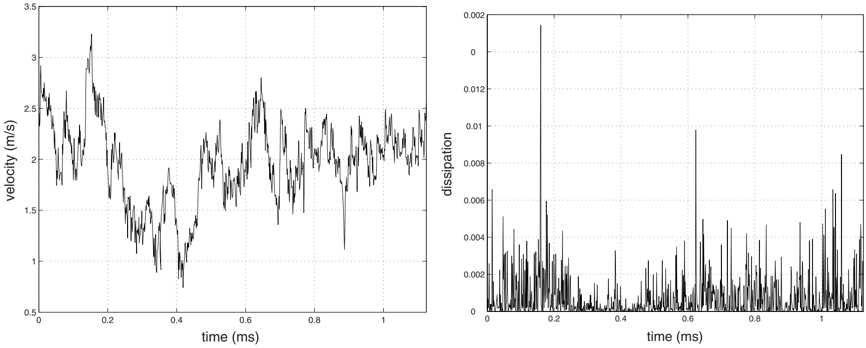


Fig. 1. Left: a typical Eulerian velocity signal $v(t)$. Right: its corresponding surrogate dissipation (derivative of the squared velocity). Those signals were measured by Pietropinto *et al.* as part of the GReC experiment [64] of fluid turbulence in low temperature gaseous helium.

Experimental Eulerian velocity.

A sample velocity signal $v(t)$ is shown on Figure 1 as an illustration. This signal was obtained during the experiment GReC [64] in a jet at high Reynolds number (up to 10^7) in helium at 4.5 K (so that its viscosity is very low). Experiments of turbulence consists in studying high speed motions in a fluid where laminarity is broken, for instance by means of a grid or by creating a jet and the flow becomes turbulent; here it is a jet. Common apparatus are hot-wire probes that measure one component $v(t)$ of the Eulerian velocity at one point (we choose to discuss only single probe measurements here). The erratic fluctuations are typical of such signals and one can see numerous points where the signal appears almost singular. The singular fluctuations are clearer from the time derivative of its energy $v(t)^2/2$, which is an estimation of the dissipation, the so-called surrogate dissipative signal [53]. This dissipative signal seems made of numerous peaks of variable amplitudes, separated by periods of almost no activity: the density of peaks is fluctuating from time to time.

Measurements of Eulerian velocity provide signals recorded along time. The Taylor hypothesis postulates that, if Re is large enough, the velocity field is advected quicker than it changes so that the evolution of $\boldsymbol{v}$ during a short time τ is mainly given by the dominant convection term. This way, the velocity at position $\boldsymbol{x}$ and at time $t-\tau$ is essentially the same as velocity $\boldsymbol{v}(\boldsymbol{x}+\tau\boldsymbol{v};t)$. This is an hypothesis of "frozen" turbulence during the time scale τ of the measurement. This hypothesis is the basic fact behind the choice of modeling spatial structure of the velocity field in the following, instead of its evolution along time.

1.2 Statistical modeling of Eulerian turbulence

A simple approach is to forget about the dynamical equation and find only the statistical properties of the velocity field [7, 13]. Knowing the complete initial velocity field of a turbulent fluid, and then following its proper evolution in time is hopeless because of the high number of modes and of the non-linearity of the equation. Experimental observations support this assertion: the fluid seems erratic, with many ever changing currents and eddies, and typical measurements of the Eulerian velocity as a function of time are strongly shambled signals. Forgetting about the initial conditions and exact geometrical setting, one can find simple models to describe statistical properties of the signal, assuming that $\boldsymbol{v}(\boldsymbol{r};t)$ is a random process indexed by $\boldsymbol{r}$ and t. We will review shortly main results obtained by this way. Many textbooks on the subject exist, see for instance [56, 33, 54], and we sketch here major steps on the subject that are especially relevant for signal processing questions.

Summary of the statistical properties of turbulence.

From the phenomenological and dimensional analysis of signals of turbulence, several properties are clear-cut. The signals have relevant characteristics at many scales: the fluctuations should be accounted for at small, intermediate and large scales in space and time. The signals are said intermittent both in space: a complex geometric structure of eddies; and time: irregularity and unpredictability of the time series. Added to that, the signals are also intermittent in statistics: there are large deviations that are evident in the dissipative signals displayed here. The distribution of the velocity is near to Gaussian. This does not mirror the apparent burstiness of experimental signals, that is related to the existence of a broad band of characteristic scales. Actually large fluctuations, far from the mean, exist both at large and small scales.

To encompass all these properties, turbulence is studied through the velocity increments $\delta v(\boldsymbol{r};\boldsymbol{x},t)$ over distance $|\boldsymbol{r}|$, and at time t and position $\boldsymbol{x}$: $\delta v(\boldsymbol{r};\boldsymbol{x},t)=v(\boldsymbol{x}+\boldsymbol{r};t)-v(\boldsymbol{x};t)$. One studies a component of the velocity $\boldsymbol{v}$, for instance the component parallel to $\boldsymbol{r}$. The velocity increment was introduced to probe the velocity at different scales, because turbulence is, before all, a multi-scale phenomenon. In question is then a model of the statistical properties of the random variables $\delta v(\boldsymbol{r};\boldsymbol{x},t)$.

Scale invariance and Self-similarity.

Kolmogorov proposed in 1941 [40, 41] a first statistical description of velocity increments (hereafter named K41 theory). He postulated that the velocity increment obeys statistically some symmetries that are compatible with the NS equation: time stationarity (independence from t), spatial homogeneity (independence from $\boldsymbol{x}$; note that this is valid because it models turbulence far from the boundaries) and isotropy (invariance under rotations). Let us recall that stationarity is the invariance under time-shifts $\mathcal{S}_\tau$; a stochastic process Y is stationary if and only if $(\mathcal{S}_\tau Y)(t) = Y(t+\tau) \overset{d}{=} Y(t)$ for any $\tau \in \mathbb{R}$. The equality $\overset{d}{=}$ should be understood as equality for all finite-dimensional probability distributions of the random variables. The other symmetries are defined in the same way with the corresponding operators.

To those symmetries, a property of scale invariance, or self-similarity, is added. Let us recall the definition of self-similarity: it is a statistical invariance under the action of dilations. Let $\mathcal{D}_{H,\lambda}$ be a dilation of scale ratio λ so that $(\mathcal{D}_{H,\lambda}X)(t) = \lambda^{-H}X(\lambda t)$. A random process $\{X(t), t \in \mathbb{R}_*^+\}$ is self-similar with exponent H (H-ss) if and only if for any $\lambda \in \mathbb{R}_*^+$, one has ([68])

$$\{(\mathcal{D}_{H,\lambda}X)(t), t \in \mathbb{R}_*^+\} \overset{d}{=} \{X(t), t \in \mathbb{R}_*^+\}.$$

For the velocity increments, this reads:

$$\delta v(\lambda\boldsymbol{r};\boldsymbol{x},t) \overset{d}{=} \lambda^{-h}\delta v(\boldsymbol{r};\boldsymbol{x},t) \quad \text{if} \quad \lambda \to 0^+. \tag{2}$$

This last property is also a prescription of the regularity of the solution because, if this relation holds for small separations $|\boldsymbol{r}|$, one solution is to have $\delta v(\boldsymbol{r}) = |\boldsymbol{r}|^h$ for small $\boldsymbol{r}$, and that rules the behavior of the derivative, and consequently of the dissipation. This defines the singularities and peaks expected in the dissipation signal.

With those symmetries, the only parameters left to describe the velocity are the mean dissipation $\overline{\varepsilon}$, the viscosity ν, the self-similarity exponent h and the length-scale $r = |\boldsymbol{r}|$ one considers. Kolmogorov supposes a full scale invariance so that all spatial scales behave the same, sharing the same mean dissipation so that for any r, $\overline{\varepsilon} = C[\delta v(r)]^2/[r/\delta v(r)]$, where C is some constant. Thus $\delta v(r) = c_1\overline{\varepsilon}^{1/3}r^{1/3}$: the velocity has a unique exponent of self-similarity $h = 1/3$. The moment of order p of δv is called the structure function of order p and obeys, according to this theory, the following relation:

$$\mathbb{E}\left\{|\delta v(r;\boldsymbol{x},t)|^p\right\} = c_p(\overline{\varepsilon}r)^{p/3} \qquad \text{if} \qquad \eta \ll r \ll L. \tag{3}$$

$\mathbb{E}$ is the mathematical expectation, i.e. the mean of the quantity; c_p are constants. Here L is an integral scale, that is a characteristic distance of the whole flow, for instance the scale of the forcing. The scales between η and L for which the scaling of Eq. (3) holds, are called the inertial zone because friction is small at those scales and the inertial effects are dominant for the NS equation, especially the convection term. Note that for order 3, the exponent is 1 and this is fortunate because the Kármán-Howarth equation derived from the NS equation imposes so [33]. The scaling law for $p = 2$ imposes the spectrum of the velocity by means of the Wiener-Khinchin relation. Kolmogorov's well-known prediction is that the spectrum should be: $S_v(k) = c_2 \nu^{5/4} \overline{\varepsilon}^{1/4} (k\eta)^{-5/3}$ if k is in-between $1/L$ and $1/\eta$ (the inertial zone). This is a property of long-range dependence: the spectrum and the correlation of the process decrease slowly. This is in this model related to scale invariance, or self-similarity.

The prediction for the spectrum holds well, as seen on Figure 2. The structure functions as a function of r are also roughly power laws r^{ζ_p}, but not exactly [9]. But the general prediction of (3) is found failing for other orders. Indeed, experimental exponents ζ_p depart from linearity predicted in $p/3$. We report in Figure 2 some properties of the structure functions: they look like power laws over the inertial range. On the right, we display the evolution of the exponents ζ_p of this power law with the order p of the moment, and the probability density function of the increments $\delta v(r)$ for different r.

Multifractality: Characterization in terms of singularities.

The failure of the previous theory is related to the spatial and temporal intermittency of the dissipation: random bursts of activity exist and the regularity of the signal changes from one point to another, and so does ε from one scale to another. The statistical self-similarity property (2) is now true only if h is also a random variable that depends on $\boldsymbol{x}$ and t. If this property holds for $\lambda \to 0^+$, h is called the Hölder exponent of the signal at point $\boldsymbol{x}$. The set of points sharing the same Hölder exponent is a complicated random set that is a fractal set with dimension $D(h)$. This is a multifractal model [32, 33] that describes the signal in terms of singularities at small scale. The underlying hypothesis is that all the statistics are ruled by those singularities. The complementary property of the multifractality is the conjecture of a relation between the singularity spectrum $D(h)$ and the scaling exponents ζ_p, by means of a Legendre transform: $D(h) = \inf_p(hp + 1 - \zeta_p)$. Mathematical aspects of multifractality and of its equality can be found in [37, 38]. Experimentally, in order to measure the multifractal spectrum that is the core of this model, one has first to compute a multiresolution quantity, then use a Legendre transform that is a statistical measure of $D(h)$ from the exponents ζ_p. Experiments now agree with $\zeta_p \simeq c_1 p - c_2 p^2/2$, where $c_1 \simeq 0.370$ and $c_2 \simeq 0.025$; this is a development in a power series p^n and terms p^n with $n \geq 3$ are too small to be correctly estimated nowadays. The corresponding singularity spectrum $D(h)$ is $1 - (h - c_1)^2/2c_2$, for values of h such that $D(h) \geq 0$. The expected value of h on a set of dimension 1 in the signal is 0.37, close to the 1/3 exponent predicted by Kolmogorov, but the local exponent fluctuates.

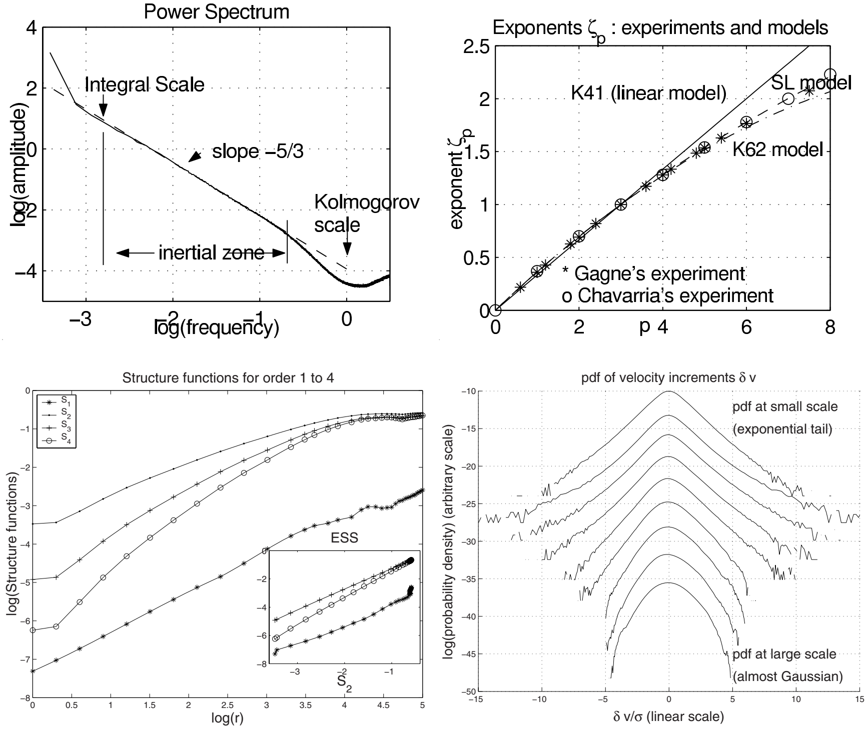


Fig. 2. Statistical analysis of one-point velocity measurements. Top left: spectrum $S_v(k)$ of the velocity that follows the K41 prediction of a power law of exponent 5/3. Bottom left: structure functions $S_p(r) = \mathbb{E}|\delta v(r)|^p$ for $p = 1$ to 4; inserted is shown the Extended Self-Similarity property [9]: the structure functions are not really power laws of r, but are acceptable power laws if drawn as a function of one another (on a log-log diagram). Top right : exponents ζ_p of the higher-order statistics (taken from [35] and [19]) are shown different from the K41 model, and closer to a multifractal models (here the Kolmogorov-Obhukov model of 1962 (K62) and the She-Lévêque (SL) model). Bottom right: pdf of the increments figured at various scales, from small scale (a few η) where the pdf is non-Gaussian with heavy tails, to large scale (around L) where the pdf is almost Gaussian; note that the scale is logarithmic for the pdf. The experimental spectrum and pdf are from the data of the GReC experiment [64].

Because the physical velocity should be a continuous signal at small scales (smaller than η), a further refinement in modeling is that singularities appear only in an analytic continuation of the velocity for complex times. The singularities in the velocity signal have the form $|t - z_0|^{h(t_0)}$, with $z_0 = t_0 + \mathrm{i}\zeta \in \mathbb{C}$, and are then a basis for multifractal interpretation. Such a distribution of singularities, having each a spectrum $k^{-2h-1}\mathrm{e}^{-2\zeta U k}$, leads to a mean spectrum consistent with quantitative measurements. Yet the existence of such isolated singularities was not proved nor derived from the NS equation, but only in simpler dynamical systems [30, 31].

Another approach was to relate the fluctuations of the exponents h to the dissipative scales η. Beneath the dissipative scale, the velocity is differentiable: $\delta v(r) = r\partial v/\partial x$. This small scale regularisation is obtained via a local dissipative scale [65, 34], defined as the scale where the local Reynolds number $\mathrm{Re}(r) = r\delta v(r)/\nu$ equals 1. In fact we have $\delta v(r) = U(r/L)^{h(\boldsymbol{x})}$ if $r > \eta(\boldsymbol{x})$ so that $\mathrm{Re}(r) = r\delta v(r)/\nu = (l/L)^{1+h(\boldsymbol{x})}\mathrm{Re}$. The dissipative scale is fluctuating locally as $\eta(h) = L\mathrm{Re}^{-1/(1+h)}$, whereas K41 uses a fixed dissipative scale $\eta = (\nu^3/\overline{\varepsilon})^{1/4}$ which is now the mean of the $\eta(h)$. Given this behavior, a unified description of the statistics $\mathbb{E}\left\{|\delta v(r;\boldsymbol{x},t)|^p\right\}$ was derived, valid both in the inertial and dissipative scales [22].

Characterization as random cascades.

We hereby test further statistical aspects of the intermittency of the flows; for this we stick with modeling only the statistics of the flows. A feature of equation (3) is notable: if the equation were true, the random variable $\delta v(r)/(\overline{\varepsilon}r)^{1/3}$ should be independent of r [18]. However experimental measurements of the probability density function (pdf) of $\delta v(r)$ shows that its shape changes with r, even in the inertial domain; see Figure 2. At large scale (close to L), the pdf is almost a Gaussian; when probing smaller scale, exponential tails become more and more prominent: rare intense events are more frequent at small scale – this is the statistical face of intermittency.

This property is best modeled as a multiplicative random process, where each scale is derived from the larger one. The general class of this model comes from the Mandelbrot martingales [42, 69] and was also developed from the experimental data in turbulence [59, 23, 63]. The challenge is to model dependencies between scale, for instance by means of multipliers between scales $W(r_1, r_2)$ defined by $\delta v(r_2) = W(r_1, r_2)\delta v(r_1)$. For the density probability function $P_{r_1}(\log|\delta v|)$ at scale r_1, this equation is a convolution between P_{r_1} and the pdf of the multipliers $\log(W(r_1, r_2))$. Because the relation holds for every couple of scales, the relevant solutions are infinitely divisible distributions. For instance, one can explicitly write [18]: $P_{r_2}(\log|\delta v|) = G^{\star[n(r_2)-n(r_1)]} \star P_{r_1}(\log|\delta v|)$, where $\star$ is a convolution. G is here the kernel of the cascade, that is the operator that maps the fluctuations from one scale r_1 to another r_2; equivalently, it gives the distribution of $\log(W(r_1, r_2))$. Derived from this, the structure functions read:

$$\mathbb{E}\left\{|\delta v(r;\boldsymbol{x},t)|^p\right\} = \mathrm{e}^{H(p)n(r)} \text{ with } H(p) = -\log\tilde{G}(p), \qquad (4)$$

where $\tilde{G}$ is the Laplace transform of G. The interest of multiplicative cascades seen as infinitely divisible processes is that this leads to elegant construction of stochastic processes satisfying exactly the relations (4) [14], and they can be used as benchmark for the estimation tools of multifractality [15, 21]. A consequence of the model is that if $n(r)$ is close to $\log r$, the structure function obeys a power law with exponents $\zeta_p = H(p)$. If not, the property is the so-called Extended Self-Similarity because all orders share the same law $\mathrm{e}^{n(r)}$ and for instance, with $\zeta_3 = 1$: $\mathbb{E}\left\{|\delta v(r;\boldsymbol{x},t)|^p\right\} = (\mathbb{E}\left\{|\delta v(r;\boldsymbol{x},t)|^3\right\})^{H(p)}$, as illustrated on Figure 2.

1.3 Vortex modeling for turbulence and oscillating singularities

The models reported were built on multi-scale properties of the velocity and on its singularities, and they are good descriptions of the data. Nevertheless, these models lack connections with the NS equation and with the structured organization of turbulent flows which are not purely random flows. One would like to characterize a flow from its own structures. Experiments of turbulence show that there are intense vortices: objects similar to stretched filaments around which the particles are mainly swirling [25]. The singularities in velocity signals could then be understood as features of a few organized objects with a complex inner structuration and a singular behavior near their core [51, 36]. A mechanism could be spiraling structures, analogous to the phenomenon of a Kelvin-Helmholtz instability [52]. Lundgren studied a specific collection of elongated vortices having a spiraling structure in their orthogonal section, and that are solution of the NS equation given a specified strain [46]. It was shown that such a collection could be responsible for a spectrum in $k^{-5/3}$ and intermittency of the structures functions consistent with modern measurements of ζ_p [66]. Turbulence is understood in this case as some superposition of building objects with complex geometrical characteristics, such as oscillations or fractality (now in a geometrical, not statistical, way).

• A simple model for corresponding Eulerian velocity signals would be an accumulation of complex singularities. This is different from modeling singularities in complex times in the sense that here the exponent is complex $(t-t_0)^{h+\mathrm{i}\beta}$, not the central time t_0 of the singularity. See some examples of those functions on Figure 5. The exponent β is responsible for oscillations in the signal and multifractal estimation is perturbed by such oscillations [1].

The Fourier spectrum of a function $\mathrm{e}^{-a(t-t_0)}(t-t_0)^{h+\mathrm{i}\beta}$ behaves like $\mathrm{e}^{4\pi\beta\,\mathrm{atan}(2\pi\nu/a)}|4\pi^2\nu^2+a^2|^{-h-1}$; except at low frequencies, so when $\nu \gg a$, the spectrum scales like $|\nu|^{-2h-2}$. This is a power law so they can be used as basis functions to built a synthetic signal with properties of turbulence. A sum of many functions of this kind may have multifractal properties that depend on the distribution of the h and β exponents [17]. One is then interested to find whether or not there are such oscillations in velocity signals.

• The consequences of the existence of spiraling structures for Lagrangian velocity would be the existence of swirling motions when a particle is close to a vortex core. Far from vortices, the motion should be almost ballistic, with small acceleration. A consequence is that expected Lagrangian trajectories will go through periods of large acceleration and periods of almost no acceleration. Non-stationary descriptions would then give interesting characterizations of the velocity.

• The vortices and the swirling motions are described by the vorticity $\boldsymbol{\omega} = \nabla \wedge \boldsymbol{v}$. Vorticity is related to dissipation since $\overline{\varepsilon} = -\nu \langle\, |\boldsymbol{\omega}|^2 \,\rangle_{\boldsymbol{r}}$. If vortices are relevant features of a flow, vorticity should be strongly organized in those specific structures. One expects that they can be detected as isolated objects and a question is their role in intermittency. Hereto the non-stationary evolution of those objects is an expected feature.

To sum up, the general problem is that one can not easily track at the same time the three kinds of interesting properties for turbulence: non-stationarity of the signals; the inner oscillating or geometric structure; and the statistical self-similar properties (exponent h or multifractality) of the spiraling vortices or their consequence for velocity.

Alternative representations of signals.

Dealing with these three properties, we know how to construct a representation jointly suited to two of them at the same time. The third one is then difficult to assess.

1) Time evolution and self-similarity: statistical methods using wavelets are adapted to multifractal models or random cascades because they probe statistical quantities of stationary signals with relevant self-similar properties but no inner oscillations [1, 39].

2) Time evolution and Fourier analysis: modern Lagrangian and vorticity measurements are made possible by following the instant variation of the Fourier spectrum of some non-stationary signal. Neither the temporal nor the spectral representation is enough: time-frequency representations that unfold the information jointly in time and frequency [28] are needed.

3) Self-similarity and inner geometry: one may be interested in oscillations and self-similarity at the same time. It is known that wavelets are not well adapted to study oscillations [1]. A variant is to measure geometry in a non-stationary context (since self-similarity implies non-stationarity). Ad-hoc procedures constructed on the wavelet transform [43] or on the Mellin-time representations [17] were considered, but for now without clear-cut results. The third part of this section is devoted to the Mellin representation that is adapted to probe self-similarity and some features of geometry because it is based on self-similar oscillating functions $(t - t_0)^{h+i\alpha}$.

To conclude this overview of turbulence, let us summarize the complexity of fluid turbulence. The problem is driven by a non-linear PDE that is reluctant to mathematical analysis. Still we dispose of strong phenomenological properties to build stochastic modeling of the velocity. The signals are irregular, intermittent and one would like to question their (multi)-fractal aspects, their singularities but also situations where their geometrical organization or some non-stationary properties are more relevant. Because there exists no single method that capture all these features, multiple tools of signal processing are useful.

2 Signal Processing Methods for Experiments on Turbulence

2.1 Some limitations of Fourier analysis

Physicists often describe signal in term of their harmonic Fourier analysis; it first relies on order 2 statistics, through the spectral analysis of the signal. The well-known Fourier representation reads: $v(t) = \int \mathrm{e}^{-2\mathrm{i}\pi t\nu} \mathrm{d}\xi(\nu)$. This decomposition as spectral increments $\mathrm{d}\xi(\nu)$ on the cosine basis is especially suited if v is stationary. In this case, its spectrum S_v is given by $\mathbb{E}\left\{\mathrm{d}\xi(\nu_1)\overline{\mathrm{d}\xi(\nu_2)}\right\} = S_v(\nu)\delta(\nu_1 - \nu_2)\mathrm{d}\nu_1\mathrm{d}\nu_2$. The spectral increments are thus uncorrelated. In the context of turbulence, one sees on Figure 2 an estimate of $S_v(\nu)$, using standard signal processing tools. The support of the spectrum is broad-band and $S_v(\nu)$ follows roughly a power law with cut-offs at the inertial scale L, and at the small scale were dissipation becomes dominant (around η). This corresponds to the lack of a single time scale of evolution. On Figure 2 is displayed the spectrum of Eulerian velocity. The lack of separate characteristic frequencies (or times) is evident. The spectrum follows closely the Kolmogorov $k^{-5/3}$ law.

A first difficulty to use Fourier representation in turbulence, is that the higher-order statistics and the geometric organization is coded in the phase of the Fourier transform. This information is hard to recover. For instance, realizations of the random Weierstrass functions (that will be defined later), that are fractal, and of oscillating singularities (for instance the function $|t - t_0|^{h+\mathrm{i}\beta}$), that are not, would share the same Fourier spectrum but not the same statistical and fractal properties [36]. Redundant representations, depending jointly of time and frequency or scale variables, will be found to capture those properties in a clearer way.

A second problem in the context of turbulence is the long-range dependence in the times series [10]. Generally, this reduces the performance of estimation of all classical quantities from the time series, including the Fourier transform. For instance, let us suppose that $X(t)$ is a stationary process with long memory, or long-range dependence, i.e. its correlation decreases like τ^{2H-2} when τ is large. If the process is known for n samples, the periodogram $I_n(\nu)$ is computed from the Fourier transform:

$$I_n(\nu) = \frac{1}{n}\left|\sum_{t=1}^{n} X(t)\mathrm{e}^{-2\mathrm{i}\pi\nu t}\right|^2 . \tag{5}$$

The periodogram is an estimate of the spectral density $S_X(\nu)$. For frequencies ν that are close to zero, the asymptotic variance of the periodogram is of the order: $\mathrm{Var}[I_n(\nu)] = O(n^{4H-2})$. It means that for low frequencies, it fluctuates much more than for short-range correlated processes, and provides a cruder estimation.

These short-comings of the Fourier transform are motivations to study different kinds of representations, which will be more adapted to multi-scale properties, singularities and long-memory.

2.2 Multiresolution characterization and estimation of scaling laws

Velocity increments.

In order to question all the time scales in a turbulent signal, the velocity increment over the time separation was introduced as a more relevant quantity: $\delta v(\tau; \boldsymbol{x}, t) = v(\boldsymbol{x}; t - \tau) - v(\boldsymbol{x}; t) = v(\boldsymbol{x} + \boldsymbol{r}; t) - v(\boldsymbol{x}; t)$. The second equality is obtained from the Taylor hypothesis, provided that $\boldsymbol{r} = \tau \boldsymbol{v}(\boldsymbol{x}; t)$. Velocity increments are a multiresolution quantity in the sense that they describe the velocity at the varying resolution τ. They are relevant to capture both long-time evolution of the signal that are dominated by the statistics of v because $v(t-\tau)$ and $v(\tau)$ are then almost independent, and short time behaviors where the dominant features are intermittent peaks of activity seen in the derivative of the signal.

Wavelet decompositions.

More general multiresolution quantities exist beside velocity increment, which are not the most well-behaved for estimation in presence of long dependence. A class of multiresolution representation is the wavelet transform [49, 48]:

$$T_v(a,t) = \int v(u)\psi([u-t]/a)\mathrm{d}u/a, \tag{6}$$

provided that the mother-wavelet $\psi(t)$ has zero integral. The wavelet is further characterized by an integer $N \geq 1$, the number of vanishing moments. The representation is then blind to polynomial trends of order less than N, and this gives robustness to the representation regarding the slow, large-period excursions that one finds in signals of turbulent velocity. Velocity increments are “the poor man’s wavelet”, setting $\psi(u) = \delta(u+1) - \delta(u)$ and letting τ be the scale variable a. This wavelet has only one vanishing moment, $N = 1$. With a larger number of vanishing moments, wavelets give good methods of estimation when one expects power law statistics, such as self-similarity or multifractality. We report briefly two estimation methods.

Wavelet transforms and singularities.

A property of the wavelet transform is that it captures the singularities of a signal on the maximum of the wavelet [48, 37]. Let us assume that $v(t-\tau) - v(t)$ behaves like $|\tau|^h$ near point t. If the number N of vanishing moments of ψ is higher than h, the wavelet transform will have a maximum in the scale-time cone (a, u) defined by $|u - t| \leq Ca$ for some constant C. This maximum behaves as a^h when $a \to 0^+$. This property permits one to estimate directly the Hölder exponents of singularities. In the context of multifractality, the singularities are not isolated and it is helpless to try to estimate separately the exponents. Combining the multifractal formalism with the properties on maxima of wavelet transform, it was proposed to estimate the spectrum $D(h)$ by computing the moments of velocity on those maxima only. This is called the Wavelet Transform Maxima Modulus method [4] and it gives reliable estimations of $D(h)$ in turbulence [5]. A limit is that there are few theoretical calculations of $D(h)$ that validate the WTMM method, and so there is not a complete mathematical justification of it.

Wavelet transform and estimation of scaling laws.

Another possibility is to take advantage of discrete orthogonal wavelet basis (see the text of J.R. Partington, page 95). Let $\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k)$ denote its dilated and translated templates on the dyadic grid, and $d_X(j,k) = \int \psi_{j,k}(u)X(u)\mathrm{d}u$, the corresponding discrete wavelet coefficients. For any second order stationary process X, its spectrum $S_X(\nu)$ can be related to its wavelet coefficients through:

$$\mathbb{E}\left\{d_X(j,k)^2\right\} = \int S_X(\nu)2^j|\Psi(2^j\nu)|^2\mathrm{d}\nu, \tag{7}$$

where Ψ stands for the Fourier transform of ψ. This is an estimation of the spectrum. One can also recover statistics of all orders. If X is a self-similar process, with parameter H, they behave as:

$$\mathbb{E}\left\{d_X(j,k)^p\right\} = C2^{jpH}, \ \text{ if } \ 2^j \to +\infty. \tag{8}$$

Moreover, it has been proven that the $\{d_X(j,k), k \in \mathbb{N}\}$ form short range dependent sequences as soon as $N - 1 > H$. This means that they no longer suffer from statistical difficulties implied by the long memory property. In particular, the time averages $S(j;p) = 1/n_j \sum_{k=1}^{n_j} |d_X(j,k)|^p$ can then be used as relevant, efficient and robust estimators for $\mathbb{E}\left\{d_X(j,k)^p\right\}$. The possibility of varying N brings robustness to these analysis and estimation procedures. The performance of the estimators was studied, see for instance [3]. One can then characterize all the statistics of X from the following estimation procedure: For a velocity signal v, a weighted linear regression of $\log_2 S(j;p)$ against $\log_2 2^j = j$, performed in the limit of the coarsest scales, provides with an estimate of the exponents ζ_p of the structure functions $\mathbb{E}\left\{|\delta v(r)|^p\right\}$.

Combining the WTMM idea and the properties of discrete wavelet transform, Jaffard proposed an exact characterization of multifractal signals using wavelet leaders (local maxima of discrete wavelet coefficients) [39] that are now developed as signal processing tools [45].

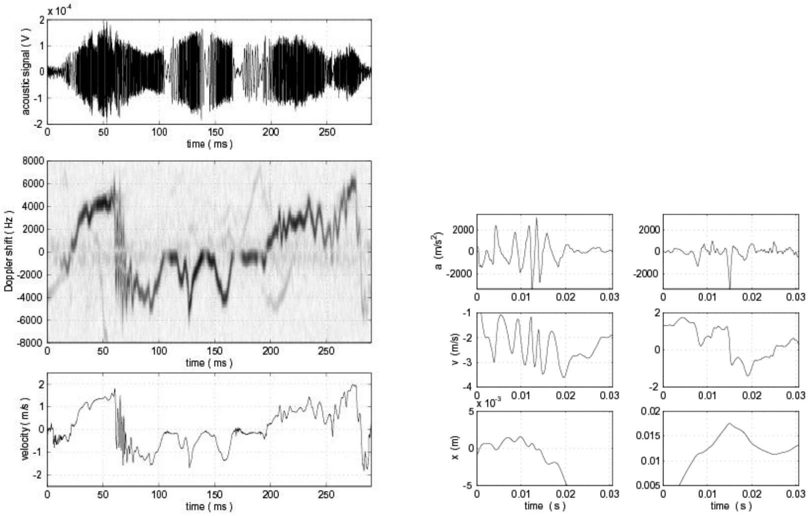


Fig. 3. Lagrangian velocity of a particle in turbulence (from [50]). Left: the Doppler signal whose instantaneous frequency gives the velocity of the tracked solid particle in a turbulent fluid, and its time-frequency representation. Right: acceleration, velocity and trajectory, reconstructed for two components from the measurement of velocity by Doppler effect.

2.3 Time-frequency methods for Lagrangian and Vorticity measurements

Time-frequency representations.

A linear time-frequency decomposition is achieved in the same manner as a wavelet transform, using a basis built from shifts in time and frequency of a small wave packet:

$$v(t) = \iint r_v(u,\nu) b_{u\nu}(t)\, \mathrm{d}u\, \mathrm{d}\nu, \text{ with } b_{u\nu}(t) = b_0(t-u)\mathrm{e}^{-2\mathrm{i}\pi\nu t}.$$

The variable ν is indeed a frequency and $r_v(u,\nu)$ gives the component of v at frequency ν and time u. The time-frequency spectrum is $\mathbb{E}\left\{|r_v(u,\nu)|^2\right\}$. Note that instead of time and frequency shifts, the wavelet transform uses time-shifts and dilation on the mother wavelet, so that the variables are time and scale rather than of time and frequency.

If one is interested in the time-frequency spectrum, it is possible to achieve better estimation using bilinear densities that are time-frequency decompositions of the energy [28]. They derive from the Wigner-Ville distribution:

$$W_v(t,\nu) = \int v(t+\tau/2)\overline{v(t-\tau/2)}\mathrm{e}^{-2\mathrm{i}\pi\nu\tau}\mathrm{d}\tau.$$

A general class is obtained by applying some smoothing in time and/or frequency. Such a distribution represents well the energy of the signal because of the following physical properties.

1. Marginals in time and frequency: $\int W_v(t,\nu)\mathrm{d}\nu = |v(t)|^2$; $\int W_v(t,\nu)\mathrm{d}t = |V(\nu)|^2$ if V is the Fourier transform of v.
2. Covariances with time and frequency shifts: $W_v(t-\tau,\nu)$ is the transform of $v(t-\tau)$ and $W_v(t,\nu-f)$ is transform of $\mathrm{e}^{2\mathrm{i}\pi ft}v(t)$.
3. Instantaneous frequency: the mean frequency $\int W_v(t,\nu)\nu\mathrm{d}\nu$ is equal to the instantaneous frequency of the signal $v(t)$, that is the derivative of the phase of the analytic signal associated to $v(t)$.

Representations of this kind are used, because of the properties, to analyze the non-stationary signals of Lagrangian experiments and of vorticity measurements.

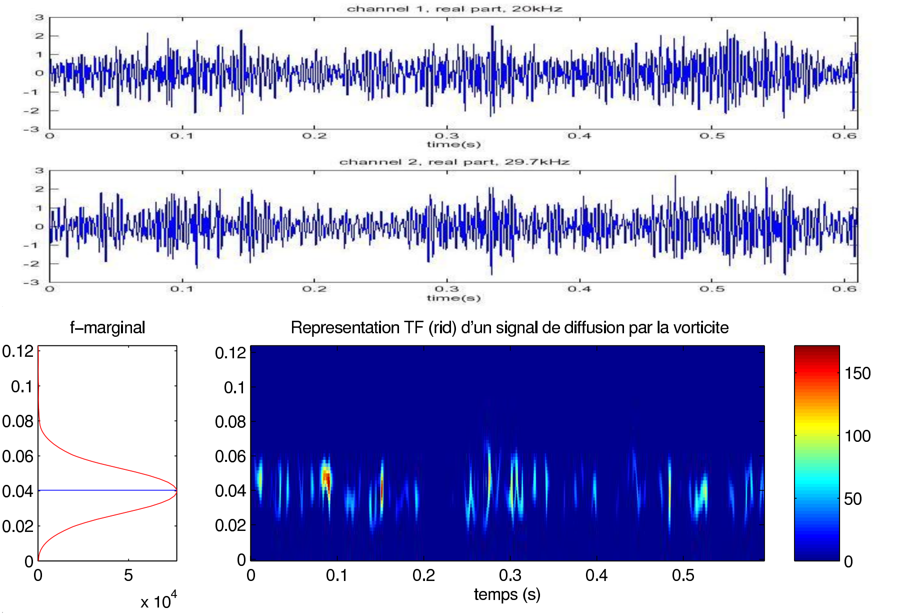


Fig. 4. Measurement of vorticity by acoustic scattering [16]. Up: examples of recorded signals for two different scattered waves at the same time by the same volume: they both represent the same $\tilde{\boldsymbol{\omega}}_i(\boldsymbol{k},t)$ along time t. Down: quadratic time-frequency representation of one signal, exhibiting packets of structured vorticity advected through the measurement volume. [16]

Measurements of Lagrangian velocity.

Recent experiments have been able to find characteristics of Lagrangian velocity. Solid particles are released in a turbulent fluid, then tracked to record their Lagrangian velocities $\boldsymbol{u}(t)$ [47, 50]. One solution uses high-speed detectors to record the trajectories, and the second one relies on tracking by sonar methods. In both cases the experiment deals with a non-stationary signal that should be tracked in position and value along time. In the second experiment, ultra-sonor waves are reflected by the particle and the Doppler effect catches its velocity. Figure 3 shows a sample experimental signal whose instantaneous frequency is the Lagrangian velocity. A time-frequency analysis follows the instantaneous frequency and thus $u(t)$. Acceleration, velocity and trajectory are reconstructed from this data. The signals contain many oscillating events such as the one figured here, and many more trajectories which are almost smooth and ballistic between short periods of times with strong accelerations. This is consistent with the existence of a few swirling structure but a clear connection between oscillations and intermittency is not made. By now, statistical analysis of the data show that Lagrangian velocity is intermittent [50], and this is well described by a multifractal model analogous to the one for Eulerian velocity [22].

Measurements of vortices and of vorticity.

Instead of trying to find indirect effects of the vortices, the intermittency of turbulence was looked after directly in vorticity. Measuring locally $\boldsymbol{\omega}$ is difficult and by now not reliable. Using the sound scattering property of vorticity, an acoustic spectroscopy method was developed [16]. The method measures a time-resolved Fourier component of vorticity, $\tilde{\boldsymbol{\omega}}_i(\boldsymbol{k}, t) = \int \boldsymbol{\omega}_i(\boldsymbol{r}, t) \mathrm{e}^{-2\mathrm{i}\pi \boldsymbol{k}\cdot\boldsymbol{r}} \mathrm{d}\boldsymbol{r}$, summed all over some spatial volume. Figure 4 shows recorded signals of scattering amplitudes for two different incident waves; they look alike because both are measurements of the same quantity, $\tilde{\boldsymbol{\omega}}_i(\boldsymbol{k}, t)$. The intermittency here is the existence of bursts of vorticity that cross the measurement volume; those packets are characteristic of some structuration of vorticity, which could be vortices. They are revealed in the time-frequency decomposition of one signal on the right. The intermittency is well captured by the description of a slow non-stationary activity that drives many short-time bursts, and so causes multi-scale properties [62].

2.4 Mellin representation for self-similarity

Another signal processing method uses oscillating functions as basis functions: the Mellin transformation. Its interest is that it is encompasses both self-similar and oscillating properties in one description. Because those tools are less known, we will survey some of their properties with more mathematical details.

Dilation and Mellin representation.

We aim at finding a formalism suited to scale invariance. Self-similarity is a statistical invariance under the action of dilations. Given exponent H, the group $\{\mathcal{D}_{H,\lambda}, \lambda \in \mathbb{R}_*^+\}$ is a continuous unitary representation of $(\mathbb{R}_*^+, \times)$ in the space $L^2(\mathbb{R}_*^+, t^{-2H-1}\mathrm{d}t)$. The associated harmonic analysis is the Mellin representation. Indeed, the hermitian generator of this group is $\mathcal{C}$ defined as: $2\mathrm{i}\pi(\mathcal{C}X)(t) = (-H + t\mathrm{d}/\mathrm{d}t)X(t)$, so that $\mathcal{D}_{H,\lambda} = \mathrm{e}^{2\mathrm{i}\pi\lambda\mathcal{C}}$. The operator $\mathcal{C}$ characterizes a scale because its eigenfunctions are unaffected by scale changes (dilations), so the eigenvalues are a possible measure of scale. Those eigenvalues $E_{H,\beta}(t)$ satisfy $\mathrm{d}E_{H,\beta}(t)/E_{H,\beta}(t) = (H + 2\mathrm{i}\pi\beta)\mathrm{d}t/t$, thus $E_{H,\beta}(t) = t^{H+2\mathrm{i}\pi\beta}$ up to a multiplicative constant. One obtains the basis of Mellin functions with associated representation:

$$\begin{aligned}(\mathcal{M}_H X)(\beta) &= \int_0^{+\infty} t^{-H-2\mathrm{i}\pi\beta} X(t)\frac{\mathrm{d}t}{t}\\ X(t) &= \int_{-\infty}^{+\infty} E_{H,\beta}(t)(\mathcal{M}_H X)(\beta)\mathrm{d}\beta.\end{aligned} \tag{9}$$

A signal processing view of several applications the Mellin transform may be found in [20, 8, 29, 58], and mathematical aspects are documented in [24, 71]. Relevant features here are, first, that β is a meaningful scale, and, second, the oscillating aspects of the Mellin functions $E_{H,\beta}(t)$. Those functions are chirps of instantaneous frequency β/t. See a drawing of such a function on Figure 5. One can disregard the behavior of those functions near 0; the important feature is the chirp part and it holds even if the function is filtered by some window, as seen on this figure. By this means we may describe both self-similarity and oscillations, as long as they can be well approximated by smoothed Mellin function.

Interpretation for self-similarity

When introducing self-similarity [44], J. Lamperti noticed a specific property of the invertible transformation $\mathcal{L}_H$, now called the Lamperti transformation and defined as:

$$(\mathcal{L}_H Y)(t) = t^H Y(\log t),\ t > 0; \qquad ({\mathcal{L}_H}^{-1} X)(t) = \mathrm{e}^{-Ht} X(\mathrm{e}^t),\ t \in \mathbb{R}. \tag{10}$$

This transformation maps stationary processes onto self-similar processes, and the converse for its inverse. The Lamperti transformation is a unitary equivalence between the group of time shifts $\mathcal{S}_\tau$ and the group of dilations $\mathcal{D}_{H,\lambda}$:

$${\mathcal{L}_H}^{-1}\mathcal{D}_{H,\lambda}\mathcal{L}_H = \mathcal{S}_{\log\lambda} \qquad \text{and} \qquad \mathcal{L}_H\mathcal{S}_\tau{\mathcal{L}_H}^{-1} = \mathcal{D}_{H,\mathrm{e}^\tau}. \tag{11}$$

This equivalence has interesting consequences: a natural representation of a self-similar process X is to use its stationary generator ${\mathcal{L}_H}^{-1}X$. Signal processing for stationary signals is a well-known field and methods can then be converted in tools for self-similar processes by applying equivalence (11) [27, 11]. In this context, Mellin representation is suited to H-ss processes in the same way as Fourier representation is suited to stationary processes, since $\mathcal{M}_H = \mathcal{F}{\mathcal{L}_H}^{-1}$:

$$(\mathcal{M}_H X)(\beta) = \int_0^{\infty} t^{-H} X(t) t^{-2\mathrm{i}\pi\beta - 1} \mathrm{d}t \tag{12}$$

$$= \int_{-\infty}^{+\infty} ({\mathcal{L}_H}^{-1} X)(u) \mathrm{e}^{-2\mathrm{i}\pi\beta u} \mathrm{d}u = (\mathcal{F}{\mathcal{L}_H}^{-1} X)(\beta). \tag{13}$$

Canonical spectral analysis of self-similar processes.

A H-ss process $X(t)$ has a covariance that reads necessarily as:

$$R_X(t,s) \hat{=} \mathbb{E}\left\{X(t)X(s)\right\} = (ts)^H c_X(t/s).$$

This comes from the correlation function $\gamma_Y(\tau)$ of its stationary generator $Y = ({\mathcal{L}_H}^{-1} X)$, with $\gamma_Y(\log k) = c_X(k)$. The Mellin spectral density $\varXi_X(\beta)$ of X is then simply introduced by means of the spectrum of Y:

$$\begin{aligned} \varGamma_Y(\beta) &= \int_{-\infty}^{+\infty} \gamma_Y(\tau) \mathrm{e}^{-2\mathrm{i}\pi\beta\tau} \mathrm{d}\tau \\ &= \int_0^{+\infty} c_X(k) k^{-2\mathrm{i}\pi\beta-1} \mathrm{d}k = (\mathcal{M}_0 c_X)(\beta) \hat{=} \varXi_X(\beta). \end{aligned} \tag{14}$$

H-ss processes admit also an harmonisable decomposition on the Mellin basis so that $X(t) = \int t^{H+2\mathrm{i}\pi\beta} \mathrm{d}\underline{X}(\beta)$, with uncorrelated spectral increments $\mathrm{d}\underline{X}(\beta)$. Thus we have $\mathbb{E}\{\mathrm{d}\underline{X}(\beta_1)\overline{\mathrm{d}\underline{X}(\beta_2)}\} = \delta(\beta_1 - \beta_2)\ \varXi_X(\beta_1)\mathrm{d}\beta_1\mathrm{d}\beta_2$.

Among the tools coming from the Lamperti equivalence, there are scale invariant filters. A linear operator $\mathcal{G}$ is invariant for dilations if it satisfies $\mathcal{G}\mathcal{D}_{H,\lambda} = \mathcal{D}_{H,\lambda}\mathcal{G}$ for any scale ratio $\lambda \in \mathbb{R}_*^+$. Using equation (11), we may replace $\mathcal{D}_{H,\lambda}$ by $\mathcal{S}_{\log\lambda}$ and we obtain the equality:

$$({\mathcal{L}_H}^{-1}\mathcal{G}\mathcal{L}_H)\mathcal{S}_{\log\lambda} = \mathcal{S}_{\log\lambda}({\mathcal{L}_H}^{-1}\mathcal{G}\mathcal{L}_H).$$

Thus, ${\mathcal{L}_H}^{-1}\mathcal{G}\mathcal{L}_H = \mathcal{H}$ is a linear stationary operator, so it acts as a filter by means of a convolution. The Lamperti transformation maps addition onto multiplication so that $\mathcal{G}$ will act by means of a multiplicative convolution instead of the usual one:

$$(\mathcal{G}X)(t) = \int_0^{\infty} g(t/s) X(s) \frac{\mathrm{d}s}{s} = \int_0^{\infty} g(s) X(t/s) \frac{\mathrm{d}s}{s}. \tag{15}$$

Let us consider $A = \mathcal{G}X$ with $\{X(t), t > 0\}$ and H-ss process and $\mathcal{G}$ a scale invariant filter. Then $A(t)$ is also self-similar because

$$\mathcal{D}_{H,\lambda} A = \mathcal{D}_{H,\lambda}\mathcal{G}X = (\mathcal{G}\mathcal{D}_{H,\lambda})X = \mathcal{D}_{H,\lambda}X.$$

This filter acts on the Mellin spectrum as a multiplication:

$$\varXi_A(\beta) = |(\mathcal{M}_H g)(\beta)|^2\ \varXi_X(\beta).$$

By means of the Bochner theorem, any H-ss process may be represented by the output of a scale-invariant linear system:

$$X(t) = \int_0^{+\infty} g(t/s)V(s)\frac{\mathrm{d}s}{s}, \text{ with } \mathbb{E}\left\{V(t)\overline{V(s)}\right\} = \sigma^2 t^{2H+1}\delta(t-s). \quad (16)$$

The random noise $V(t)$ is white and Gaussian but non-stationary; it is the image by $\mathcal{L}_H$ of the Wiener process. The self-similar process X is defined by g; the second-order properties are covariances given by means of

$$c_X(k) = \sigma^2 k^{-H} \int g(k\theta)\overline{g(\theta)}\theta^{-2H-1}\mathrm{d}\theta,$$

and Mellin spectrum which is $\varXi_X(\beta) = \sigma^2|(\mathcal{M}_H g)(\beta)|^2$. Models of this kind were studied in [70, 57].

Other methods are derived in the same way. For instance, time-frequency methods that were suited to measure jointly time and frequency components of a signal will be converted in time-Mellin scale representations that measure contents as a joint function of time and Mellin scale.

Examples of self-similar processes.

A fractional Brownian motion B_H is defined as a H-ss process with Gaussian stationary increments [55]. Its covariance is necessarily:

$$R_{B_H} = \sigma^2(|t|^{2H} + |s|^{2H} - |t-s|^{2H})/2$$

which satisfies the general expected structure with

$$c_{B_H}(k) = \sigma^2[k^H + k^{-H} - |\sqrt{k} - 1/\sqrt{k}|^{2H}]/2.$$

The corresponding Mellin spectrum is obtained by a straightforward calculus (Γ is the Euler function):

$$\varXi_{B_H}(\beta) = \frac{\sigma^2}{H^2 + 4\pi^2\beta^2}\left|\frac{\Gamma(1/2 + 2\mathrm{i}\pi\beta)}{\Gamma(H + 2\mathrm{i}\pi\beta)}\right|^2. \quad (17)$$

Here, we have a representation of fractional Brownian motions alternative to its harmonic or moving-average representations [67]. From this spectral representation, one can synthesize exact samples of fractional Brownian motions: it is enough to prescribe Mellin spectral increments satisfying equation (17) with random i.i.d. phases in $[0, 2\pi[$. An inverse Mellin transform gives then a fractional Brownian motion. Classical methods of whitening, prediction and interpolation for this process were derived from this Mellin representation in [60, 61]. Developments of the synthesis method from the Mellin spectrum for other self-similar processes without stationary increments were studied also in [17].

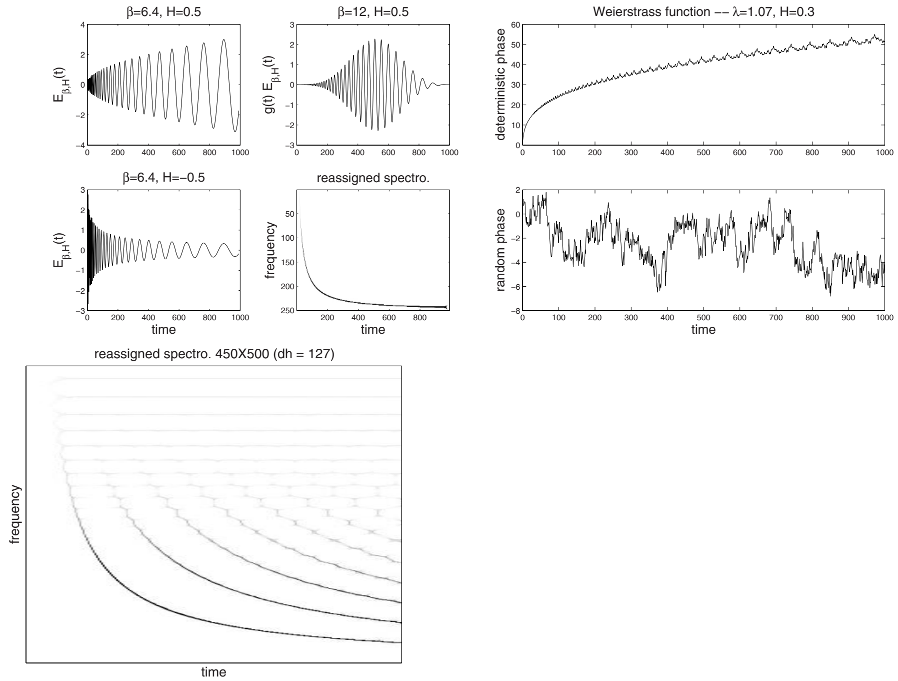


Fig. 5. Left: Mellin functions with various H, and spectrogram of one smoothed Mellin function (where $g(t)$ is a Kaiser window) that shows the instaneous frequency path, chirp behavior of the Mellin functions. Middle: samples of Weierstrass-Mandelbrot functions, both deterministic and random ($H = 0.3$, $\lambda = 1.07$). Right: spectrogram of the empirical variogram of a Weierstrass-Mandelbrot function (adapted from [26]). Spectograms are computed here using reassignment techniques for time-frequency distributions [2].

The random Weierstrass-Mandelbrot function is a good model of inexact self-similarity that can be studied by means of a Mellin decomposition. It is a step towards properties closer to turbulence than pure self-similarity. It is defined [12] as $W(t) = \sum_{n\in\mathbb{Z}} \lambda^{-nH}(1 - \mathrm{e}^{\mathrm{i}\lambda^n t})\mathrm{e}^{\mathrm{i}\phi_n}$, with i.i.d. phases ϕ_n. The function is given here as a sum of Fourier modes. This is possible since it has stationary increments. But another feature is more obvious if one considers its decomposition on a Mellin basis, namely its scale invariance. $W(t)$ has Discrete Scale Invariance [11] because $W(\lambda^k t) \overset{d}{=} \lambda^{-kH} W(t)$, scale invariance for dilations with a scale ratio that is a power of λ only. Using $\mathcal{L}_H$, one can find up the Mellin representation for the deterministic version of the function, with $\phi_n = 0$, [12, 26]:

$$W(t) = \sum_m \frac{-\Gamma(-H - m/\ln\lambda)}{\ln\lambda} \mathrm{e}^{[-\mathrm{i}\pi(H+m/\ln\lambda)/2]} E_{H,m/\ln\lambda}(t).$$

The two writings of $W(t)$ are its time-frequency representation and its time-Mellin scale representation. Both methods of analysis are valid as tools to assess the characteristics of the function. The relevance comes from the joint properties of stationary increments and self-similarity (even in the weakened sense of Discrete Scale Invariance). A time-frequency analysis illustrates this, see Figure 5. Deterministic and randomized versions of $W(t)$ have a spectrogram (from the detrended empirical variogram) that is made partly of pure tones, and partly of chirps, that are localized on the Mellin modes $\beta = m/\ln\lambda$. Here both aspects are shown, depending on the width of the smoothing window with respect to the rapidity of variation of the chirp (one see the chirp when its frequency does not change quickly over the length of the window) [26].

Concluding remarks.

We lectured here a signal processing view of turbulence. We have surveyed how the complexity of turbulence, and the need to understand various models and experiments, is linked to a great diversity of signal processing methods that are useful for turbulence: time-scale analysis, time-frequency analysis, self-similarity and Mellin analysis, and geometrical characterizations.

Concerning the last point, we are far from having at disposal convenient tools for estimation of the geometry (fractal sets, oscillations,...) of a self-similar process. We have proposed here a framework adapted to self-similarity and based on the oscillating Mellin functions $t^{h+2i\pi\beta}$ but a tractable extension to oscillating singularities of the form $|t-t_0|^{h+2i\pi\beta}$ is yet to be found. To be relevant for turbulence, the central point t_0 of the singularity has to be a variable, whereas the Lamperti framework is for a fixed central time, $t_0 = 0$, of the Mellin functions. Consequently, though a mixture of oscillating functions such as $|t-t_0|^{h+2i\pi\beta}$ may have multifractal properties close to the one measured in turbulence, one lack signal processing tools to inverse the mixture and estimates the various parameters (t_0, h, β) of each object.

Finally, turbulence is an active, challenging and open field with many problems that are interesting from a mathematical, physical or signal processing point of view. This is a subject where one needs to establish fruitful interactions between models, tools of analysis and experimental measurements.

Thanks.

I would like to thank people that helped me by their competence and their willingness to share their knowledge and ideas. Many thanks thus to Olivier Michel, Patrick Flandrin and Pierre-Olivier Amblard, with whom I have the pleasure to work. I am also thankful to C. Baudet, B. Castaing, L. Chevillard, N. Mordant, J.F. Pinton, and J.C. Vassilicos.

References

1. A. Arneodo, E. Bacry, S. Jaffard, J.F. Muzy. Singularity spectrum of multifractal functions involving oscillating singularities. *J. Four. Anal. Appl.*, 4(2):159–174, 1998.
2. F. Auger, P. Flandrin. Improving the readability of time-frequency and time-scale representations by reassignment methods. *IEEE Trans. on Signal Proc. V*, SP-43(5):1068–1089, 1995.
3. P. Abry, P. Flandrin, M. Taqqu, D. Veitch. Wavelets for the analysis, estimation, and synthesis of scaling data. In K. Park and W. Willinger, editors, *Self-Similar Network Traffic and Performance Evaluation*. Wiley, 2000.
4. A. Arneodo, E. Bacry, J.F. Muzy. The thermodynamics of fractals revisited with wavelets. *Physica A*, 213:232–275, 1995.
5. A. Arneodo, J.F. Muzy, S. Roux. Experimental analysis of self-similarity and random cascade processes: applications to fully developped turbulence data. *J. Phys. France II*, 7:363–370, 1997.
6. G. Barenblatt. *Scaling, self-similarity, and intermediate asymptotics*. CUP, Cambridge, 1996.
7. G.K. Batchelor. *The theory of homogeneous turbulence*. Cambridge University Press, 1953.
8. J. Bertrand, P. Bertrand, J.P. Ovarlez. The Mellin transform. In A.D. Poularikas, editor, *The Transforms and Applications Handbook*. CRC Press, 1996.
9. R. Benzi, S. Ciliberto, R. Tripicione, C. Baudet, F. Massaioli. Extended self-similarity in turbulent flows. *Phys. Rev. E*, 48:R29–R32, 1993.
10. J. Beran. *Statistics for Long-memory processes*. Chapman & Hall, New York, 1994.
11. P. Borgnat, P. Flandrin, P.-O. Amblard. Stochastic discrete scale invariance. *Signal Processing Lett.*, 9(6):181–184, June 2002.
12. M. Berry, Z. Lewis. On the Weierstrass-Mandelbrot fractal function. *Proc. Roy. Soc. Lond. A*, 370:459–484, 1980.
13. A. Blanc-Lapierre, R. Fortet. *Théorie des fonctions aléatoires*. Masson, Paris, 1953.
14. J. Barral, B. Mandelbrot. Multifractal products of cylindrical pulses. *Probab. Theory Relat. Fields*, 124:409–430, 2002.
15. E. Bacry, J.F. Muzy. Log-infinitely divisible multifractal processes. *Comm. in Math. Phys.*, 236:449–475, 2003.
16. C. Baudet, O. Michel, W. Williams. Detection of coherent vorticity structures using time-scale resolved acoustic spectroscopy. *Physica D*, 128:1–17, 1999.
17. P. Borgnat. *Modèles et outils pour les invariances d'échelle brisée : variations sur la transformation de Lamperti et contributions aux modèles statistiques de vortexen turbulence*. PhD thesis, École normale supérieure de Lyon, November 2002.
18. B. Castaing. Turbulence: Statistical approach. In B. Dubrulle, F. Graner, and D. Sornette, editors, *Scale Invariance and Beyond*, pages 225–234. Springer, 1997.
19. G. Chavarria, C. Baudet, S. Ciliberto. Hierarchy of the energy dissipation moments in fully developed turbulence. *Phys. Rev. Lett.*, 74:1986–1989, 1995.
20. L. Cohen. The scale representation. *IEEE Trans. on Signal Proc.*, 41(12):3275–3292, December 1993.

21. P. Chainais, R. Riedi, P. Abry. On non scale invariant infinitely divisible cascades. to appear in IEEE Trans. on Info. Theory, 2004.
22. L. Chevillard, S. Roux, E. Lévêque, N. Mordant, J.F. Pinton, A. Arneodo. Lagrangian velocity statistics in turbulent flows: Effects of dissipation. *Phys. Rev. Lett.*, 91:214502, 2003.
23. A. Chhabra, K. Sreenivasan. Scale-invariant multiplier distributions in turbulence. *Phys. Rev. Lett.*, 68(18):2762–2765, 1992.
24. B. Davies. *Integral transforms and their applications.* Springer-Verlag, New York, 1985.
25. S. Douady, Y. Couder, M.E. Brachet. Direct observation of the intermittency of intense vorticity filaments in turbulence. *Phys. Rev. Lett.*, 67:983–986, 1991.
26. P. Flandrin, P. Borgnat. On the chirp decomposition of Weierstrass-Mandelbrot functions, and their time-frequency interpretation. *Applied and Computational Harmonic Analysis*, 15:134–146, September 2003.
27. P. Flandrin, P. Borgnat, P.-O. Amblard. From stationarity to self-similarity, and back : Variations on the Lamperti transformation. In G. Raganjaran and M. Ding, editors, *Processes with Long-Range Correlations: Theory and Applications*, volume 621 of *Lectures Notes in Physics*, pages 88–117. Springer-Verlag, June 2003.
28. P. Flandrin. *Temps-Fréquence (1ère ed.).* Hermes, 1993.
29. P. Flandrin. Inequalities in Mellin-Fourier signal analysis. Newton Institute Preprint NI98030-NSP, Cambridge, UK, November 1998.
30. U. Frisch, R. Morf. Intermittency in nonlinear dynamics and singularities at complex times. *Phys. Rev. A*, 23(5):2673–2705, May 1981.
31. U. Frisch, M. Mineev-Weinstein. Extension of the pole decomposition for the multidimensional Burgers equation. *Phys. Rev. E*, 67:067301, 2003.
32. U. Frisch, G. Parisi. On the singularity structure of fully developped turbulence. In M. Ghil, R. Benzi, and G. Parisi, editors, *Proc. of Int. School of Phys. on Turbulence and predictability in geophysical fluid dynamics*, pages 84–87, Amsterdam, 1985. North-Holland.
33. U. Frisch. *Turbulence.* CUP, Cambridge, 1995.
34. U. Frisch, M. Vergassola. A prediction of the multifractal model: the intermediate dissipative range. *Europhys. Lett.*, 14:429, 1991.
35. Y. Gagne. *Étude expérimentale de l'intermittence et des singularités dans le plan complexe en turbulence développée.* PhD thesis, INP Grenoble, 1987.
36. J.C.R. Hunt, N.K.-R. Kevlahan, J.C. Vassilicos, M. Farge. Wavelets, fractals and fourier transforms: detection and analysis of structures. In M. Farge, J.C.R. Hunt, and J.C. Vassilicos, editors, *Wavelets, Fractals, and Fourier Transforms*, pages 1–38. Oxford : Clarendon Press, 1993.
37. S. Jaffard. Multifractal formalism for functions, part 1 and 2. *SIAM J. of Math. Anal.*, 28(4):944–998, 1997.
38. S. Jaffard. On the Frisch-Parisi conjecture. *J. Math. Pures Appl.*, 79(6):525—552, 2000.
39. S. Jaffard. Wavelet techniques in multifractal analysis. Proceedings of Symposia in Pure Mathematica, 2004.
40. A.N. Kolomogorov. The local structure of turbulence in incompressible viscuous fluid for very large Reynolds numbers. *Dokl. Akad. Nauk SSSR*, 30:9–13, 1941.

41. A.N. Kolomogorov. On degeneration of isotropic turbulence in a incompressible viscuous liquid. *Dokl. Akad. Nauk SSSR*, 31:538–540, 1941.
42. J.-P. Kahane, J. Peyrière. Sur certaines martingales de Benoit Mandelbrot. *Adv. Math.*, 22:131–145, 1976.
43. N. Kevlahan, J.C. Vassilicos. The space and scale dependencies of the self-similar structure of turbulence. *Proc. R. Soc. Lond. A*, 447:341–363, 1994.
44. J. Lamperti. Semi-stable stochastic processes. *Trans. Amer. Math. Soc.*, 104:62–78, 1962.
45. B. Lashermes. PhD thesis, ÉNS Lyon, in preparation, 2004.
46. T.S. Lundgren. Strained spiral vortex model for turbulent fine structure. *Phys. Fluids*, 25(12):2193–2203, 1982.
47. A. La Porta, G.A. Voth, A.M. Crawford, J. Alexander, E. Bodenschatz. Fluid particle accelerations in fully developped turbulence. *Nature*, 409:1017–1019, 2001.
48. S. Mallat. *A Wavelet tour of signal processing.* Academic Press, 1999.
49. Y. Meyer. *Ondelettes et opérateurs.* Hermann, 1990.
50. N. Mordant, P. Metz, O. Michel, J.F. Pinton. Scaling and intermittency of Lagrangian velocity in fully developped turbulence. *Phys. Rev. Lett.*, 87:21–24, 2001.
51. H. Moffatt. Simple topological aspects of turbulent velocity dynamics. In T. Tatsumi, editor, *Turbulence and chaotic phenomena in fluids*, pages 223–230. Elsevier, 1984.
52. H. Moffatt. Spiral structures in turbulent flows. In M. Farge, J.C.R. Hunt, and J.C. Vassilicos, editors, *Wavelets, Fractals, and Fourier Transforms*, pages 317–324. Oxford : Clarendon Press, 1993.
53. C. Meneveau, K.R. Sreenivasan. The multifractal nature of turbulent energy-dissipation. *J. Fluid Mechanics*, 224:429–484, March 1991.
54. J. Mathieu, J. Scott. *An introduction to Turbulent Flow.* Cambridge University Press, 2000.
55. B. Mandelbrot, J. W. Van Ness. Fractional Brownian motions, fractional Brownian noises and applications. *SIAM review*, 10:422–437, 1968.
56. A.S. Monin, A.S. Yaglom. *Statistical fluid mechanics (vol. 1 and 2).* The MIT Press, 1971.
57. E. Noret, M. Guglielmi. Modélisation et synthèse d'une classe de signaux auto-similaires et à mémoire longue. In *Proc. Conf. Delft (NL) : Fractals in Engineering*, pages 301–315. INRIA, 1999.
58. J.M. Nicolas. Introduction aux statistiques de deuxième espèce : Applications aux lois d'images SAR. Rapport interne, ENST, Paris, February 2002.
59. E.A. Novikov. Intermittency and scale-similarity in the structure of a turbulent flow. *P.M.M. Appl. Math. Mech.*, 45:231–241, 1971.
60. C. Nuzman, V. Poor. Transformed spectral analysis of self-similar processes. In *Proc. CISS'99*, May 1999.
61. C. Nuzman, V. Poor. Linear estimation of self-similar processes via Lamperti's transformation. *J. of Applied Probability*, 37(2):429–452, June 2000.
62. C. Poulain, N. Mazelllier, P. Gervais, Y. Gagne, C. Baudet. Lagrangian vorticity and velocity measurements in turbulent jets. *Flow, Turbulence, and Combustion*, 72(2–4):245–271, 2004.
63. G. Pedrizzetti, E. Novikov, A. Prakovsky. Self-similarity and probability distributions of turbulent intermittency. *Physical Review E*, 53(1):475–484, 1996.

64. S. Pietropinto, C. Poulain, C. Baudet, B. Castaing, B. Chabaud, Y. Gagne, B. Hébral, Y. Ladam, P. Lebrun, O. Pirotte, P. Roche. Superconducting instrumentation for high Reynolds turbulence experiments with low temperature gaseous helium. *Physica C*, 386:512–516, 2003.
65. G. Paladin, A. Vulpiani. Degrees of freedom of turbulence. *Phys. Rev. A*, 35:1971, 1987.
66. P.G. Saffman, D.I. Pullin. Calculation of velocity structure fonctions for vortex models of isotropic turbulence. *Phys. Fluids*, 8(11):3072–3077, 1996.
67. G. Samorodnitsky, M. Taqqu. *Stable Non-Gaussian Random Processes.* Chapman & Hall, 1994.
68. W. Vervaat. Properties of general self-similar processes. *Bull. of International Statistical Inst.*, 52:199–216, 1987.
69. E. Waymire, S. Williams. A general decomposition theory for random cascades. *Bull. Amer. Math. Soc.*, 31:216–222, 1994.
70. B. Yazici, R. L. Kashyap. A class of second-order stationary self-similar processes for $1/f$ phenomena. *IEEE Trans. on Signal Proc.*, 45(2):396–410, 1997.
71. A. Zemanian. *Generalized integral transforms.* Dover, 1987.

Control of Interferometric Gravitational Wave Detectors

François Bondu and Jean-Yves Vinet

Laboratoire Artemis
Observatoire de la Côte d'Azur,
BP 4229, 06304 Nice (France).
Francois.Bondu@obs-nice.fr
vinet@obs-nice.fr

1 Introduction

Interferometric gravitational wave detectors are promising instruments to make the first direct detection of gravitational waves, and later to permanently open a new window on the universe [1, 2, 3, 4]. A detector like Virgo aims at observing signals in the 10 Hz - 10 kHz band. The detectors have very strong noise requirements, bringing challenging designs of control loops.

It was directly observed in 1918 that gravity has some effect on light propagation, and in particular is able to bend light rays nearby massive objects like the sun. This effect was predicted by A. Einstein as a consequence of General Relativity, a relativistic theory of gravitation describing gravitational fields as the geometry of space-time. Non static gravitational fields can, in this theory, have some time variable effects on space-time, and be considered as "gravitational waves". Highly energetic astrophysical events are expected to produce such waves, the observation of which would be of the highest interest for our understanding of the Universe. These waves in all theoretical studies, are foreseen very weak (analogous, in an interferometric length measurement, to a distortion of one interferometer arm $\Delta L/L \sim 10^{-22}$ in best cases).

2 Interferometers

Interferometric detection of gravitational waves amounts to continuously measure the length difference between two orthogonal paths. The right topology of the instrument is that of a Michelson interferometer [5]. Virgo is the nearest (Pisa, Italy) example of such an interferometric detector of gravitational waves. It consists essentially of large mirrors suspended by wires in a vacuum, and of light beams partially reflected and/or transmitted by these mirrors over long distances. The right sensitivity is reached by two steps of light

J.-D. Fournier et al. (Eds.): Harm. Analysis and Ratio. Approx., LNCIS 327, pp. 303–311, 2006.

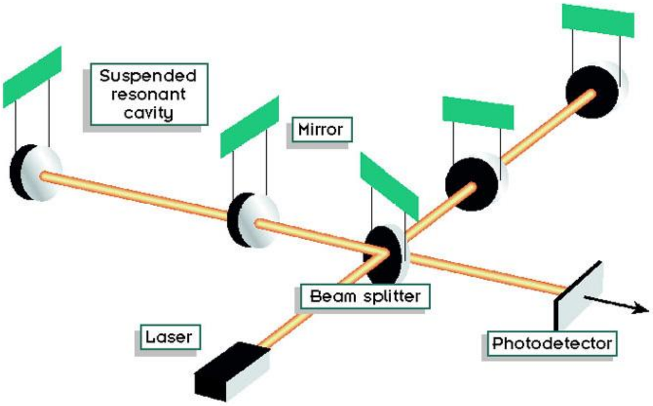


Fig. 1. Principles of an interferometer to detect gravitational waves. A laser beam lights a Michelson interferometer. The mirror suspensions are efficient enough to filter out the seismic noise in the detection band. Resonant Fabry-Perot cavities in the arms allow to enhance the interferometer sensitivity.

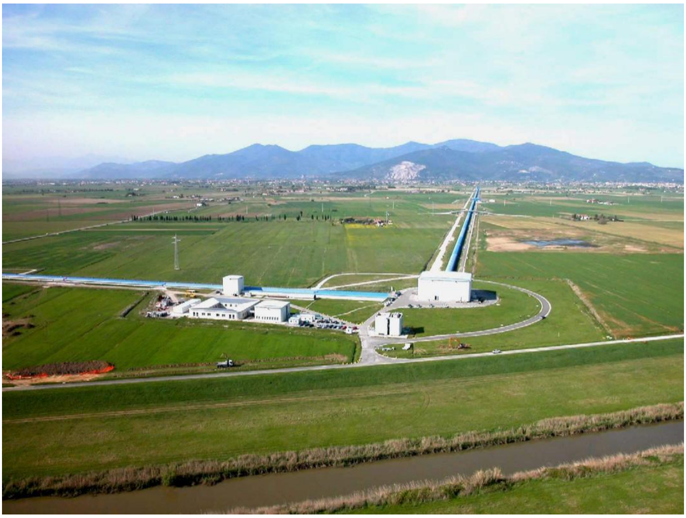


Fig. 2. Aerial view of the Virgo interferometer

power build-up: one step is the resonance caused by the so-called recycling mirror, a second one is due to the use of Fabry-Perot cavities on each arm. Recycling has the effect of increasing the light power, Fabry-Perot's have the effect of enlarging the effective lengths of the arms.

3 Servo systems

3.1 Introduction

To make the complex optical structure of an interferometer work, an ensemble of servo-systems is needed in order to lock the resonant cavities and the laser's frequency at the right place. The design of the open loop transfer functions has to make a trade-off between large gain for frequencies below 1 Hz (where seismic noise is large), stability of the system, and large attenuation of loop gain above 10 Hz, where the aim is to detect gravitational waves. This is the reason why sophisticated servo-loops have been studied for several years.

An interferometer as a complex optical structure

An interferometer to detect gravitational waves, is, today, typically made up of 6 main mirrors. Each arm of the interferometer is made of a long resonant optical cavity, and an additional mirror, in front of the interferometer, helps to build up the light power.

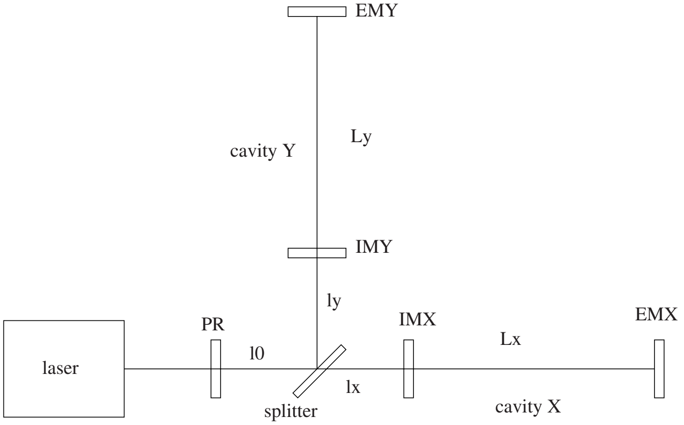


Fig. 3. Main distances of an interferometer to detect gravitational waves

Here we will simplify the analysis to the displacements of the mirrors along the optical axes.

The main output of the interferometer, the so-called “dark fringe” sensitive to gravitational waves, is measuring the variations of the difference of the two cavity lengths $Lx - Ly$; this is the differential mode of the interferometer (the static difference is close to zero). In order for the cavities to work close to their optimal point, both Lx and Ly have to be controlled at the picometer level.

The common mode of the interferometer, $Lx + Ly$, can be used to control the laser frequency fluctuations.

The short Michelson difference, $lx - ly$, and the recycling cavity length, $l0 + (lx + ly)/2$, should be controlled at the order of the picometer as well.

The error signals for these controls are provided by additional outputs of the interferometer (other ports than the dark fringe), and actuation is done by means of currents in coils facing magnets glued on the mirrors.

On these four lengths, only the interferometer differential mode can be monitored with high signal to noise ratio. Thus, on the other lengths, the unity gain and gain loop should be designed so that it does not introduce noise on the dark fringe.

Seismic noise and the suspensions

The spectrum of the seismic noise, in the 10 Hz-10 kHz bandwidth, where one expects to detect gravitational waves, is orders of magnitude higher than what is necessary to detect gravitational waves. Thus, seismic noise isolators are necessary [6, 7].

The suspensions act as low pass filter transfer functions: at the test mass mirror level, the seismic noise is much attenuated for frequencies above 10 Hz. But the seismic noise is amplified on resonances (several resonances in the 0.1 Hz - few Hz band), and still quite high on low frequencies. As a result, the free suspended mirror motion is about 1 μm on a 1 second timescale. There is a need for a loop gain of $\sim 10^6$ below 1 Hz. When controlling the degrees of freedom other than the differential one, the loop gain should be very low for frequencies above 10 Hz, in order to not re-inject the error signal noise.

Laser frequency stabilization

A similar feedback loop issue exists for the laser frequency stabilization [8].

The Virgo instrument requires a very stable laser frequency with a relative level as low as $10^{-20}/\sqrt{\text{Hz}}$, in the 10 Hz – 10 kHz band; the frequency has also to be reasonably stable for frequencies below 10 Hz, so that the Fabry-Perot cavities are kept resonant. The stability is ensured by the quality of the reference oscillator.

The laser frequency in the 10 Hz – 10 kHz band is locked on the common mode of the two long Fabry-Perot cavities. The seismic isolation ensures that this level of stability can be reached.

The low frequency stability is defined by locking the common mode to a rigid and very stable cavity, manufactured in a material with a low expansion thermal coefficient. The spectral resolution of such a cavity is about 4 orders of magnitude higher than the one of the long Fabry-Perot. Therefore, the lock of the common mode of the long arms to the short cavity should have a negligible action at 10 Hz.

Other constraints

Each mirror of the interferometer has also to be controlled in the angular degrees of freedom, with similar, although less stringent, constraints to the longitudinal ones.

3.2 Mathematical requirements on open loop transfer function

The noise requirements on the mirror motions are very strong, and impose the design of efficient corrector filters for the various feedback loops.

Model of servo loops

All servo loops are defined by their frequency domain open loop transfer function. This allows to impose the constraints on the gain at various frequencies, since this is a fundamental issue in the control of the locking point of the resonant cavities, in order not to re-inject the noise of the error signals.

The open loop transfer function G is a complex function of the Laplace variable s, or, equivalently, of a Fourier frequency f. This function is the product of the various elements of the loop: the system to be corrected, including the actuator transfer functions; the error signal transfer function, related to the optical properties; the corrector filter, whose design must satisfy the requirements for the open loop transfer function.

System identification

The error signal transfer function and the actuator transfer function can be easily approximated by rational functions of the variable s. The approximation of resonances, poles, quality factors, etc. is done so that the difference between the modeled transfer function and the measured one is not bigger than a few percent in amplitude and a few degrees in the phase. This is done by manually fitting the measured curves.

The corrector filter is implemented in a DSP by a description of gain, poles and zeros together with their quality factors. Thus, the corrector filter is also a rational fraction of the variable s, with a reasonable order (could be 10, exceptionally 20).

Mathematical description of the requirements

We make the assumption that the error signal and actuator transfer functions are unity at all frequencies. Of course, this is not true, and exact compensation of pole-zero pairs is not possible either. But we assume that we can deal with this problem once corrector filters for unity system transfer functions are available.

The customary variable is the frequency f, with $s = 2\mathrm{i}\pi f$

Ideally, we would like to find a solution for a function G modeled as a rational function of a reasonable order (for example numerator and denominator order not bigger than 20). The constraints become:

$$\forall f < f_1, \qquad |G(f)| > G_1 \tag{1}$$

$$\forall f > f_2, \qquad |G(f)| < G_2 \tag{2}$$

The closed loop should be stable, i.e. $1/(1+G)$ should not have any pole with positive real parts.

The closed loop should be robust. We could define as commonly done gain and phase margins, but unfortunately this can still allow low effective margins for complicated functions. We then require that:

$$\forall f, \qquad |1/(1+G(f))| < k \tag{3}$$

Example of aimed values: $f_1 = 0.1$ Hz, $f_2 = 10$ Hz, $G_1 = 10^6$, $G_2 = 10^{-4}$, $k = 2$ or, even better $f_1 = 1$ Hz, $f_2 = 10$ Hz, $G_1 = 10^4$, $G_2 = 10^{-4}$, $k = 2$.

3.3 The Coulon's solution

At the 'Observatoire de la Côte d'Azur', J.-P. Coulon has built a program to find solutions such that $f_2 = 10$ Hz, $f_1 = 0.1$ Hz or lower. The program actually looks for functions made with a function $G_s(f)$ so that $G(f) = G_s(f)/G_s(1/f)$. $G_s(f)$ is a rational fraction with real coefficients, with m zeroes and n poles.

The program looks over the coefficients of the G_s function, with a successive approximation technique, and some "culinary" recipes to save computing time.

The algorithm allows to find solutions with an order up to 6 zeroes and $n = 8$ poles for the G_s function in a reasonable time. Despite this modest complexity, the solutions given are much better than the "current engineer design".

The open loop transfer function in the Nichols plot helps to check the stability:

Such a filter has been tried successfully, on the real Virgo suspensions.

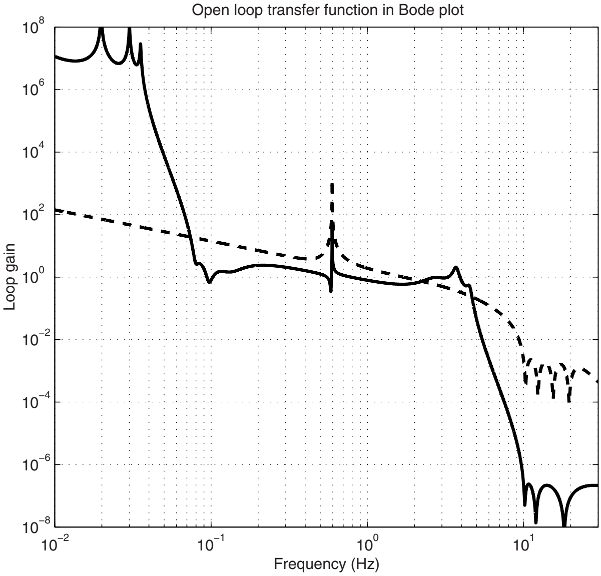


Fig. 4. Two open loop transfer functions, based on the same simplified suspension system (one resonant pole at 0.6 Hz): Coulon's filter (continuous line), "engineer" design (dashed line). The Coulon's filter is defined for $k = 2$, $f_2/f_1 = 278$.

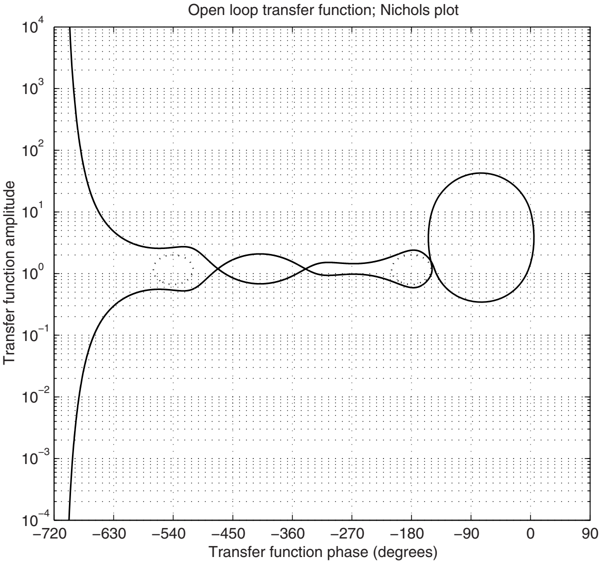


Fig. 5. Coulon's filter used to stabilize a pendulum in the Nichols plot. The dotted circles correspond to a closed loop overshoot of 2 (corresponding to a gain margin of 2 and a phase margin of 30°.)

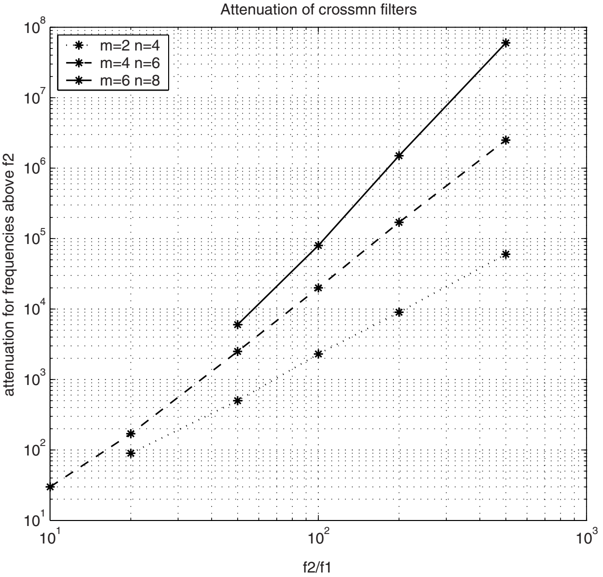


Fig. 6. Performances of various Coulon's filters, with m zeroes and n poles for the $G_s(f)$ function.

The performances of the filters have been computed, depending of the ratio $\frac{f_2}{f_1}$, for various orders of poles and zeroes of the $G_s(f)$ function: The figure 6 seems to indicate that the filter performances will not be very high on a short frequency span $f_2/f_1 = 10$, if one increases the order of the filter.

4 Conclusion and perspectives

The Coulon's filters give very good result, when f_2/f_1 is about two decades. This makes already very good attenuation at 10 Hz (more than 10^6).

Yet, the loop gain in the 0.1 Hz - 1 Hz band is not enough to attenuate the real seismic noise. One would need Coulon-like filters with f_2/f_1 of the order of 10, if possible. New kind of mathematical functions might be investigated to improve the feedback filter performances.

References

1. F. Acernese et al. Status of Virgo. *Class. Quantum Grav.*, 21(5):S385+, March 2004.
2. D. Sigg. Commissioning of Ligo detectors. *Class. Quantum Grav.*, 21(5):S409+, March 2004.

3. B. Willke et al. Status of Geo 600. *Class. Quantum Grav.*, 21(5):S417+, March 2004.
4. R. Takahashi and the TAMA collaboration. Status of Tama 300. *Class. Quantum Grav.*, 21(5):S403+, March 2004.
5. P. Saulson. *Fundamentals of interferometric gravitational wave detectors.* World Scientific Publishing Company, Singapore, 1994.
6. F. Acernese et al. The last stage suspension of the mirrors for the gravitational wave antenna Virgo. *Class. Quantum Grav.*, 21(5):S245+, March 2004.
7. F. Acernese et al. Properties of seismic noise at the Virgo site. *Class. Quantum Grav.*, 21(5):S433+, March 2004.
8. F. Bondu, A. Brillet, F. Cleva, H. Heitmann, M. Loupias, C.N. Man, H. Trinquet, and the VIRGO collaboration. The Virgo injection system. *Class. Quantum Grav.*, 19(7), April 2002.

Lecture Notes in Control and Information Sciences

Edited by M. Thoma and M. Morari

Further volumes of this series can be found on our homepage:
springer.com

Vol. 326: Wang, H.-S.; Yung, C.-F.; Chang, F.-R.
H_∞ Control for Nonlinear Descriptor Systems
164 p. 2006 [1-84628-289-6]

Vol. 325: Amato, F.
Robust Control of Linear Systems Subject to Uncertain Time-Varying Parameters
180 p. 2006 [3-540-23950-2]

Vol. 324: Christofides, P.; El-Farra, N.
Control of Nonlinear and Hybrid Process Systems
446 p. 2005 [3-540-28456-7]

Vol. 323: Bandyopadhyay, B.; Janardhanan, S.
Discrete-time Sliding Mode Control
147 p. 2005 [3-540-28140-1]

Vol. 322: Meurer, T.; Graichen, K.; Gilles, E.D. (Eds.)
Control and Observer Design for Nonlinear Finite and Infinite Dimensional Systems
422 p. 2005 [3-540-27938-5]

Vol. 321: Dayawansa, W.P.; Lindquist, A.; Zhou, Y. (Eds.)
New Directions and Applications in Control Theory
400 p. 2005 [3-540-23953-7]

Vol. 320: Steffen, T.
Control Reconfiguration of Dynamical Systems
290 p. 2005 [3-540-25730-6]

Vol. 319: Hofbaur, M.W.
Hybrid Estimation of Complex Systems
148 p. 2005 [3-540-25727-6]

Vol. 318: Gershon, E.; Shaked, U.; Yaesh, I.
H_∞ Control and Estimation of State-muliplicative Linear Systems
256 p. 2005 [1-85233-997-7]

Vol. 317: Ma, C.; Wonham, M.
Nonblocking Supervisory Control of State Tree Structures
208 p. 2005 [3-540-25069-7]

Vol. 316: Patel, R.V.; Shadpey, F.
Control of Redundant Robot Manipulators
224 p. 2005 [3-540-25071-9]

Vol. 315: Herbordt, W.
Sound Capture for Human/Machine Interfaces: Practical Aspects of Microphone Array Signal Processing
286 p. 2005 [3-540-23954-5]

Vol. 314: Gil', M.I.
Explicit Stability Conditions for Continuous Systems
193 p. 2005 [3-540-23984-7]

Vol. 313: Li, Z.; Soh, Y.; Wen, C.
Switched and Impulsive Systems
277 p. 2005 [3-540-23952-9]

Vol. 312: Henrion, D.; Garulli, A. (Eds.)
Positive Polynomials in Control
313 p. 2005 [3-540-23948-0]

Vol. 311: Lamnabhi-Lagarrigue, F.; Loría, A.; Panteley, E. (Eds.)
Advanced Topics in Control Systems Theory
294 p. 2005 [1-85233-923-3]

Vol. 310: Janczak, A.
Identification of Nonlinear Systems Using Neural Networks and Polynomial Models
197 p. 2005 [3-540-23185-4]

Vol. 309: Kumar, V.; Leonard, N.; Morse, A.S. (Eds.)
Cooperative Control
301 p. 2005 [3-540-22861-6]

Vol. 308: Tarbouriech, S.; Abdallah, C.T.; Chiasson, J. (Eds.)
Advances in Communication Control Networks
358 p. 2005 [3-540-22819-5]

Vol. 307: Kwon, S.J.; Chung, W.K.
Perturbation Compensator based Robust Tracking Control and State Estimation of Mechanical Systems
158 p. 2004 [3-540-22077-1]

Vol. 306: Bien, Z.Z.; Stefanov, D. (Eds.)
Advances in Rehabilitation
472 p. 2004 [3-540-21986-2]

Vol. 305: Nebylov, A.
Ensuring Control Accuracy
256 p. 2004 [3-540-21876-9]

Vol. 304: Margaris, N.I.
Theory of the Non-linear Analog Phase Locked Loop
303 p. 2004 [3-540-21339-2]

Vol. 303: Mahmoud, M.S.
Resilient Control of Uncertain Dynamical Systems
278 p. 2004 [3-540-21351-1]

Vol. 302: Filatov, N.M.; Unbehauen, H.
Adaptive Dual Control: Theory and Applications
237 p. 2004 [3-540-21373-2]

Vol. 301: de Queiroz, M.; Malisoff, M.; Wolenski, P. (Eds.)
Optimal Control, Stabilization and Nonsmooth Analysis
373 p. 2004 [3-540-21330-9]

Vol. 300: Nakamura, M.; Goto, S.; Kyura, N.; Zhang, T.
Mechatronic Servo System Control
Problems in Industries and their Theoretical Solutions
212 p. 2004 [3-540-21096-2]

Vol. 299: Tarn, T.-J.; Chen, S.-B.; Zhou, C. (Eds.)
Robotic Welding, Intelligence and Automation
214 p. 2004 [3-540-20804-6]

Vol. 298: Choi, Y.; Chung, W.K.
PID Trajectory Tracking Control for Mechanical Systems
127 p. 2004 [3-540-20567-5]

Vol. 297: Damm, T.
Rational Matrix Equations in Stochastic Control
219 p. 2004 [3-540-20516-0]

Vol. 296: Matsuo, T.; Hasegawa, Y.
Realization Theory of Discrete-Time Dynamical Systems
235 p. 2003 [3-540-40675-1]

Vol. 295: Kang, W.; Xiao, M.; Borges, C. (Eds)
New Trends in Nonlinear Dynamics and Control,
and their Applications
365 p. 2003 [3-540-10474-0]

Vol. 294: Benvenuti, L.; De Santis, A.; Farina, L. (Eds)
Positive Systems: Theory and Applications (POSTA 2003)
414 p. 2003 [3-540-40342-6]

Vol. 293: Chen, G. and Hill, D.J.
Bifurcation Control
320 p. 2003 [3-540-40341-8]

Vol. 292: Chen, G. and Yu, X.
Chaos Control
380 p. 2003 [3-540-40405-8]

Vol. 291: Xu, J.-X. and Tan, Y.
Linear and Nonlinear Iterative Learning Control
189 p. 2003 [3-540-40173-3]

Vol. 290: Borrelli, F.
Constrained Optimal Control
of Linear and Hybrid Systems
237 p. 2003 [3-540-00257-X]

Vol. 289: Giarré, L. and Bamieh, B.
Multidisciplinary Research in Control
237 p. 2003 [3-540-00917-5]

Vol. 288: Taware, A. and Tao, G.
Control of Sandwich Nonlinear Systems
393 p. 2003 [3-540-44115-8]

Vol. 287: Mahmoud, M.M.; Jiang, J.; Zhang, Y.
Active Fault Tolerant Control Systems
239 p. 2003 [3-540-00318-5]

Vol. 286: Rantzer, A. and Byrnes C.I. (Eds)
Directions in Mathematical Systems
Theory and Optimization
399 p. 2003 [3-540-00065-8]

Vol. 285: Wang, Q.-G.
Decoupling Control
373 p. 2003 [3-540-44128-X]

Vol. 284: Johansson, M.
Piecewise Linear Control Systems
216 p. 2003 [3-540-44124-7]

Vol. 283: Fielding, Ch. et al. (Eds)
Advanced Techniques for Clearance of
Flight Control Laws
480 p. 2003 [3-540-44054-2]

Vol. 282: Schröder, J.
Modelling, State Observation and
Diagnosis of Quantised Systems
368 p. 2003 [3-540-44075-5]

Vol. 281: Zinober A.; Owens D. (Eds)
Nonlinear and Adaptive Control
416 p. 2002 [3-540-43240-X]

Vol. 280: Pasik-Duncan, B. (Ed)
Stochastic Theory and Control
564 p. 2002 [3-540-43777-0]

Vol. 279: Engell, S.; Frehse, G.; Schnieder, E. (Eds)
Modelling, Analysis, and Design of Hybrid Systems
516 p. 2002 [3-540-43812-2]

Vol. 278: Chunling D. and Lihua X. (Eds)
H_∞ Control and Filtering of
Two-dimensional Systems
161 p. 2002 [3-540-43329-5]

Vol. 277: Sasane, A.
Hankel Norm Approximation
for Infinite-Dimensional Systems
150 p. 2002 [3-540-43327-9]

Vol. 276: Bubnicki, Z.
Uncertain Logics, Variables and Systems
142 p. 2002 [3-540-43235-3]

Vol. 275: Ishii, H.; Francis, B.A.
Limited Data Rate in Control Systems with Networks
171 p. 2002 [3-540-43237-X]

Vol. 274: Yu, X.; Xu, J.-X. (Eds)
Variable Structure Systems:
Towards the 21$^{\rm st}$ Century
420 p. 2002 [3-540-42965-4]

Vol. 273: Colonius, F.; Grüne, L. (Eds)
Dynamics, Bifurcations, and Control
312 p. 2002 [3-540-42560-9]

Vol. 272: Yang, T.
Impulsive Control Theory
363 p. 2001 [3-540-42296-X]

Vol. 271: Rus, D.; Singh, S.
Experimental Robotics VII
585 p. 2001 [3-540-42104-1]

Vol. 270: Nicosia, S. et al.
RAMSETE
294 p. 2001 [3-540-42090-8]

Vol. 269: Niculescu, S.-I.
Delay Effects on Stability
400 p. 2001 [1-85233-291-316]