Asymptotic Combinatorics with Application to Mathematical Physics

Edited by

Vadim Malyshev and Anatoly Vershik

NATO Science Series

II. Mathematics, Physics and Chemistry – Vol. 77

Asymptotic Combinatorics with Application to Mathematical Physics

NATO Science Series

A Series presenting the results of scientific meetings supported under the NATO Science Programme.

The Series is published by IOS Press, Amsterdam, and Kluwer Academic Publishers in conjunction with the NATO Scientific Affairs Division

Sub-Series

I. Life and Behavioural Sciences	IOS Press
II. Mathematics, Physics and Chemistry	Kluwer Academic Publishers
III. Computer and Systems Science	IOS Press
IV. Earth and Environmental Sciences	Kluwer Academic Publishers
V. Science and Technology Policy	IOS Press

The NATO Science Series continues the series of books published formerly as the NATO ASI Series.

The NATO Science Programme offers support for collaboration in civil science between scientists of countries of the Euro-Atlantic Partnership Council. The types of scientific meeting generally supported are “Advanced Study Institutes” and “Advanced Research Workshops”, although other types of meeting are supported from time to time. The NATO Science Series collects together the results of these meetings. The meetings are co-organized bij scientists from NATO countries and scientists from NATO’s Partner countries – countries of the CIS and Central and Eastern Europe.

Advanced Study Institutes are high-level tutorial courses offering in-depth study of latest advances in a field.
Advanced Research Workshops are expert meetings aimed at critical assessment of a field, and identification of directions for future action.

As a consequence of the restructuring of the NATO Science Programme in 1999, the NATO Science Series has been re-organised and there are currently Five Sub-series as noted above. Please consult the following web sites for information on previous volumes published in the Series, as well as details of earlier Sub-series.

http://www.nato.int/science
http://www.wkap.nl
http://www.iospress.nl
http://www.wtv-books.de/nato-pco.htm

Series II: Mathematics, Physics and Chemistry – Vol. 77

Asymptotic Combinatorics with Application to Mathematical Physics

edited by

Vadim Malyshev

I.N.R.I.A.,
Le Chesnay, France

and

Anatoly Vershik

Steklov Mathematical Institute of Russian Academy of Sciences,
St. Petersburg, Russia

Springer Science+Business Media, B.V.

Proceedings of the NATO Advanced Study Institute on
Asymptotic Combinatorics with Application to Mathematical Physics
St. Petersburg, Russia
9–22 July 2001

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN 978-1-4020-0793-4 ISBN 978-94-010-0575-3 (eBook)
DOI 10.1007/978-94-010-0575-3

Printed on acid-free paper

Contents

PREFACE

This volume presents some courses and series of lectures which were organized in the framework of NATO Advanced Study Institute under the title

"Asymptotic Combinatorics with Application to Mathematical Physics"

at Euler International Mathematical Institute at St. Petersburg, July 2001. At the same time it was European Summer School as a part of the activity of European Mathematical Society. The idea of organizers was to invite the best specialists in this area to give a series of lectures for young mathematicians as well as for specialisits who are interested in recent progress in wide part of mathematics including asymptotic methods in mathematical physics, combinatorics, representation theory and some applications. One of the main points in the organization of the School was to invite both mathematicians and physicists in order to emphsize deep connections and interrelation between various approaches to the common subjects. This volume includes some courses and seminar talks which are more concerned to mathematical physics.

The areas of mathematics and mathematical physics which were the subject of the meeting have been studied very intensively last years. This intensive investigations brought about a series of striking results which could be considered together as a new higher level of understanding of related problems: theory of integrable systems, Riemann–Hilbert problem, asymptotic representation theory, spectrum of random matrices, combinatorics of Young diagrams and permutations and even some aspects of quantum field theory. The initial idea to get leading specialists in all these areas together in order to give short courses of lectures on these subjects and to attract attention to recent progress, and especially to attract young mathematicians to those areas, had met an active reaction of the specialists, and we hope it ended up being a fruitful contact between asymptotic combinatorics and mathematical physics. The list of lecturers included the pioneers in these areas — E. Brezin, P. Deift, L. Faddeev, V. Kazakov, S. Novikov, et al).

The methods of theory of integrable systems, matrix problems and theory of Riemann–Hilbert problem together with asymptotic combinatorics and representation theory give very powerful tools for solving numerous old problems: distribution of the fluctuations of the eighenvalues of random matrices, counting the number of coverings of algebraic curves, universality of the distribution of the spacing and other statistical characteristics of Young diagrams, etc. It make sense to mention the lectures and courses about spectacular new results in mathematical physics (P. Deift, L. Pastur), matrix problems (E. Brezin, V. Kazakov), applications of combinatorics (A. Borodin, R. Kenyon), representation theory (A. Vershik, G. Olshansky) algebraic geometry (A. Okounkov), probability theory of maps (V. Malyshev) as well as other distinguished lectures (L. Faddeev, S. Novikov, I. Krichever, V. Korenpin, J. Jacobsen et al)

We do not want to comment each paper in this volume but nevertheless emphasize that combinatorial and probabilistic methods started to play an important role in mathematical physics. Random matrix models, combinatorics of maps, and Young diagrams as well as traditional methods of generating functions and related part of analysis become more and more popular and powerful in applications to mathematical physics. Simultaneously these needs give a serious additional impulse for the pure mathematical theory of these objects. We can watch also a fast developing of nontraditional probabilistic theories based on new achievements. All this shows the fruitful interrelation between these new parts of mathematics and physics.

The program of the school was even overfull with additional talks (besides lectures). The Round table on the current problems and some special session with the questions and answers for young people and short discussions about their studies were organized during the school. We decided to include in this volume some papers of authors who unfortunately could not attend to the meeting but have sent us their papers (V. Maslov, D. Aldous and J. Pitman etc).

The NATO ASI — EMS School was organized by International Euler Mathematical Institute, St. Petersburg branch of Mathematical Institute of Russian Academy of Sciences (POMI RAN) and St. Petersburg Mathematical Society. The preparation of the School started at 1999; application to NATO was sent in July 2000 when the School was already in preparation as European Summer School in the framework of Summer schools of the European Mathematical Society. The support which came from the NATO Scientific Affairs Division was extremely important. Without this support the School could not be so representative and successful. Below we put the more details on the school.

V. A. MALYSHEV
A. M. VERSHIK

Asymptotic combinatorics with application to mathematical physics

PROGRAM

July 8–20

July 8, Sunday (Euler International Mathematical Institute, Pesochnaya nab., 10)
15.00–20.00 REGISTRATION

1st day, July 9, Monday (Steklov Mathematical Institute, Fontanka, 27)
08.30–09.45 REGISTRATION
10.00–10.30 Opening Session
10.30–11.30 ***Brezin E.*** An introduction to matrix models – 1
12.00–13.00 ***Vershik A.*** Introduction to asymptotic theory of representations – 1
15.00–15.50 ***Korepin V.*** Quantum spin chains and Riemann Zeta function with odd arguments
16.10–17.00 ***Bozejko M.*** Positive definite functions on Coxeter groups and second quantization of Yang–Baxter type

2nd day, July 10, Tuesday (Euler International Mathematical Institute, Pesochnaya nab., 10)
9.30–10.30 ***Ol'shanski G.*** Harmonic analysis on big groups, and determinantal point processes – 1
10.40–11.40 ***Brezin E.*** An introduction to matrix models – 2
12.10–13.10 ***Malyshev V.*** Combinatorics and probability for maps on two dimensional surfaces
15.00–15.50 ***Hora A.*** An algebraic and combinatorial approach to central limit theorems related to discrete Laplacians
16.00–16.50 ***Nazarov M.*** On the Frobenius rank of a skew Young diagram
17.20–18.10 ***Kenyon R.*** Hyperbolic geometry and the low-temperature expansion of the Wulff shape in the $3D$ Ising model
19.40 ***Brezin E.*** Informal discussion on matrix models

3rd day, July 11, Wednesday
9.30–10.30 ***Okounkov A.*** Combinatorics and moduli spaces of curves – 1
10.40–11.40 ***Brezin E.*** An introduction to matrix models – 3
12.10–13.10 ***Biane Ph.*** Asymptotics of representations of symmetric groups, random matrices and free cumulants
14.00–17.00 Excursion over St. Petersburg
19.00 Mariinski Theatre

4th day, July 12, Thursday
9.30–10.30 ***Deift P.*** Random matrix theory and combinatorics: a Riemann–Hilbert approach – 1
10.40–11.40 ***Okounkov A.*** Combinatorics and moduli spaces of curves – 2

12.10–13.10 ***Kazakov V.*** Matrix quantum mechanics and statistical physics on planar graphs – 1
15.00–15.50 ***Faddeev L.*** 3-dimensional solitons and knots
16.30–18.10 ***Krichever I.*** τ-functions of conformal maps
19.30 Boat trip

5th day, July 13, Friday
9.30–10.30 ***Deift P.*** Random matrix theory and combinatorics: a Riemann–Hilbert approach – 2
10.40–11.40 ***Okounkov A.*** Combinatorics and moduli spaces of curves – 3
12.10–13.10 ***Ol'shanski G.*** Harmonic analysis on big groups, and determinantal point processes – 2
15.00–15.50 ***Kazakov V.*** Matrix quantum mechanics and statistical physics on planar graphs – 2
16.00–16.50 Round table on combinatorics of the configurations and limit shapes
17.20–18.10 ***Liskovets V.*** Some asymptotic distribution patterns for planar maps

July 14 Excursions, Museums, etc.

July 15 Excursion to Petergof 10:00 (from Euler Institute)

6th day, July 16, Monday
9.30–10.30 ***Deift P.*** Random matrix theory and combinatorics: a Riemann–Hilbert approach – 3
10.40–11.30 ***Kazakov V.*** Combinatorics of planar graphs in matrix quantum mechanics – 3
12.10–13.10 ***Borodin A.*** Asymptotic representation theory and Riemann–Hilbert problem – 1
15.00–15.50 ***Novikov S.*** On the weakly nonlocal Poisson and Symplectic Structures
16.00–16.50 ***Spiridonov V.*** Special functions of hypergeometric type associated with elliptic beta integrals
17.20–18.10 ***Missarov M.*** Exactly solvable renormalization group model

7th day, July 17, Tuesday
9.30–10.30 ***Deift P.*** Random matrix theory and combinatorics: a Riemann–Hilbert approach – 4
10.40–11.40 ***Borodin A.*** Asymptotic representation theory and Riemann–Hilbert problem – 2
12.10–13.10 ***Okounkov A.*** Combinatorics and moduli spaces of curves – 4
15.00–15.50 ***Speicher R.*** Free probability and Random matrices
16.00–16.50 ***Korotkin D.*** Riemann–Hilbert problems related to branched coverings of GP^1, τ-function and Liouville action
17.20–18.10 ***Litvinov G.*** Representation theory in Idempotent (asymptotic) Mathematics

8th day, July 18, Wednesday

9.30–10.30 ***Smirnov S.*** Critical percolation is conformally invariant – 1

10.40–11.40 ***Pastur L.*** Eigenvalue distribution of unitary invariant ensembles of random matrices of large order – 1

12.10–13.10 ***Pevzner M.*** On tensor products and Beresin kernels

15.00–17.00 ***Memorial session devoted to Sergei Kerov and Anatoly Izergin***

9th day, July 19, Thursday

9.30–10.30 ***Speicher R.*** Free probability and Random matrices – 2

10.40–11.40 ***Bozejko M.*** White noise associated to the characters of the infinite symmetric group — Hopf–Kerov deformation

12.10–13.00 ***Jacobsen J. L.*** Enumerating coloured tangles

13.05–13.35 ***Sniady P.*** Random matrices and free probability

15.00–15.30 ***Mlotkowski W.*** Λ-free probability

15.30–16.00 ***Petrogradskij V.*** Asymptotical theory of infinite dimensional Lie algebras

16.30–17.00 ***Kuznetsov V.*** On explicit formulae for special Macdonald polynomials

17.05–17.35 ***Dubrovskiy S.*** Moduli space of symmetric connections

17.40–18.10 ***Stukopin V.*** Representations of Yangians of Lie Superalgebras $A(m,n)$ type

10th day, July 20, Friday

9.30–10.30 ***Pastur L.*** Eigenvalue distribution of unitary invariant ensembles of random matrices of large order – 2

10.40–11.10 ***Vershik A.*** Introduction to asymptotic theory of representations – 2

11.15–11.45 ***Yambartsev A.*** Two-dimensional Lorentzian Models

12.15–13.00 ***Round table***: Problems of the theory of integrable operators and determinant processes

13.00–13.30 ***Closing the school***

List of participants

Name	Affiliation	email
Ahmedov Ruslan	Moscow State University, Russia	ahmedov@mccme.ru
Alexandrov Sergej	SPhT, France	alexand@spht.saclay.cea.fr
Aptekarev Alexander	Keldysh Institute, Russia	aptekaa@spp.keldysh.ru
Balog Jozsef	University of Memphis, USA	balogj@msci.memphis.edu
Batchourine Pavel	Moscow State University, Russia	bachurin@ipac.ac.ru
Belogrudov Alexander	Inst. of Mathematics, Ufa Science Centre of RAS; Russia	belan@anrb.ru
Beliaev Dmitri	KTH, Sweden	beliaev@math.kth.se
Biane Philippe	CNRS, ENS, DMA, France	Philippe.Biane@ens.fr
Bloznelis Mindaugas	Vilnius University, Lithuania	mblozn@ieva.maf.vu.lt
Borodin Alexey	University of Pennsylvania, USA	borodine@math.upenn.edu
Bozejko Marek	Instytut Matematyczny, Wroclaw University, Poland	bozejko@math.uni.wroc.pl
Brezin Edouard	Ecole Normale Superiore, France	brezin@corto.lpt.ens.fr
Calkin Neil	Clemson University; USA	calkin@math.clemson.edu
Carlet Guido	SISSA, Italy	carlet@sissa.it
Chebalov Sergej	UIUC, USA	chebalov@math.uiuc.edu
Chou Ching-Sung	Dept of Math, National Central University, Taiwan, R.O.C.	chou@math.ncu.edu.tw
Cislo Jerzy	Univ of Wroclaw, Institute of Theoretical Physics, Poland	cislo@ift.uni.wroc.pl
Cutler Johnatan	University of Memphis, USA	cutlerj@msci.memphis.edu
Deift Percy	Courant Institute and University of Pennsylvania, USA	deift@math1.cims.nyu.edu
Dobrynin Sergei	St. Petersburg State University, Russia	
Dubrovskiy Stanislav	Northeastern University, USA	dubr@research.neu.edu
Duzhin Sergei	Steklov Institute of Mathematics at St. Petersburg (POMI), Russia	duzhin@pdmi.ras.ru
Erschler Anna	Steklov Institute of Mathematics at St. Petersburg (POMI), Russia	erschler@pdmi.ras.ru
Faddeev Ludvig	Steklov Institute of Mathematics at St. Petersburg (POMI), Russia	faddeev@pdmi.ras.ru
Fedotov Sergej	UMIST, UK	sergei.fedotov@umist.ac.uk
Galkin Sergej	Moscow Independent University, Russia	
Gioev Dimitri	KTH, Sweden	
Gorbulski Alexander	St. Petersburg State University, Russia	alexander@MG4812.spb.edu
Gorbulski Mikhail	St. Petersburg State University, Russia	mike@MG4812.spb.edu
Gordin Mikhail	Steklov Institute of Mathematics at St. Petersburg (POMI), Russia	gordin@pdmi.ras.ru
Grebenkov Denis	Ecole Normale Superiore, France	dg@pmc.polytechnique.fr
Gurin Alexey	ILT NAF Ukraina, Ukraine	gurin@ilt.kharkov.ua
Hora Akihito	Okayama University, Japan	hora@ems.okayama-u.ac.jp
Igonine Serguei	University of Twente, the Netherlands	igonin@mccme.ru
Jacobsen Jesper Lykke	Lab de Physique Theorique et Modeles Statistiques, France	jacobsen@ipno.in2p3.fr
Kalita Eugen	Inst. of Applied Mathematics and Mechanics NAS Ukraine	kalita@iamm.ac.donetsk.ua
Kazakov Vladimir	Ecole Normale Superiore, France	Vladimir.Kazakov@lpt.ens.fr
Kenyon Richard	CNRS UMR 8628, France	kenyon@topo.math.u-psud.fr
Khimshiashvili Giorgi	A.Razmadze Mathematical Institute, Georgia	khimsh@rmi.acnet.ge

Name	Affiliation	email
Kirillov Alexander	SUNY at Stony Brook, USA	kirillov@math.sunysb.edu
Kokhas Konstantin	St. Petersburg State University, Russia	kostik@KK1437.spb.edu
Korepin Vladimir	YITP, State University of New York at Stony Brook, USA	korepin@insti.physics.sunysb.edu
Krichever Igor		
Krikun Maxim	Lab of Large Random Systems, Moscow State Univ, Russia	
Kuznetcov Vadim	University of Leeds; UK	vadim@amsta.leeds.ac.uk
Larson Craig	University of Houston; USA	clarson@math.uh.edu
Liskovets Valerij	Institute of Mathematics of NAS of Belarus	liskov@im.bas-net.by
Litvinov Grigori	International Sophus Lie Centre, Russia	islc@dol.ru
Lodkin Andrei	St. Petersburg State University, Russia	lodkin@pdmi.ras.ru
Lucena Brian	Brown University: Division of Applied Math, USA	lucena@cfm.brown.edu
Maida Mylene	ENS, France	maida@clipper.ens.fr
Malyshev Vadim	I.N.R.I.A., France	Vadim.Malyshev@inria.fr
Marciniak Macin	Institute of Mathematics, Gdansk University, Poland	matmm@univ.gda.pl
Mattera Massimiliano	Universite Paris Sud, Laboratoire de mathematiques France	Massimiliano.Mattera@math.u-psud.fr
Matveev Vladimir	Steklov Institute of Mathematics at St. Petersburg (POMI), Russia	matveev@u-bourgogne.fr
Missarov Moukadas	Kazan State University, Department of Applied Mathematics Laboratory of Probability Theory; Russia	Moukadas.Missarov@ksu.ru
Mlotkowski Wojcieh	Mathematical Institute, The University of Wroclaw, Poland	mlotkow@math.uni.wroc.pl
Mnev Nikolai	Steklov Institute of Mathematics at St. Petersburg (POMI), Russia	mnev@pdmi.ras.ru
Momar Dieng	University of California, Davis; USA	momar@math.ucdavis.edu
Nazarov Maxim	Department of Mathematics, University of York, UK	mln1@york.ac.uk
Nikitin Pavel	St. Petersburg State University, Russia	pasha@AN4760.spb.edu
Novikov Sergej	Landau Institute RAN, Moscow, Russia	
Okounkov Andrej	Univ of California at Berkeley, Dept of Mathematics, USA	okounkov@math.berkeley.edu
Olshanlki Grigori	Institute for Information Transmission Problems, Russia	olsh@online.ru
Panov Dmitri	Ecole Polytchnique, CMAT, France	panov@ihes.fr
Pastur Leonid	CPT, France/FTINT NANU, Ukraine; France	pastur@cpt.univ=mrs.fr
Petrogradskij Viktor	Ulyanovsk State University, Russia	vmp@mmf.ulsu.ru
Pevzner Micha	ULB; Belgium	mpevzner@ulb.ac.be
Romanovski Yuri	St. Petersburg State University, Russia	
Rider Brian	Dept of E.E., The Technion, Israel	rider@ee.technion.ac.il

Name	Affiliation	email
Sniady Piotr	Instytut Matematyczny, Uniwersytet Wroclawski, Poland	Piotr.Sniady@math.uni.wroc.pl
Speicher Roland	Queen's University Jeffery Hall, Canada	speicher@mast.queensu.ca
Spiridonov Vyacheslav	Joint Institute for Nuclear Research; Dubna, Russia	svp@thsun1.jinr.ru
Spitz Grigory	International Sophus Lie Centre, Moscow, Russia	
Starr Shannon	U.C. Davis, USA	sstarr@math.ucdavis.edu
Stenlund Mikko	Department of Mathematics, University of Helsinki, Finland	mikko.stenlund@helsinki.fi
Stevens Laura	UNC Chapel Hill, USA	laura@jalapeno.math.unc.edu
Strahov Evgeny	Brunel University, UK	strahov@physics.technion.ac.il
Stukopin Vladimir	Don State Technical University, Russia	stukopin@math.rsu.ru
Suidan Toufic	Princeton University; USA	tmsuidan@math.princeton.edu
Szenes Andras	MIT, USA	szenes@math.mit.edu
Tarasov Vitaly	Steklov Institute of Mathematics at St. Petersburg (POMI), Russia	tarasov@pdmi.ras.ru
Terilla John	UNC, USA	terillaj@wlu.edu
Tsilevich Natalia	St. Petersburg State University, Russia	natalia@pdmi.ras.ru
Vasiliev Vladimir	VNIPIET, Russia	
Vershik Anatoly	Steklov Institute of Mathematics at St. Petersburg (POMI), Russia	Vershik@pdmi.ras.ru
Wojakowski Lukazc	University of Wroclaw, Mathematical Institute, Poland	Lukasz.Wojakowski@math.uni.wroc.pl
Yakubovich Yuri	Steklov Institute of Mathematics at St. Petersburg (POMI), Russia	yuyakub@mail.ru
Yambartsev Anatoly	Moscow State University, Russia	yambar@lbss.math.msu.su
Yakobson Dmitri		
Zamyatin Andrej	Lab. of Large Random Systems, Moscow State Univ., Russia.	zamyatin@lbss.math.msu.su
Zinn-Justin Paul	LPTMS, University Paris-Sud, ORSAY, FRANCE	Paul.Zinn-Justin@lpt.ens.fr
Zvonarev Mikhail	Steklov Institute of Mathematics at St. Petersburg (POMI), Russia	zvonarev@pdmi.ras.ru
Zvonkine Dimitri	Ecole Normale Superieure, France	zvonkine@clipper.ens.fr

Part ONE
Matrix Models and Graph Enumeration

MATRIX QUANTUM MECHANICS

Lectures at the European Summer School "Asymptotic combinatorics with applications to mathematical physics", St.Petersburg (Russia), July 2001

VLADIMIR KAZAKOV (kazakov@physique.ens.fr)
Laboratoire de Physique Théorique
l'Ecole Normale Supérieure
75231 Paris CEDEX
France

Abstract. We give an introduction into the basic features and applications of a quantum mechanical "particle" with the Hermitian matrix valued coordinate. The system enjoys certain integrability properties which allows exact analytical calculations of some interesting physical quantities and counting of planar graphs embedded into the one dimensional line or the circle.

1. Introduction

The matrix quantum mechanics (MQM) is defined by the Hamiltonian $\hat{H}_M = -\Delta_M + \mathrm{Tr}\, V(M)$, where Δ_M is the standard $SU(N)$ invariant Laplacian on the homogeneous space of Hermitian $N \times N$ matrices $M_{ij}(t)$, $i, j = 1, \ldots, N$ and the potential $V(M)$ is usually a function having a regular expansion around $M = 0$. In the functional integral version (in imaginary, or Euclidean, time) this model is described by the Feynman functional integral over the matrix coordinates $M(t)$.

The MQM attracted a considerable attention of physicists starting from the seminal work [6] where its ground state was found exactly in the limit of the infinite size N of the matrix. This so called 't Hooft limit appears to describe an interesting problem of graph combinatorics: "counting" of graphs with a given two-dimensional topology and having special weights with respect to the "time" coordinates of their vertices. This combinatorial interpretation of the model leads to important results in the two-dimensional string theory and statistical mechanics of the bosonic field on random lattices, including the statistics of vortices in case of the compactified bosonic field. From the mathematical point of view, the MQM enjoys interesting integrability properties: its partition function appears to be, under certain circumstances, a τ-function of the Toda lattice hierarchy. MQM has also interesting applications in the mesoscopic physics. When the size of the matrix $N \to \infty$ the system has an infinite number of degrees of freedom and can

V. A. Malyshev and A. M. Vershik (eds.),
Asymptotic Combinatorics with Application to Mathematical Physics, 3–21.

exhibit the phenomena typical for large statistical-mechanical systems like phase transitions, spontaneous symmetry breakdown etc.

The model has an obvious $SU(N)$ symmetry property with respect to the rotations $M \to \Omega^+ M \Omega$, where Ω is an $SU(N)$ rotation matrix. Hence one can consider the Gibbs partition function of the MQM at finite temperature, defined as

$$Z_N(\beta, \Omega) = \mathrm{Tr}[e^{-\beta \hat{H}_M} \hat{\Theta}] \tag{1}$$

where the twist operator $\hat{\Theta}$ acts on the wave function as the $SU(N)$ similarity transformation of the matrix argument: $\hat{\Theta}\Psi(M) = \Psi(\Omega^+ M \Omega)$.

In the simplest case of singlet representation and infinite time interval (solved long ago by E. Brezin et al [6]) the wave function is an $SU(N)$ scalar and thus a function only of N eigenvalues of M. The problem appears to be equivalent to the quantum mechanical system of N non-interacting fermions in the potential $V(x)$. The eigenfunctions are Slater determinants of the wave functions of individual fermions.

As in the case of the one matrix model (see the lectures of E. Brezin) the expansion with respect to the non-Gaussian terms of the potential $V(M)$ can be represented in terms of sums over Feynman graphs classified according to their two-dimensional topologies due to the weight N^{2-2h}, where h is the genus of a graph, arising from the matrix structure of the model. So in the large N limit only the planar graphs (with the topology of a sphere for the free energy of the model) survive. But in contrast to the one matrix model each propagator of a graph is weighted with the weight $\exp[-|t_v - t_{v'}|]$, where t_i and t_j are the time coordinates of the neihbouring vertices v, v' of the graph, and the integration over each coordinate is done along the real axis. This picture corresponds to the singlet representation sector of the model with respect to the $SU(N)$.

The situation when the fermions fill a metastable minimum of the potential up to the top (local maximum) corresponds to the instability of the system related to the dominance of graphs of a very big size. This picture allows to define in a simple way the asymptotic combinatorics of big planar Feynman graphs (with the topology of a sphere) embedded into the one dimensional space (time), as was demonstrated in [4].

In the case of a finite temperature the weights of propagators are periodic in the imaginary time interval β. The sum over Feynman graphs describes the statistical mechanical model of Berezinski–Kosterlitz–Thouless (BKT) vortices on random planar lattices [5], [3]. The twisted Gibbs partition function already takes into account all sectors (corresponding to the particular $SU(N)$ representations for the wave functions) of the Hilbert space of the model. We will see that in the double scaling limit (DSL) dominated by large Feynman graphs this model enjous the Toda type integrability which allows to describe a new interesting physics of BKT vortices on a random dynamical surface.

In the following sections, we will try to give a brief introduction into the methods adn results concerning this model of MQM.

2. Combinatorics of planar graphs and the singlet sector in matrix quantum mechanics

The model is defined by the partition function defined as the expectation value of the time evolution operator or as a quantum mechanical path integral

$$\zeta_N(g) =< 0|e^{-\beta \hat{H}}|0 >= \int D^{N^2}\ M(t) \exp -\frac{N}{g^2} \operatorname{Tr} \int_0^\beta \ dt[\frac{1}{2}\dot{M}^2 + V(M)] \qquad (2)$$

ou $V(M) = \frac{1}{2}M^2 - \frac{1}{3}M^3$.

First, we will consider the zero temperature limit $\beta \to \infty$ when the partition function is reduced to

$$\zeta \simeq e^{-\beta E_0}$$

where E_0 is the ground state energy of the system (the free energy per time unit on the infinite euclidean "time" line)

As usually, the matrix structure allows to consider the sums over Feynman graphs of fixed topologies (planar graphs). The propagators drown by double lines (each single line conserves the corresponding index, which is reflected in the Kronecker δ-simbols)

$$t \ {}^{i}_{k} \overline{\overline{\qquad\qquad}}\ {}^{j}_{l} \ \ t' \ = e^{-|t-t'|}\ \delta_{ij}\delta_{kl}$$

already depend on the time moments t and t' attributed to the vertices which they connect. The Feynman rules impose that these correlators connect the cubic vertices in all possible ways, giving rise to the so called "fat" graphs. Various contractions of the Kronecker simbols give N-dependent factors. Their counting is absolutly identical to the one-matrix model considered in the lectures of E. Brezin in this volume. The time dependence is the only thing which is different in our actual model.

The partition function has the following expansion with respect to g et $1/N$:

$$\frac{1}{N^2} \log \zeta_N(g) = N^2 \sum_{h=0}^{\infty} N^{-2h} \sum_{n=1}^{\infty} g^{2n} \sum_{G_n^{(h)}} \int dt_1 \cdots dt_n \ e^{-\sum_{<ab>_G} |t_a - t_b|} \qquad (3)$$

Note that in the one-dimensional case there are no UV divergences for $t_a \sim t_b$. For $g \to g_c$ the functional integral is dominated by large dynamical graphs ("random surfaces") projected onto one dimensional line.

We are going to solve this matrix model and extract from the solution the universal critical properties of large random surfaces with fixed topology.

The Schroedinger equation for the wave function $\Psi(M)$ of the MQM takes the form

$$\hat{H}\Psi = E\Psi$$

where $H = -\frac{1}{2N^2}\Delta_M + V(M)$, ou Δ_M is the laplacian on the hermitian matrices. Write the Schroedinger equation in the variational form:

$$E(g) = \min_{\Psi} \frac{\int dM \ [\frac{1}{2}|\frac{\partial}{\partial M}\Psi|^2 + V(M)|\Psi|^2]}{\int dM \ |\Psi|^2} \qquad (4)$$

Diagonalizing the matrix field (introducing the generalized "polar" coordinates) $M \simeq \Omega^+ x \Omega$, $\Omega \simeq 1 + d\omega$ we have for any wave function:

$$\partial_{M_{ij}} \Psi = \delta_{ij} \partial_{x_i} \Psi + \sum_{i \neq j} \frac{\partial_{\omega_{ij}} \Psi}{x_i - x_j} = \delta_{ij} \partial_{x_i} \Psi. \tag{5}$$

For the singlet state the wave function is a symmetric function only of the eigenvalues $\Psi(M) = \Psi(x_1, \cdots, x_N) := \Psi(x)$ and the eq. (5) reduces to $\partial_{M_{ij}} \Psi = \delta_{ij} \partial_{x_i} \Psi(x)$. It is natural to introduce a totally antisymmetric function $\Phi(x) = \frac{1}{\Delta(x)} \Psi(x)$. For the first term in (4) we get by integrating by parts

$$\int dM \left| \frac{\partial}{\partial M} \Psi \right|^2 = \int \prod_k dx_k \Delta^2(x) \sum_i \left| \frac{\partial}{\partial x_i} \frac{\Phi}{\Delta(x)} \right|^2 = \int \prod_k dx_k \sum_j |\partial_j \Phi|^2$$
$$+ \int \prod_k dx_k |\Phi|^2 \left[\sum_k \sum_{i,j} \frac{1}{(x_k - x_i)(x_k - x_j)} - \sum_{k \neq i} \frac{1}{(x_k - x_i)^2} \right]$$

and since

$$\sum_{k \neq i \neq j} \frac{1}{(x_k - x_i)(x_k - x_j)} = 0$$

we finally get for the singlet

$$E(g) = \min_\Phi \frac{\int \prod_k dx_k \ \sum_j | \left[\frac{1}{2} \left| \frac{\partial}{\partial x_j} \Phi(x) \right|^2 + V(x_j) |\Phi(x)|^2 \right]}{\int \prod_k dx_k \ |\Phi(x)|^2} \tag{6}$$

We see that instead of N^2 interacting bosonic particles we now have N free fermions. In general, all the eigenstates of this system can be written as Slater determinants:

$$\Phi_{n_1, \cdots, n_N}(x_1, \cdots, x_N) = \frac{1}{\sqrt{N!}} \det_{k,m} \psi_{n_k}(x_m)$$

where the one particle functions obey the usual Schroedinger eq., but with the Plank constant $\hbar = 1/N$:

$$(-\frac{1}{2N^2} \frac{d^2}{dx^2} + V(x) - \epsilon_k) \psi_k(x) = 0 \tag{7}$$

Eventhough our potentiel is unstable in the limit $N \to \infty$ the amplitudes of the tunneling through the barrier are exponentially small: Im $\epsilon_i \sim e^{-a_i N}$ and the system reduces to the collection of N quasiclassical fermions locked in the metastable well as demonstrated in fig. 1.

The vacuum state energy can be calculated by the formula

$$E_0 = \sum_i \epsilon_i \sim N^2.$$

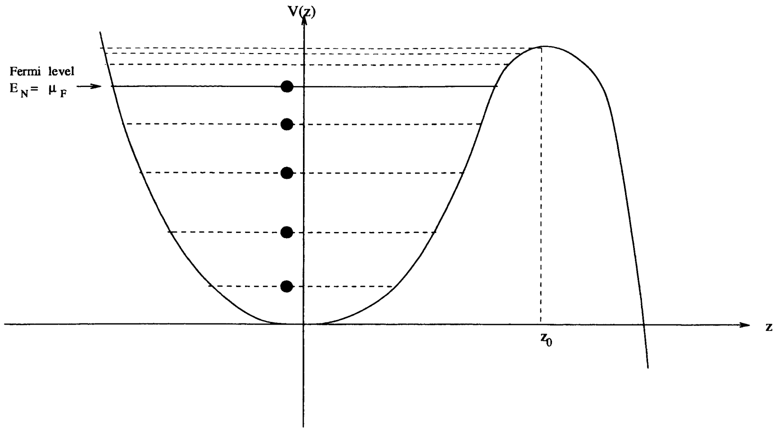


Figure 1. The potential well filled by fermions occupying first N levels.

Note that $\mu = \epsilon_N$ is the Fermi level.

3. Counting of planar (spherical) graphs projected onto the infinite line

In the zero temperature limit $\beta \to \infty$, the free energy of the system is defined by the ground state which should be the most symmetric one, namely, the singlet state (similarly to the ground state of the hydrogen atom which has the zero angular momentum). In the case of singlet one can find E_0 in the planar approximation applying the WKB formulas. One only has to fill the Fermi sea in the phase space (the Thomas–Fermi approximation, which is exact in the limit $N \to \infty$):

$$\frac{g^3}{N^2}E_0(\mu) = \int \frac{dp\,dx}{2\pi}\left(\frac{1}{2}p^2 + V(x)\right)\theta\left(\mu - \frac{1}{2}p^2 - V(x)\right) \tag{8}$$

and from the fact that we have N particles (normalisation) we have:

$$g^2 = \int \frac{dp\,dx}{2\pi}\theta\left(\mu - \frac{1}{2}p^2 - V(x)\right) \tag{9}$$

Introducing the density of energy levels

$$\frac{dN}{d\mu} = \rho(\mu) = g^{-2}\int \frac{dp\,dx}{2\pi}\delta\left(\mu - \frac{1}{2}p^2 - V(x)\right) = g^{-2}\int \frac{dx}{2\pi}\frac{1}{\sqrt{\mu - V(x)}}$$

we obtain for the ground state energy:

$$\frac{\partial}{\partial\mu}\left(\frac{g^2}{N^2}E_0(\mu)\right) = \mu g^2\rho(\mu) = \mu\frac{\partial g^2}{\partial\mu}, \tag{10}$$

This gives a useful formula for the calculation of $E_0(g)$:

$$\frac{\partial}{\partial g^2}\left(\frac{g^2}{N^2}E_0(\mu)\right) = \mu$$

Also, due to the eq. (9) we have

$$Ng^2 = \int_0^\mu d\epsilon \rho(\epsilon). \tag{11}$$

The formulas (10) and (11) give the μ-parametric representation for the free energy $E_0(g)$ allowing to calculate the sum of planar graphs on the one dimensional infinite line for arbitrary size (number of vertices) of graphs. Now let us look at the asymptotics of this sum for the graphs of a large size $n \sim \infty$. This allows us to calculate the partition function of the continuous random surfaces projected onto one dimensional space.

The singularity defining the big graphs corresponds to such a critical coupling constant g_c for which the Fermi level touches the top of the potential (the eigenvalues start to spill over the top if one increases g up to this point, forcing the Fermi level μ to increase to the critical value μ_c as well):

$$\mu_c = V(x_0), \text{ or } V'(x_0) = 0 \text{ and } V''(x_0) < 0.$$

Obviously, we have:

$$\int_0^{\mu_c} d\epsilon \rho(\epsilon) = g_c^2. \tag{12}$$

Near this point the potential can be represented as $V(x) = -\frac{1}{2}(x-x_0)^2 - \frac{g}{3\sqrt{N}}(x-x_0)^3 + \text{const}$ (we changed the variables $x \to \frac{g}{\sqrt{N}}x$ in the functional integral (2)). The cubic term plays the role of a cut-off wall on a distance $|x - x_0| \sim \sqrt{\lambda} \sim \sqrt{N}$ fom the top. Substracting (12) from (11) and calculating the integral in the logarithmic approximation

$$g^2\rho(\mu) = \int_{x_1}^{x_2} \frac{dx}{\sqrt{\mu - V(x)}} \simeq_{\mu\to\mu_c} \log(\mu_c - \mu)$$

we obtain for $\Delta = g_c^2 - g^2$

$$N\Delta \simeq (\mu_c - \mu)\log[(\mu_c - \mu)/\Lambda] \tag{13}$$

and for the singular part of the free energy $\delta E := \left(g^2 E_0(g) - g_c^2 E_0(g_c)\right)$

$$\delta E \sim -(\mu_c - \mu)^2 \log(\mu_c - \mu)$$

Finally, we deduce from the last two equations the necessary asymptotics

$$\frac{1}{N^2}\delta E_0(g) \sim \frac{\Delta^2}{\log[\Delta/\Lambda]} \tag{14}$$

For the susceptibility $\chi = \frac{1}{N^2}\frac{\partial^2 E_0}{\partial g^2}$ we find:

$$\delta\chi_0(g) \sim \frac{1}{\log[\Delta/\Lambda]} \tag{15}$$

Expanding $\delta\chi_0(g) \simeq \sum_n \chi^{(n)} g^n$, where the sum of the graphs of a big size $n \to \infty$ is given by

$$\chi^{(n)} \sim \frac{1}{n \log n} g_c^{-n}.$$

One can also estimate the average size of a random surface in the one-dimensional embedding space. Since the mass-gap (proportional to the inverse correlation time) is

$$m \sim \rho(\mu) \sim \frac{1}{\log(\mu_c - \mu)} \sim \frac{1}{\log n}$$

we have

$$\langle \Delta t)^2 \rangle \sim \log^2 n.$$

4. Double scaling limit and combinatorics of large graphs with fixed topology

It follows from the results of the previous section that in order to study the combinatorics of graphs of a very big size and a fixed genus we need to concentrate at the behaviour of the the model near a local maximum of the matrix potential which, in the corresponding limit (called the double scaling limit [23]), can be approximated by the inverted quadratic potential $V(M) = -\frac{1}{2}M^2$ (see the reviews [7], [2], [24]). This model is unstable and one needs to specify the boundary conditions for big M's. Usually one considers the boundary conditions in which the absolute value of any of the eigenvalues of M cannot exceed some maximum value $\Lambda \sim N$ (a cut-off wall). The spectral density of energy depends in a very universal (logarithmic) way on Λ. In the singlet state the spectrum is that of N independent fermions (eigenvalues) in the same potential and the eigenfunctions are the Slater determinants of the parabolic cylinder functions (see the review in [7]). Knowing all this it is easy to calculate the partition function of singlet sector in the grand canonical ensemble $Z_\mu = \sum_{N=1}^{\infty} e^{\beta\mu N} Z_N$. It allows to define explicitly the asymptotic combinatorics of big graphs of any genus.

Since we know that the critical properties of the large graphs are defined by the behaviour of the eigenvalues in the vicinity of the fermi level which should be placed close to the quadratic top of the potential, we can choose new coordinates near the top having the coordinate x_c:

$$x = x_c + a\frac{z}{\sqrt{N}}, \qquad \epsilon_k = \mu_c - b\frac{\xi_k}{N}$$

and choose a, b in such a way that the Schroedinger equation in these variables will look as:

$$\left(-\frac{d^2}{dx^2} - \frac{1}{4}x^2 + \frac{g}{3\sqrt{N}}x^3 + \xi_k\right)\psi_k(x) = 0.$$

The cubic term is small for $x \ll \sqrt{\Lambda} \sim \sqrt{N}$. One can omit it but impose the cutoff Λ for big values of the fermionic coordinates in all calculations. The singularities will appear to be very soft (logarithmic) due to the logarithmically divergent period of the classical motion:

$$T = 2\int_{2\sqrt{\xi}}^{\sqrt{\Lambda}} \frac{dx}{\sqrt{\frac{1}{4}x^2 - \xi}} \simeq \log(\Lambda/\xi).$$

This ensures the universality of the results (independence of a particular form of potentiel).

Since most of the time of their motion the particles are far away from the reflexion points one can apply the WKB method:

$$\psi_\xi(x) \simeq \frac{C}{\sqrt{x}} \sin[\frac{1}{4}x^2 + \xi \log x + \Phi(\xi) + \Phi_0](1 + O(1/x))$$

or

$$\Phi(\xi) = -\frac{1}{2} i \log \frac{\Gamma(\frac{1}{2} + i\xi)}{\Gamma(\frac{1}{2} - i\xi)}$$

(see [27]). The Bohr–Sommerfeld rule gives for the energy spectrum:

$$\frac{1}{4}\Lambda^2 + \xi \log \Lambda + \Phi(\xi) + \Phi_0 = -\pi n, \quad n = 0, 1, 2, \cdots$$

or, for the density of the energy levels:

$$\begin{aligned} \rho(\xi) &= \frac{1}{2\pi}\frac{dn}{d\xi} = Re\psi(\frac{1}{2} + i\xi) + \frac{1}{\pi}\log\Lambda \\ &= \frac{1}{2\pi}[\frac{1}{\pi}\log(\Lambda^2/\xi) + \sum_{h=1}^{\infty} C_h \xi^{-2h} + O(e^{-2\pi\xi}) \end{aligned} \quad (16)$$

where

$$C_h = -(-1)^h \frac{\frac{1}{2} - 2^{-2h}}{h} B_{2h}$$

and $B_0 = 1, B_2 = 1/6, B_4 = -1/30, B_6 = 1/42, \ldots$ are the Bernulli numbers.

Now, introducing $\nu = \frac{1}{N}(\mu_c - \mu)$, we can give a ν-parametric representation of the free energy (recall that it is also the generating function of graphs) in the form:

$$\begin{aligned} E_0(g) &= N^2 \int^{\nu} d\xi\, \rho(\xi)\xi \\ \Delta &= \int^{\nu} d\xi\, \rho(\xi) \end{aligned} \quad (17)$$

Using the expansion (16) we obtain the topological ($1/N$) expansion of the susceptibility:

$$\begin{aligned} \chi(g, N) &= \frac{1}{\rho} + \text{reg} \simeq \chi_0 + \frac{1}{N^2}\chi_1 + \cdots \\ &= \frac{1}{\log(\nu/\Lambda^2)} + \frac{1}{N^2}\frac{1}{24}\frac{1}{\nu^2 \log^2 \nu} + \frac{1}{N^4}\frac{7}{2800}\frac{1}{\nu^4 \log^2 \nu} + \cdots . \end{aligned} \quad (18)$$

Using $\nu \log \nu \sim \Delta$ one obtains the contribution of big graphs with particular topologies and two marced vertices:

$$\chi_0 = 1/\log\Delta, \qquad \chi_1 = \frac{1}{24}\frac{1}{\Delta^2}, \qquad \chi_2 = \frac{7}{2800}\frac{\log^2\Delta}{\Delta^4}.$$

Expanding these expressions in g/g_c we find the universal part of the sum over surfaces of a given genus progected onto the one-dimensional line. Note that even the numerical constants in these formulas are universal numbers. For example, these results would not change if we use the $\operatorname{tr} M^4$ interaction in the MQM model (2) giving the 4-valent graphes instead of the $\operatorname{tr} M^3$ interaction corresponding to the 3-valent graphs.

5. Matrix quantum mechanics in periodic time and related Toda hierarchy

Let us now try to investigate this model at finite temperature, or equivalently, in the periodic (compact) imaginary time. One can show that this problem is related to the statistical mechanical system of Berezinski–Kosterlitz–Thouless vortices on dynamical (random) graphs of various topologies [2], [8], [9] (see fig. 2). It was conjectured in [5] and shown in [3] that the adjoint representation describe the vortex anti-vortex sector in the 1+1 dimensional string theory with one compact dimension. Higher representation describe higher numbers of vortex anti-vortex pairs (corresponding to the number of boxes in the Young tableau of the representation).

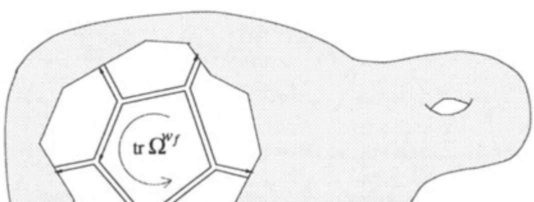


Figure 2. A typical planar graph of a fixed topology, with a vortex in one of the faces.

First, let us show that the compactification of time forces us to take into account not only the singlet sector (which for the infinite time interval of the previous sections) but also all non-singlet sectors. As was noticed earlier, the Hilbert space of the MQM is factorized into the non-interacting sectors corresponding to all irreps of $SU(N)$ symmetry of the matrix Hamiltonian. Namely, we can write the twisted partition function (1) as a sum over all irreps r of $SU(N)$:

$$Z_N(\beta,\Omega) = \sum_r \chi_r(\Omega)\,\mathrm{Tr}_r\, e^{-\beta \hat{H}_r} \tag{19}$$

where $\chi_r(\Omega)$ is the Weyl character of the irrep r, $\hat{H}_r$ is the hamiltonian of the MQM in this irrep and the trace Tr_r is taken with respect to the $r-th$ sector of the Hilbert space of the model. To calculate $\hat{H}_r$ through the matrix eigenvalues it is enough to consider in (4), (5) the wave function transforming as:

$$\Psi_r^I(\Omega^+ M\Omega) = \sum_J \Omega_r^{IJ}\Psi_r^J(M)$$

where Ω_r is a group element Ω in representation r and I, J are the indices of the representation. Such a function may be decomposed as

$$\Psi_r^I(M) = \sum_J \Omega_r^{IJ}\psi_r^J(x). \tag{20}$$

Here $\psi_r^I(x_1, \cdots, x_N)$ is a vector in the representation r.

Near the unity element on the group space $\Omega \simeq I + \omega$ we have $\Omega_r \simeq P_r + \sum_{ij}\omega_{ij}T_{ij}^r$ where P_r is a projector (unity element) in the r space, ω is a small deviation from it and T_{ij}^r are the $u(N)$ algebra generators. This gives:

$$\frac{\partial}{\partial M_{ij}} = \delta_{kj}\frac{\partial}{\partial x_k} + \sum_{m=1}^{N}\frac{1}{x_k - x_m}\frac{\partial}{\partial \omega_{mj}}$$

and we finally obtain from (4) the following variational principle:

$$\min_{\psi_r}\int \prod_k dx_k \Delta^2(x)$$
$$\times \mathrm{Tr}_r\Big(\frac{1}{2}\sum_j |\frac{\partial}{\partial x_j}\psi_r(x)|^2 + \frac{1}{2}\sum_{i\neq j}|T_{ij}^r\psi_r|^2 + \sum_m V(x_m)|\psi_r|^2\Big) \tag{21}$$

where all the quantities and operators with the subscript r are subject to the corresponding matrix operations in the matrix space of representation.

The Schroedinger equation now reads:

$$-\frac{1}{2}\sum_k \Delta^{-2}(x)\frac{\partial}{\partial x_k}\Delta^2(x)\frac{\partial}{\partial x_k}\psi_r(x) - \frac{1}{2}\sum_{i\neq j}T_{ij}^r T_{ji}^r\psi_r(x)$$
$$= \big(E - \sum_k V(x_k)\big)\psi_r(x). \tag{22}$$

It is useful to introduce a new function $\phi_r(x) = \frac{1}{\Delta(x)}\psi_r(x)$ obeying the equation

$$-\sum_k\left(\frac{1}{2}\frac{\partial}{\partial x_k}\right)^2\phi_r(x) - \frac{1}{2}\sum_{i\neq j}\frac{T_{ij}^r T_{ji}^r}{(x_i - x_j)^2}\phi_r(x) = \big(E - \sum_i V(x_i)\big)\phi_r(x). \tag{23}$$

Note that any translation $\omega_{ij} \to \omega_{ij} + \delta_{ij}\epsilon$ does not change the wave function Ψ_r. That means that we are looking only for the states on which the condition is imposed

$$T_{kk}^R\psi_r = 0, \quad k = 1, \cdots, N. \tag{24}$$

On the first sight, we fulfilled our main task for the matrix quantum mechanics: we reduced it to an eigenvalue problem and are now dealing with only N variables. But the Schroedinger equation (22) contains the Hamiltonian which is a matrix in the representation space acting on the wave function which is a vector in this space. For small representations whose Young tableaux contain $\ll N^2$ boxes the problem is still solvable in the large N limit (as we will demonstrate below). For a very interesting case of big representations ($\sim N^2$ boxes in the Young tableaux) the problem remains a serious challenge.

As we saw in section 2, in the simplest case of singlet representation the wave function is a scalar and the last term in the r.h.s. of the Schroedinger equation (23) drops out. The problem appears to be equivalent to the quantum mechanical system of N non-interacting fermions (due to the antisymmetry of $\phi(x)$) in a potential $V(x)$.

The next smallest representation is adjoint. The adjoint wave function satisfying the relation (20) should be a function of the type

$$\Psi(M;x) = \sum_{a=0}^{N-1} C_a(x) M^a$$

where the coefficients C_A possibly depend on the invariants (eigenvalues). If we denote $\phi_{adj}(x_i;x) \equiv \phi_i(x)$ (depending of course on all N x_i) we can write the Schroedinger equation for the adjoint wave function in the form:

$$\sum_i \Big(-\left(\frac{\partial}{\partial x_i}\right)^2 + V(x_i)\Big)\phi_k(x) - \frac{1}{N^2}\sum_{i(\neq k)} \frac{\phi_i(x)-\phi_k(x)}{(x_i-x_k)^2} = E\phi_k(x). \qquad (25)$$

One can see that the last term in the l.h.s. of this equation is $\sim N^2$ smaller than the other terms and can be regarded as a small perturbation on the background of the free fermion solution of the singlet sector.

Partition function (1) can be also represented in the functional integral form (2) if we take there the twisted periodic boundary conditions for the matrix field:

$$M(\beta) = \Omega^\dagger M(0)\Omega. \qquad (26)$$

It is hard to analyze the partition function (1) for an arbitrary matrix potential $V(M)$, even in the large N limit. However, we noticed in the section 3 that in the double scaling limit (describing interesting applications related to the counting of graphs of fixed topologies), we need to solve the model in the inverted oscillatorial potential $V(M) = -M^2$. The model is unstable and one needs to specify the boundary conditions for big M's. Usually one considers the boundary conditions when the absolute value of any of the eigenvalues of M cannot exceed some maximum value Λ (a cut-off wall). In the case of the large N limit one takes $\Lambda \sim N$ and it happens that the spectrum density of the model depends in a very universal and soft (logarithmic) way on Λ. In the singlet state the spectrum is that of N independent fermions (eigenvalues) in the same potential and the eigenfunctions are the Slater determinants of the parabolic cylinder functions. In the non-singlet sectors the eigenvalues start interacting and obey a more complicated statistics corresponding to the symmetry of the Young tableau of representation (see the review [15] for the details; for the large N estimates of the mass gap of adjoint representation see [16], [5] and [3]). In principle, the procedure of the double scaling limit introduced for the singlet sector should be carefully reconsidered in the case of non-singlet irreps. However, intuitively we understand that for small higher representations the deviation from the singlet sector can be treated as a $1/N^2$ perturbation which hardly can change our choise of the doubly scaled potential. Hence, at least perturbatively, our choice of the inverted oscillatorial potential should be kept for higer irreps. It is now natural to assume that this picture will be true in some finite range of parameters (such as the size of the corresponding Young tableaux) characterizing the deviation from the singlet.

Let us study the partition function (1) for the inverted oscillatororial potential. Formally (if we forget about the instability of this potential and about the existence of the cut-off wall) the functional integral (2) becomes gaussian and we can try to integrate out the matrix field $M(t)$. We will do it first for the usual stable oscillatorial potential $V(M) = \omega^2 M^2$ and then try to continue analytically the result by changing the frequency $\omega \to i\omega$ (which is equivalent to the Wick rotation of the periodic time $\beta \to i\beta$).

To calculate the gaussian integral in (2) we first notice that due to the $SU(N)$ invariance of the model the twist matrix can be taken diagonal $\Omega = (z_1, \cdots, z_N)$ without a loss of generality. Then the action in (2) factorizes into the sum of oscillatorial actions of individual matrix elements $M_{ij}(t)$ (and their complex conjugates), each with its own periodicity condition: $M_{ij}(\beta) = \frac{z_j}{z_i} M_{ij}(0)$. The functional integral over the individual matrix elements can be easily calculated (see [17] for example) and we obtain a simple expression:

$$Z_N(\Omega) = \prod_{j,k=1}^{N} \frac{1}{|z_j q^{1/2} - z_k q^{-1/2}|}, \tag{27}$$

where

$$q = e^{2\pi R i}, \text{ where } R = \frac{\beta}{2\pi}. \tag{28}$$

To find a link between this partition function and the Toda lattice hierarchy, we integrate the partition function over the twist matrix with an $SU(N)$ invariant weight depending on the infinite set of new couplings λ_n

$$Z(\mu, \lambda_{\pm 1}, \lambda_{\pm 2}, \cdots) = \int [d\Omega]_{SU(N)} \exp[\sum_{n \in \mathbf{Z}} \lambda_n \operatorname{tr} \Omega^n] Z[\mu, \Omega] \tag{29}$$

and introduce the grand canonical partition function:

$$\mathcal{Z}[\mu, \lambda] = \sum_{N=0}^{\infty} e^{2\pi R \mu N} \, \mathcal{Z}_N[\lambda]. \tag{30}$$

We will show now that the resulting partition function appears to be a τ-function of Toda hierarchy, with the couplings λ_n of the potential playing the role of "times" of the Toda commuting flows [1]. The Toda equations, together with the boundary conditions at $\lambda_{\pm 1} = \lambda_{\pm 2} = \cdots = 0$ (the known free fermion result for the singlet, vortex free sector), appear to be sufficient to find the partition functions and some correlators of the model for various genera of graphs and for different $SU(N)$ sectors of the MQM at finite temperature describing the Berezinski–Kosterlitz–Thouless vortex dynamics on dynamical (random) graphs of various topologies [8], [9].

Using the Haar measure for the $SU(N)$ integration in (29)

$$[d\Omega]_{SU(N)} = \frac{1}{N!}\prod_{k=1}^{N} dz_k \prod_{i>j}|z_i - z_j|^2$$

we write $\mathbb{Z}_N[\lambda]$ in the form

$$\mathcal{Z}_N[\lambda] = \frac{1}{N!}\oint \prod_{k=1}^{N} \frac{dz_k}{2\pi i z_k}\, \frac{e^{2u(z_k)}}{(q^{1/2} - q^{-1/2})} \prod_{j\neq j'}^{N} \frac{z_j - z_{j'}}{q^{1/2}z_j - q^{-1/2}z_{j'}}, \tag{31}$$

where $u(z) = \frac{1}{2}\sum_n \lambda_n z^n$.

Another useful representation of $\mathcal{Z}_N$, obtained by using the Cauchy identity

$$\frac{\Delta(a)\Delta(b)}{\prod_{i,j}(a_i - b_j)} = \det_{(i,j)}\left(\frac{1}{a_i - b_j}\right) \tag{32}$$

is

$$\mathcal{Z}_N[\lambda] = \prod_{k=1}^{N}\oint \frac{dz_k}{2\pi i}\ \det_{(i,j)}\left(\frac{\exp\left[u(z_i) + u(z_j)\right]}{z_i q^{1/2} - z_j q^{-1/2}}\right). \tag{33}$$

A natural way to avoid the ambiguity in the contour integration in (33) is to add a small imaginary part to the radius R of the time-circle and integrate over the $\{z_k\}$ along the unit circle.

The grand canonical partition function (30) takes then the form of a Fredholm determinant

$$\mathcal{Z}_\mu = \mathrm{Det}(1 + e^{\mu 2\pi R}\hat{K}), \tag{34}$$

where the integral operator $\hat{K}$ is defined as

$$(\hat{K}f)(z) = -\oint \frac{dz'}{2\pi i}\, \frac{e^{u(z)+u(z')}}{q^{1/2}z - q^{-1/2}z'}\, f(z').$$

The partition function can be calculated by residues for given N and t_k, but the formulae get more and more complicated when N grows. Furthermore, the analytical structure of $\mathcal{Z}_N$ makes it difficult to formulate a well defined $1/N$ expansion. To avoid these problems, let us use the powerful technique of integrability. Namely, let us demonstrate that the partition function of the inverted matrix oscillator defined by (30) and (31) (or (33)) satisfies the Toda lattice equations.

According to the results of [18], the most general solitonic τ-function solving the differential equations of the Toda lattice hierarchy can be presented in the form (adopting our notations):

$$\tau_l[\mu, \theta] = \det_{1<j,k<L} \left(\delta_{jk} + e^{\beta\mu} \hat{\mathcal{K}}(p_j, p_k) \right) \tag{35}$$

where

$$\hat{\mathcal{K}}_l(p_j, p_k) = \left(\frac{p}{q}\right)^l \frac{E(p)E(q)}{p-q} \exp\left(\sum_{n\in\mathbb{Z}} t_n(p^n - q^n)\right) \tag{36}$$

where the function $E(p)$ does not depend on t_n's and p_j is an arbitrary infinite set of numbers.

Putting $E(p) = 1$ and placing continuously p_j on the unit circle: $p_j = e^{\frac{2\pi i j}{L}}$, $L \to \infty$ we identify our partition function (34) as a solitonic τ-function of the Toda chain hierarchy. Our partition function is a particular case of the continuum limit (infinitely many solitons) of the soliton referred to as "general solution" in [18]. Explicitly the τ-function with charge l is defined as

$$\tau_l[t] = e^{-\sum_n n t_n t_{-n}} \sum_{N=0}^{\infty} (q^l\ e^{2\pi R\mu})^N \mathcal{Z}_N[t] = e^{-\sum_n n t_n t_{-n}} \mathcal{Z}_{\mu+il}[t] \tag{37}$$

where

$$t_n = \frac{\lambda_n}{q^{n/2} - q^{-n/2}} \tag{38}$$

The grand canonical partition function (30) is obtained by taking $l = 0$. It is easy to see that the τ-function with the vacuum charge l can be obtained from (30) by replacing $\mu \to \mu + il$.

As a function of the charge l and the "times" $t_1, t_2, \cdots$, this τ-function satisfies a set of the PDE's which can be summerized in the so called Hirota identity [19], [18]. The first nontrivial equation of this set is the famous Toda lattice equation:

$$\tau_l(\partial_+\partial_-\tau_l) - (\partial_+\tau_l)(\partial_-\tau_l) + \tau_{l+1}\tau_{l-1} = 0, \tag{39}$$

where we denoted $\partial_\pm = \partial/\partial\lambda_\pm$ and $\lambda_\pm = t_{\pm 1} = \pm\frac{\lambda_{\pm 1}}{2i\sin(\pi R)}$.

We will show in what follows how to use the Toda equation to extract the physical results for our MQM.

6. Free energy of the MQM of inverted twisted oscillator from Toda equation

As mentioned above, the τ-function depends on the cosmological constant μ and the charge l via the combination $\mu + il$

$$\tau_l(\mu) = \tau_0(\mu + il). \tag{40}$$

We will restrict ourselves to the case $\lambda_n = \lambda(\delta_{n,1} + \delta_{n,-1})$ (note that $\lambda_\pm = \frac{\lambda_{\pm 1}}{\pm 2i \sin(\pi R)}$), for which

$$\tau_0(\mu) = e^{-\lambda_+\lambda_-} \mathcal{Z}_\mu(\lambda). \tag{41}$$

Winding number conservation implies that the τ-function depends only on the product $\lambda^2 = \lambda_+\lambda_-$, so that

$$\partial_+\partial_- = \frac{1}{4}\lambda^{-1}\partial_\lambda\lambda\partial_\lambda, \qquad \lambda = \sqrt{\lambda_+\lambda_-}.$$

The free energy $\mathcal{F}(\lambda, \mu) = \log \mathcal{Z}_\mu(\lambda)$ of the matrix model satisfies

$$\partial_+\partial_-\mathcal{F}(\lambda, \mu) + \exp\left[\mathcal{F}(\lambda, \mu + i) + \mathcal{F}(\lambda, \mu - i) - 2\mathcal{F}(\lambda, \mu)\right] = 1. \tag{42}$$

6.1. BOUNDARY CONDITIONS

Note that the differential equation (42) do not depend on R explicitly. However, to solve them we need to supply boundary conditions. A convenient choice is to specify the free energy $\mathcal{F}$ at $\lambda = 0$, and use the differential equations to solve for the λ dependence. Since $\lambda = 0$ corresponds to the singlet sector of the MQM described in sections 2-3, $\mathcal{F}(\mu) \equiv \mathcal{F}(\lambda, \mu)|_{\lambda=0}$ is known. Let us review its form.

The corresponding grand canonical partition function (30) in the absence of vortices (all $t_m = 0$) can be expressed in therms of the Fermi–Dirac ensemble as a partition function of free fermions with the density (16) of the energy levels of the inverted harmonic oscillator (see [5]). The (universal part of the) grand canonical free energy of the MQM compactified on a circle of radius R has a well defined (asymptotic) $1/\mu$ expansion:

$$\begin{aligned}
\mathcal{F}(\mu) &= \log \mathcal{Z}_\mu[0] \\
&= \frac{1}{2}\int_{-\infty}^{\infty} dE\rho(E)\log\left(1 + e^{-2\pi R(\mu - E)}\right) \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} dE \sum_{k=0}^{\infty} \frac{k + \frac{1}{2}}{E^2 + (k + \frac{1}{2})^2}\log(1 + e^{-2\pi R(\mu - E)}) \\
&= -\frac{R}{2}\mu^2\log\mu - \frac{1}{24}(R + \frac{1}{R})\log\mu + R\sum_{k=2}^{\infty}[R\mu]^{-2(k-1)}f_k(R) + O(e^{-2\pi\mu}),
\end{aligned} \tag{43}$$

with the known coefficients $f_k(R)$ which are polynomials in R.

6.2. CANONICAL VERSUS GRAND CANONICAL PARTITION SUM

The sum over worldsheets in the continuum string theory is described in the matrix model by the $1/N$ expansion of the canonical free energy $\mathcal{F}_N = \log \mathcal{Z}_N$. The formalism described in the previous subsections allows one to compute the grand canonical free energy $\mathcal{F}(\mu) = \log \mathcal{Z}_\mu$, which is related to $\mathcal{F}_N$ via

$$\exp \mathcal{F}_N = \oint \frac{d\mu}{2\pi i} \exp[2\pi R\mu N + \mathcal{F}(\mu)] \tag{44}$$

where the integration contour encircles the point $e^{-2\pi R\mu} = 0$.

As follows from the relation (13) (see [7, 2, 25] for more details), the $1/N$ expansion of $\mathcal{F}_N(\lambda)$ can be rearranged as a $1/\mu_0$ expansion, with μ_0 defined by $\mu_0 \log(\mu_0/\Lambda) = N(g_c - g)$. The grand canonical free energy $\mathcal{F}(\lambda = 0, \mu)$ has a $1/\mu$ expansion given by (43). Similarly, we will see in the next section that $\mathcal{F}(\lambda, \mu)$ has a $1/\mu$ expansion which can be generated from (42). One can show that in the double scaling limit, the $1/\mu_0$ expansion of $\mathcal{F}_N$ coincides with the $1/\mu$ expansion of $\mathcal{F}(\mu)$, up to a flip of the sign[1] of the leading (spherical) contribution to $\mathcal{F}$ (see [3], revised version, section 6).

In the next section we will construct the $1/\mu$ expansion of the grand canonical free energy, and use the above discussion to identify it with the genus expansion of the partition sum of the continuum theory.

7. Some results on the genus expansion of the free energy

The purpose of this section is to solve the differential equations (42) with the boundary conditions (43). This will allow us to calculate explicitely the planar limit of the partition function of the periodic MQM with twisted boundary conditions. We have not found an exact solution of these equations for any μ, but as we will see below, one can solve them iteratively, in a genus expansion,

$$\mathcal{F}(\lambda, \mu) = \sum_{h=0}^{\infty} \mathcal{F}_h(\lambda, \mu). \tag{45}$$

In the planar limit we know that $\mu \to \infty$. It is natural to expect that for a sensible physical limit we have to send also $\lambda \to \infty$ in such a way that a ratio λ/μ^ν with a certain ν be fixed. We will fix $\nu = \frac{1}{2}R - 1$ from the analysis of the planar limit of the equation (42) together with the boundary conditions (43). This scaling appears to correspond well to the so called KPZ-DDK physical scaling of our problem—the vortices on planar graphs [21], [2]. The genus h partition sum $\mathcal{F}_h$ conforme to this scaling must have the form[2]

$$\mathcal{F}_h(\lambda, \mu) = -\delta_{h,0}\frac{R}{2}\mu^2 \log\mu - \frac{\delta_{h,1}}{24}(R + \frac{1}{R}) \log\mu + \mu^{2-2h} A_h(z), \tag{46}$$

where $z = (R-1)\lambda_+\lambda_-\mu^{R-2}$. The boundary conditions (43) determine $A_h(z)$ at $z = 0$. Plugging the ansatz (46) into (42) one can solve for $A_h(z)$. It is remarkable that the Toda equations allow such an ansatz at all; this is due to their conformal symmetry.

At tree level one finds an ODE for $A_0(z)$ which together with the boundary conditions fixes it uniquely. At genus h one finds a differential equation that mixes $A_h(z)$ with $A_{h'}(z)$ with $h' < h$. These equations can be used to determine all A_h iteratively given $A_h(0)$.

[1] This sign flip is a standard feature of the Legendre transform.

[2] The fact that the r.h.s. of (42) is one, leads to an additive contribution $\lambda_+\lambda_-$ to $\mathcal{F}$. This contribution is non-universal and will be dropped below.

Let us demonstrate how it works for the spherical (large μ) approximation. Introducing the "susceptibility" as:

$$\chi(\lambda,\mu) := \partial_\mu^2 \mathcal{F}(\lambda,\mu) \tag{47}$$

we can rewrite it in the spherical limit as

$$\chi_0(\lambda,\mu) = -\frac{2R}{2-R}\log\left(\lambda\sqrt{R-1}\right) + X_0(y), \tag{48}$$

where

$$y = z^{-\frac{1}{2-R}} = \frac{\mu}{\lambda^{2/(2-R)}\kappa}, \qquad \kappa = (R-1)^{1/(2-R)}$$

and

$$X_0(y) = 2A_0(z) - R\log y.$$

In the Toda eq. (42) we can use now the expansion:

$$\mathcal{F}(\lambda,\mu+i) + \mathcal{F}(\lambda,\mu-i) - 2\mathcal{F}(\lambda,\mu) = \sum_{n=1}^{\infty}\frac{(-1)^n}{(2n)!}\partial_\mu^{2n}\mathcal{F}(\lambda,\mu). \tag{49}$$

In the limit $\mu \to \infty$ it takes then the form:

$$\alpha\left(y\partial_y\right)^2 X_0 + \partial_y^2 e^{-X_0} = 0 \tag{50}$$

where $\alpha = (R-1)/(R-2)^2$. The solution is

$$y = e^{-\frac{1}{R}X_0} - e^{\frac{1-R}{R}X_0}. \tag{51}$$

The integration constant is fixed by comparing to the leading term in (43).

From (51) we can find for example the λ expansion of the free energy of our model:

$$\begin{aligned} F_0(\lambda,\mu) &= -\frac{R}{2}\mu^2\log\frac{\mu}{\Lambda} \\ &\quad + R\mu^2\sum_{n=1}^{\infty}\frac{1}{n!}\left[(1-R)\mu^{R-2}\lambda_+\lambda_-\right]^n\frac{\Gamma(n(2-R)-2)}{\Gamma(n(1-R)+1)}. \end{aligned} \tag{52}$$

reproducing the result conjectured by G. Moore [20]. The integration constant is fixed by comparing to the leading term in (43).

One can find in this way the partition function for the torus (next order of $1/N^2$ approximation), as well as a triangular system of the 2-nd order linear ODE's allowing (in principle) to define recursively the results genus by genus (see [8]).

8. Discussion and unsolved problems

There are many interesting physical problems behind the nice mathematics related to the twisted periodic MQM. These problems involve, for example, various aspects of the 2D string theory, 2D black hole physics and 2D statistical mechanics of vortices. We will not discuss the unsolved physical questions here but ruther formulate the underlying mathematical problems which wait for their resolution:

1. The description of spectra and states of higher representations of $SU(N)$. This problem is advanced very much by the Toda integrability described in the previous sections but it seems to be only the beginning of the work.

2. There are two types of "physical" operators in this theory which are in some sense dual to each other: one correspond to the hermitean matrix field $M(t)$ and the other—to the twist $SU(N)$ matrix Ω. The Toda integrability allows to calculate effectively (at least in the large N approximation corresponding to the dispersionless limit of Toda eqations) various correlators of each of these variables, separately.

3. An interesting problem is to calculate the partition functions of the twisted MQM beyond the torus and sphere approximation of the previous section. Even the results for a few next orders of the topological expansion, beyond the sphere and the torus partition functions found in [8], would be very instructive. Even more ambitious (but not hopeless, in our opinion) is the problem of finding of the exact nonperturbative (in this sense) solution which would certainly shed a new light on very important problems of the string theory. A useful tool on this way might be the "string equation" for the described system proposed in [26].

To conclude, we would like to underline that the matrix quantum mechanics appeared to be a very important physical system with many impressive applications and at the same time obaying a very beautiful mathematical structure promissing many new discoveries.

References

1. Hoppe, J., Kazakov V. and Kostov, I. (2000) Dimensionally Reduced SYM_4 as Solvable Matrix Quantum Mechanics, *Nucl.Phys.* B, **571**, 479; hep-th/9907058.
2. Klebanov, I. (1991) String Theory in Two Dimensions, Lectures delivered at the ICTP Spring School on String Theory and Quantum Gravity, Trieste, April 1991; hep-th/9108019. (Review sur la corde a une dimension - surfaces aleatoires en 1 dimension, les approches matriciel et conforme)
3. Boulatov, D. and Kazakov, (1993) One dimensional string theory with vortices as an upside down matrix oscillator, *J. Mod. Phys.* A, **8**, 809; hep-th/0012228.
4. Kazakov, V. and Migdal, A. (1988) Recent Progress in the Theory of Non-critical Strings, *Nucl. Phys.* B, **311**, 171. (Review sur la gravite quantique a 2 dimensions et scaling physique pour les surfaces aleatoires en 1 dimension).
5. Gross, D. and Klebanov, I. (1990) Vortices and the non-singlet sector of the c=1 matrix model, *Nucl. Phys.* B, **354**, 459.
6. Brezin, E., Itzykson, C., Parisi, C. and Zuber, J.-B. (1978) Planar approximation, *Comm. Math. Phys.*, **59**, 35. (Le papier classique sur le modele d'une matrice et la mechanique quantique matricielle.)
7. Kazakov, V. A. (1991) Bosonic strings and string field theories in one-dimensional target space, in O. Alvarez, E. Marinari, P. Windey (Eds.) *Random surfaces and quantum grav-*

ity, Carg se 1990. (Review sur les surfaces aleatoires et $D = 1$—mechanique quantique matriciele).

8. Kazakov, V., Kostov, I. and Kutasov, D. (2001) A Matrix Model for the Two Dimensional Black Hole; `hep-th/0101011`.
9. Alexandrov, S. and Kazakov V. (2001) Correlators in 2D string theory with vortex condensation, *Nucl. Phys.* B, **610**, 77; `hep-th/0104094`.
10. Boulatov, D. V., Kazakov, V. A., Kostov, I.K. and Migdal, A.A. (1986) Analytical and numerical study of a model of dynamically triangulated random surfaces, *Nucl. Phys.* B, **275**, [FS17], 641. (Tout sur les surfaces aleatoires en D dimensions (y compris les resultats numeriques) et les resultats analytiques sur $D = -2, 0, 1$.)
11. Boulatov, D. V. and Kazakov, V. A. (1987) The Ising model on a random planar lattice: the structure of phase transition and the exact critical exponents, *Phys. Lett.* B, **186** 379. (La solution et l'analyse du modele d'Ising sur les reseaux aleatoires=le modele de 2 matrices).
12. Ginsparg, P. and Moore, G. (1993) Proceedins of TASI 1992, *hep-th/9304011*. (Review sur la gravite quantique a 2 dimensions: les approches matriciel et conforme).
13. Di Francesco, P., Ginsparg, P. and Zinn-Justin, J. (1995) 2D gravity and random matrices, *Phys. Rept.*, **254** 1–133. (Review sur la gravite quantique a 2 dimensions: les approches matriciel et conforme).
14. Mehta, M. L. (1991) *Random Matrices*, Acad. Press, NY. (le livre classique sur les matrices aleatoires).
15. Polychronakos, A. P. (1999) Generalized statistics in one dimension, `hep-th/9902157`, Les Houches 1998 Lectures; 54 pages.
16. Marchesini, G. and Onofri, E. (1980) Planar limit for $SU(N)$ symmetric quantum dynamical systems, *J. Math. Phys.*, **21**, 1103.
17. Feynman, R. P. (1972) *Statistical Mechanics*, W. A. Benjamin, Inc., Advanced book program reading, Massachusetts.
18. Ueno, K. and Takasaki, K. (1984) *Adv. Stud. Pure Math.*, **4**, 1; Takasaki, K. (1984) *Adv. Stud. Pure Math.* **4** (1984) 139.
19. Hirota, R. (1980) Direct Method in Soliton Theory, in R. K. Bullogh and R. J. Caudrey (Eds.) *Solitons*, Springer, 1980; Jimbo, M. and Miwa, (1983) T. Publ. RIMS, Kyoto Univ. **19**, No. 3, 943.
20. Moore, G. (1992) `hep-th/9203061`.
21. Knizhnik, V., Polyakov, A. and Zamolodchikov, A. (1988) *Mod. Phys. Lett.* A, **3**, 819; David F. (1988) *Mod. Phys. Lett.* A, **3**, 1651; Distler, J. and Kawai, H. (1989) *Nucl. Phys.* B, **321**, 509.
22. Das, S. and Jevicki, A. (1990) *Mod. Phys. Lett.* A, **5**, 1639.
23. Brezin, E., Kazakov, V. and Zamolodchikov, A. (1990) *Nucl. Phys.* B, **338**, 673; Parisi, G. (1990) *Phys. Lett.* B, **238**, 209, 213; Gross, D. and Miljkovic, N. (1990) *Phys. Lett.* B, **238**, 217; Ginsparg, P. and Zinn-Justin, J. (1990) *Phys. Lett.* B, **240**, 333.
24. Polchinski, J. (1994) What is String Theory?, Lectures presented at the 1994 Les Houches Summer School "Fluctuating Geometries in Statistical Mechanics and Field Theory", `hep-th/9411028`.
25. Ginsparg, P. and Moore, G. (1993) Proceedings of TASI 1992, `hep-th/9304011`.
26. Kostov, I. K. `hep-th/0107247`.
27. Landau, L. and Lifchits, E. (1951) Quantum Mechanics, vol. 3, ch. 7, the problem of E. C. Kemble in sect. 50.

INTRODUCTION TO MATRIX MODELS

E. BRÉZIN (edouard.brezin@lpt.ens.fr)
Laboratoire de Physique Théorique
Ecole Normale Supérieure
24 rue Lhomond 75231
Paris Cedex 05
*France**

1. Examples of physical problems involving random matrices

Nuclear levels. In 1951 E. Wigner [1] proposed, as a first approach to the understanding of the structure of the eigenstates of complex nuclei, to sustitute to the Schrödinger equation, in which the forces between the nucleons are not very well-known, and whose solution requires drastic simplifying assumptions, a simple random Hamiltonian drawn from a Gaussian ensemble. The only constraints are imposed by symmetries, such as time-reversal symmetry, and the need to focus on a sequence of levels with given angular momentum and parity. This approach was then developped considerably by Dyson [2, 3], Mehta [4] and hundreds of followers. Numerous review articles on nuclear data show that the level spacing distribution, as drawn from random matrix theory, fits well the measured level spacing on a large sequence of nuclei, see, for instance, [4].

Quantum chaos. The original approach of Wigner was then followed by the study of numerous Schrödinger equations for non-integrable Hamiltonians, such as the hydrogen atom in a magnetic field [5], free motion in non integrable shapes such as "stadium", Sinai billiards, etc. In all those cases the statistics of levels was strikingly fitted by simple ranom matrix results, although there is nothing random in the problem itself. This led Bohigas et al. [6] to the conjecture of *quantum chaos*, namely that the statistics of the eigenlevels of the Schrödinger operators, corresponding to classicaly non-integrable Hamiltonians, were always asymptotically given by random matrix theory, provided the energies were large compared to the mean level spacing. This striking conjecture, which resembles the substitution of statistical mechanics to a hopeless detailed dynamical approach, except that in quantum chaos we may be simply dealing with just a few degrees

* Unité Mixte de Recherche 8549 du Centre National de la Recherche Scientifique et de l'École Normale Supérieure.

V. A. Malyshev and A. M. Vershik (eds.),
Asymptotic Combinatorics with Application to Mathematical Physics, 23–50.

of freedom, has been attracting a number of interesting attempts to explain it [7], but the final word may still be missing.

Universal fluctuations in mesoscopic systems. One is dealing there with small metallic particles at very low temperatures. The inelastic scattering of the electrons is so rare that the electron propagate with full phase coherence. The sample to sample fluctuations of the conductance are then well described by random matrix theory, with the appropriate symmetry (orthogonal or unitary if a magnetic field is present). A extensive description of the use of random matrix theory in mesoscopic systems may be found in Beenakker's review article [8].

Lattice QCD in the large N limit. Quantum chromodynamics on a space-time lattice is much used for numerical calculations in strong coupling, needed for finding the spectrum of the observed colour singlets particles. It consists of a lattice in which to each link is associated a unitary matrix drawn from the group $SU(N)$. Eguchi and Kawai have shown that for N large the model on an infinite lattice may be reduced to one single hypercube, with thus 16 coupled matrices in four dimensions (and appropriate boundary conditions). This is an example in which the reduction of a problem to an integral over a few large matrices, instead of the full initial problem on an infinite lattice, is somewhat illusory. The number of degreees of freedom for N large goes also to infinity, and it is only in well chosen cases that the resulting integrals may be asymptotically computed in the large N limit.

Growth models. Recently Prahőfer and Spohn [9] have considered a model describing the growth of an interface in $(1 + 1)$ dimension. It consists of growth through random nucleation of terraces, which propagate after nucleation, and eventually merge. They have shown that their model may be mapped into the problem of the probability distribution of the longest increasing subsequence of a random permutation, a problem recently solved in the mathematical literature, and connected to the main lines of this school. The connection between random permutations and random matrices goes as follows. One draws at random an integer n with Poissonian probability $p_n = \frac{\lambda^n}{n!}e^{-\lambda}$, one then takes an random one of the $n!$ permutations of S_n. The probability that the longest increasing subsequence has a length l which does not exceed m has been shown to be an integral over the group $U(m)$:

$$\mathrm{Prob}\{l \leq m\} = e^{-\lambda} \int_{m\times m} dU \, \exp\{\sqrt{\lambda}\mathrm{Tr}(U + U^{\dagger})\}.$$

This integral well-known in the physics litterature (it is QCD for one plaquette) has been solved by Gross and Witten. It has a critical point at $\lambda/m^2 = 1/4$, and its "double" scaling limit has been found by Periwal. It is described by the solution of a Painlevé II equation. We shall describe in these lectures not this particular integral, which may be analyzed by very similar techniques, but how such non linear differential equations arise when one studies the asymptotic behaviour of integrals over matrices.

Number theory. Let us also mention outside of the world of physics, the amazing connections between number theory and random matrices. The most studied example concerns the zeros of Riemann zeta-function on the critical line, namely the spacing of the t_n given by

$$\zeta(\frac{1}{2} + it_n) = 0.$$

The celebrated numerical studies of Odlyzko [10] show, for instance, that the spacing distribution of 70 million zeros around the 10^{20}th zero is indistinguishable from the spacing distribution of the eigenvalues of random Hermitian matrices.

2. Random matrices and and random surfaces

We have to go here over very elementary and well-known algebraic techniques used in field theory, which the reader should of course skip if he or she is familiar with Wick theorem and Feynman diagrams.

2.1. WICK THEOREM FOR GAUSSIAN AVERAGES

Let us consider the (normalized) Gaussian average

$$\langle x_{i_1} x_{i_2} \cdots x_{i_{2k}} \rangle = \frac{1}{Z} \int_{R^n} dx_1 \cdots dx_n \; x_{i_1} x_{i_2} \cdots x_{i_{2k}} \exp\left(-\frac{1}{2} \sum_{i,j=1}^{n} x_i A_{ij} x_j\right), \tag{1}$$

in which A is a real symmetric positive matrix. The normalization is such that $\langle 1 \rangle = 1$, i.e.

$$Z = (2\pi)^{n/2} (\det A)^{-1/2}. \tag{2}$$

Wick theorem gives this Gaussian average as follows

- One chooses one pairing of the variables $x_{i_1} x_{i_2} \cdots x_{i_{2k}}$. Let us consider for instance the pairing
$$[x_{i_1} x_{i_l}] \cdots [x_{i_m} x_{i_{2k}}];$$
- to a pair $[x_i x_j]$ one associates the factor $(A^{-1})_{ij}$;
- then
$$[x_{i_1} x_{i_l}] \cdots [x_{i_m} x_{i_{2k}}] = A^{-1}_{i_1 i_l} \cdots A^{-1}_{i_m i_{2k}};$$
- one sums over the successive results of the
$$(2k-1)(2k-3) \cdots 1$$
possible pairings.

Example.

$$\langle x^4y^2\rangle = \frac{1}{Z}\int_{R^2} dx\, dy\, x^4y^2 \exp\left(-\frac{1}{2}(x^2+2xy+2y^2)\right).$$

Thus

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}\ A^{-1} = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$$

i.e. the pairings are

$$[xx] = 1, \quad [xy] = -1, \quad [yy] = 2.$$

Then

$$\langle x^4y^2\rangle = 3[xx]^2\cdot[yy] + 12[xx]\cdot[xy]^2 = 18.$$

The derivation is elementary:

$$\begin{aligned}&\langle x_{i_1}x_{i_2}\cdots x_{i_{2k}}\rangle \\ &= \frac{\partial}{\partial b_{i_1}}\cdots\frac{\partial}{\partial b_{i_{2k}}}\frac{1}{Z}\int_{R^n} dx_1\cdots dx_n \exp\left(-\frac{1}{2}\sum_{i,j=1}^{n} x_iA_{ij}x_j + \sum_{1}^{n} b_ix_i\right)\Big|_{b_i=0}.\end{aligned}$$

Therefore

$$\begin{aligned}\langle x_{i_1}x_{i_2}\cdots x_{i_{2k}}\rangle &= \frac{\partial}{\partial b_{i_1}}\cdots\frac{\partial}{\partial b_{i_{2k}}}\exp\left(\frac{1}{2}\sum_{i,j=1}^{n} b_iA^{-1}_{ij}b_j\right)\Big|_{b_i=0} \\ &= \frac{1}{2^k k!}\frac{\partial}{\partial b_{i_1}}\cdots\frac{\partial}{\partial b_{i_{2k}}}\left(\sum_{i,j=1}^{n} b_iA^{-1}_{ij}b_j\right)^k.\end{aligned}$$

The derivations do pair the indices as announced and it is elementary combinatorics to verify that the prefactor $1/2^k k!$ is cancelled.

2.2. WICK THEOREM AND FEYNMAN DIAGRAMS

As a first elementary example of application let us consider the perturbative series for the integral

$$Z(g) = \int_{R^n} dx_1\cdots dx_n \exp\left(-\frac{1}{2}\sum_{i,j=1}^{n} x_iA_{ij}x_j + g\sum_{1}^{n} x_i^4\right).$$

Then

$$\frac{Z(g)}{Z(0)} = \sum_{k=0}^{\infty}\frac{g^k}{k!}\left\langle\left(\sum_{1}^{n} x_i^4\right)^k\right\rangle$$

in which the brackets denote the Gauswsian expectation values, which may be computed through Wick theorem. It is convenient to represent x_i^4 by four lines carrying an index i, meeting at a vertex. Thus, at order k, there are k vertices. Wick theorem connects in all possible way those vertices by lines associated to the pairing of the x's. A line connecting a vertex x_i^4 and a vertex x_j^4 will thus carry a factor $(A^{-1})_{ij}$. To each pairing implied par Wick theorem one thus associates a diagram following these rules. Note that those diagrams may be connected (by the lines) or disconnected.

It is a simple exercise to show that the logarithm of Z is the sum of those diagrams which are connected:

$$\log \frac{Z(g)}{Z(0)} = \sum_{k=1}^{\infty} \frac{g^k}{k!} \Big\langle \Big(\sum_1^n x_i^4\Big)^k \Big\rangle_{\text{connected}}.$$

2.3. WICK THEOREM FOR INTEGRALS OVER MATRICES

A simple generalization of the above diagrammatic rules is needed for integrals over matrices. Let us consider for instance the integral

$$Z(g,N) = \int dX \exp -N \operatorname{Tr}(X^2 - gX^n) \tag{3}$$

in which the measure dX is the Euclidean measure over the N^2 matrix of the $N \times N$ Hermitian matrix X

$$dX = C \prod_{i=1}^{N} dX_{ii} \prod_{1 \le i < j \le N} d\Re X_{ij} d\Im X_{ij} \tag{4}$$

and the normalization constant C is chosen such that

$$Z(0,N) = 1. \tag{5}$$

Expanding again in powers of g

$$Z(g,N) = \sum_{k=0}^{\infty} \frac{(Ng)^k}{k!} \langle (\operatorname{Tr} X^n)^k \rangle \tag{6}$$

in which the bracket stand again for the Gaussian expectation value. Those averages satify again a Wick theorem with a pairing

$$[X_{ij} X_{kl}] = \langle X_{ij} X_{kl} \rangle = \int dX \, X_{ij} X_{kl} \exp -N \operatorname{Tr} X^2 = \frac{1}{N} \delta_{il} \delta_{jk}. \tag{7}$$

The vertices are now of the form $gN X_{i_1 i_2} X_{i_2 i_3} \dots X_{i_{n-1} i_n}$, and they are represented by n double lines which carry successively the pairs of indices $i_1 i_2$, etc. These doubles lines are connected by the pairing (7) of double lines.

Again $\log Z(g,N)$ is the sum of connected diagrams constructed with the previous rules.

2.4. T'HOOFT TOPOLOGICAL EXPANSION

The diagrams constructed from the previous matrix rules carry a dependence in N, the size of the matrices, which come from three different sources

- Each vertex carry a factor gN. Thus a diagram with V vertices receives a factor N^V from there.
- Each propagator for a double line connecting two vertices carries a factor N^{-1} from (7). Therefore a diagram with I lines connecting the vertices receives a factor N^{-I} from those lines.
- Each index carried by a line is conserved at a vertex. Therefore internal indices make closed loops. Each loop may be viewed as enclosing a face of a polyhedron. The sum over the N possible values taken by the internal index which runs around a closed loop, gives a factor N. Therefore if this polyhedron consists of F faces, the contribution of the sums over internal indices is proprotional to N^F.
- The combined dependences in N gives for the graph $N^{V-I+F} = N^{2-2h}$ in which h is the number of handles of the surface (related to the Euler character by $\chi = 2-2h$). Therefore the leading diagrams for large N are all those which are planar ($h=0$); the diagrams which may be drawn on a one-torus ($h=1$) are suleading by a factor $1/N^2$, etc.

2.5. RELATION TO POLYAKOV DISCRETIZED BOSONIC STRING

In the discretized version of Polyakov's bosonic string [17] the partition function is a sum over random surfaces of arbitrary genera and area; for given genus h and area A one sums over all triangulations $T(h,A)$, with the number of cells, i.e. triangles or squares or polygons, fixed by the area and the Euler character by the genus. Once the triangulation is given one performs the sum over the matter field X, a conformally invariant field of central charge c. The discretized partition function is thus

$$Z(\mu,\lambda) = \sum_{h=0}^{\infty} e^{-\mu h} \sum_{A=1}^{\infty} e^{-\lambda A} \sum_{T(h,A)} \int DX \exp -\beta \sum_{\langle ij\rangle} (Xi - Xj)^2 \tag{8}$$

in which $\langle ij\rangle$ restricts the sum to neighboring (dual) sites of the triangulation.

This whole sum may be mapped into an integral [11–13] over $N \times N$ Hermitian matrices X_a, $a = 1,\dots,p$. If we compute the matrix partition function

$$\zeta(N,g) = \int \prod_{\alpha=1}^{p} dX_a \exp[-\tfrac{N}{g} V(X_a)] \tag{9}$$

one recovers the string partition function $Z(\mu,\lambda)$ as

$$Z(\mu,\lambda) = \frac{1}{N^2} \operatorname{Log} \zeta(N,g) \tag{10}$$

with N related to the string coupling constant and g to the cosmological constant by:

$$e^{-\mu} = \frac{1}{N^2} \quad \text{and} \quad e^{-\lambda} = g. \tag{11}$$

For pure gravity (no matter field), one may take one matrix only with

$$V(X) = \mathrm{Tr}[\mathrm{X}^2 - \tfrac{1}{12}\mathrm{X}^4] \tag{12}$$

which corresponds to a "triangulation" by squares rather than by triangles, but it is indeed equivalent in the continuum limit to an X^3-triangulation, as expected and as it should for consistency, and it leads to simpler equations.

Remarks. 1. The numerical coefficients in (12) such as $1/12$ are chosen for later convenience; note also that if one rescaled X by a factor $g^{1/2}$ the coupling constant would be in front of the X^4.

2. The action (12) is unbounded below: the integral (9) is meaningless; it is simply a bookkeeping of the perturbation expansion.

Within one matrix models, for a more general probability distribution, we shall consider

$$P(X) = \frac{1}{\zeta}\exp\left[-\tfrac{N}{g}V(X)\right] \tag{13}$$

in which V is a polynomial of degree $2k$. We shall see that one can "tune" by an appropriate choice of the coefficients of V, multicritical points [12] of order $2, 3, \dots, k$. The k-th multicritical point corresponds to a $(2, 2k-1)$ minimal conformal matter field coupled to gravity, with central charge

$$c = 1 - \frac{3(2k-3)^2}{2k-1}. \tag{14}$$

Pure gravity is the $k = 2$, i.e. $c = 0$, case.

2.6. SCALING LAWS

The replacement of the string partition function by an integral over a matrix, or a finite number of matrices, is still far from a complete solution of the problem. It is easy to recover the planar limit of string perturbation theory: it is the large N limit of the matrix integral, which can easily be obtained by saddle-point techniques. It is more cumbersome, but possible, to expand in powers of $1/N^2$ to include higher genera. However the main interest of this matrix representation is to exhibit a continuum limit of the theory, the so-called "double scaling limit", in which the coupling constants, here N and g, are simultaneously tuned to criticality. In this limit it is expected, and one does show, that there is universality at long distance, independently of the specific discretization of the manifold. Although this double scaling limit has been discovered within matrix models, it was implicitly present in the work of Knizhnik, Polyakov, Zamolodchikov [14], extended through the Liouville theory, to arbitrary topologies by David [15] and Distler–Kawai [16]. They proved that the sub-partition function $Z_{h,A}$, for manifolds of given genus h and area A, behaves, for A large compared to the regularization (squared) length scale, as

$$Z_{h,A} \simeq a_h e^{\lambda_c A} A^{\gamma_h - 3} \tag{15}$$

with

$$\gamma_h = \gamma_0 + h(2 - \gamma_0) \tag{16}$$

and

$$\gamma_0 = 2 - \frac{1}{12}\left[25 - c + \sqrt{(1-c)(25-c)}\,\right] \tag{17}$$

in which c is the central charge of the conformal unitary matter field (which is smaller than one).

The non-universal exponential factor is expected; in a lattice approach it is related to the number of plaquettes which meet at one link and, very much as for random walks of given length, the interesting factor is the universal power of A.

Setting (15) into the string partition function

$$Z(\lambda, \mu) = \sum_h e^{-\mu h} \int_0^\infty dA e^{-\lambda A} Z_{h,A} \tag{18}$$

we first find that the continuum theory is recovered only when $\lambda$is in the vicinity of λ_c (with $\lambda > \lambda_c$), since the mean area $\frac{\partial \log Z}{\partial \lambda}$ is singular at λ_c. From the asymptotic formulae we extract the singular part of the string partition function under the scaling form

$$Z(\lambda, \mu) = e^{-\mu} F(x) \tag{19}$$

with

$$x = (\lambda - \lambda_c) \exp\left[\mu/(2-\gamma_0)\right]. \tag{20}$$

Note that the scaling theory does not fix, or constraint, the function $F(x)$ in any way.

The continuum limit is thus given by a double limit in which λ is close to λ_c, μ is large and x is fixed. Trading λ and μ for the matrix variables g and N according to (11), and with the explicit result (17) for γ_0, we have to verify that for a matrix model representing a unitary minimal model coupled to gravity, $\log \zeta$ is a function of the scaling variable

$$x \sim N^{1-\sqrt{(1-c)(25-c)}}\,(g_c - g) \tag{21}$$

in the large-N, small-$(g_c - g)$ limit.

The DDK derivation holds for unitary minimal models, since it assumes that the operator with the lowest dimension has dimension zero. For non unitary models one does have operators with negative dimensions, and the DDK derivation has to be repeated with the appropriate modification. This has the effect of replacing the KPZ-DDK exponent by

$$2 - \gamma_0 = \frac{25 - c + \sqrt{(25-c)(1-c+24D)}}{12(1-\Delta)} \tag{22}$$

in which Δ is the lowest (negative) dimension. For a (p, q) model with

$$c = 1 - 6\frac{(p-q)^2}{pq}, \tag{23}$$

the Kac table of dimensions is given by

$$\Delta_{r,s} = \frac{(rp - sq)^2 - (p-q)^2}{4pq}, \tag{24}$$

whose minimum is

$$\Delta = \frac{1-(p-q)^2}{4pq}$$

and we obtain

$$\gamma_0 = -\frac{2}{p+q-1}. \tag{25}$$

Therefore the scaling exponents are functions of $p+q$ alone, and not p and q separately, and one has thus to be careful in the identification of the matter content at multicritical points.

3. Matrix integrals in the large N limit

Before studying the scaling limit (often called the "double" scaling limit because it will require a simultaneous tuning of N and g) let us consider the large N limit of the matrix integrals, which corresponds to planar diagrams. We begin with one matrix integrals of the type

$$\zeta = \int dX \exp[-N\mathrm{TrV(X)}] \tag{26}$$

in which V is an (even) polynomial of degree $2k$. The unitary invariance of the action and of the measure dX allows us to trace out immediately the unitary degrees of freedom of X: writing

$$X = U\Lambda U^{\dagger} \tag{27}$$

in which Λ is a real diagonal matrix $\Lambda = (\lambda_1, \lambda_2, \ldots, \lambda_N)$ and U is unitary, we can integrate out U, yielding an unimportant factor that we normalize to unity. The representation (27) gives a Jacobian which is a function of the λ's:

$$dX = dU d\lambda_1 d\lambda_2 \cdots d\lambda_N \; J(\Lambda). \tag{28}$$

It is easy to prove that

$$J(\Lambda) = \Delta^2(\Lambda) \tag{29}$$

with

$$\Delta(\Lambda) = \prod_{i>j}(\lambda_i - \lambda_j) = \det \begin{pmatrix} 1 & \lambda_1 & \ldots & \lambda_1^{N-1} \\ 1 & \lambda_2 & \ldots & \lambda_2^{N-1} \\ \ldots & \ldots & \ldots & \ldots \\ 1 & \lambda_N & \ldots & \lambda_N^{N-1} \end{pmatrix}, \tag{30}$$

but it is immediate to convince oneself of this result by the following argument: the parametrization (27) is singular iff two eigenvalues coincide. Therefore $J(\Lambda)$ is proportional to a power of $\Delta(\Lambda)$. A simple dimensional counting shows that if we give dimension one to all the matrix elements of X, the dimension of J is $N(N-1)$, and thus J is proportional to Δ^2, and again we normalize the measure so that the proportionality constant is one. We end up with an integral over N eigenvalues:

$$\zeta = \int d^N \Lambda \Delta^2(\Lambda) \exp\Big[-N \sum_{i=1}^{N} V(\lambda_i)\Big]. \tag{31}$$

Note that we have integrated most of the N^2 degrees of freedom since we are left with an integral over only N eigenvalues. This is a pre-requisite for any solution of matrix models; when one cannot integrate out the unitary group, one does not know how to solve the model.

Let us recover first the lowest order of string perturbation theory, the planar limit [18] that we recover here by letting N go to infinity. We introduce the density of eigenvalues

$$\rho(\lambda) = \frac{1}{N} \sum_{i=1}^{N} \delta(\lambda - \lambda_i) \tag{32}$$

and, taking into account the Jacobian, we obtain an effective action

$$S_{\text{eff}} = N^2 \Big[\int \rho(\lambda) V(\lambda)\, d\lambda - \int \rho(\lambda)\rho(\lambda') \log|\lambda - \lambda'|\, d\lambda\, d\lambda'\Big]. \tag{33}$$

The expression (33) makes it clear that the eigenvalues would all lie at the bottom of the well provided by V, if they did not experience a logarithmic repulsion coming from the measure. The resulting distribution is a compromise between these two antagonistic factors. In the large N limit, the distribution ρ approaches a continuum function which is an extremum of S_{eff}; varying ρ (with a Lagrange multiplier for the constraint $\int \rho(\lambda)\, d\lambda = 1$), and differentiating once with respect to λ, we arrive at the integral equation

$$\frac{1}{2} V'(\lambda) = P \int d\lambda' \, \frac{\rho(\lambda')}{\lambda - \lambda'}. \tag{34}$$

Introducing the function

$$f(z) = \int d\lambda' \frac{\rho(\lambda')}{z - \lambda'} \tag{35}$$

which is analytic in a cut plane along the support of ρ we see that its real part on the cut is equal to $\frac{1}{2}V'$ and its imaginary part for $z = \lambda + i\epsilon$ is $-i\pi\rho(\lambda)$. It is immediate to verify that the solution is

$$f(z) = \frac{1}{2} V'(z) - P(z)\sqrt{z^2 - a^2} \tag{36}$$

in which P is a polynomial. (We limit here ourselves to even potentials: the cut is the symmetric interval $(-a, +a)$. For some potentials the cut will break up into several segments. We do not discuss these cases, but the method is staightforward; they can lead to different string equations but the matter content of the corresponding "gravity" is unclear).

The polynomial P and the parameter a are uniquely determined by demanding that $f(z)$ falls-off for large z as $1/z$. Through straightforward algebra we find that if the potential is given by

$$V(\lambda) = \sum_{m=1}^{k} \frac{g_m}{2m} \lambda^{2m}, \tag{37}$$

then P is the even polynomial of degree $2k-2$ given by

$$P(\lambda) = \frac{1}{2} \sum_{m=1}^{k} g_m \sum_{n=0}^{m-1} \binom{2n}{n} \left(\frac{a^2}{4}\right)^n \lambda^{2m-2n-2} \tag{38}$$

and a, the end point of the spectrum, is given by the algebraic equation

$$\frac{1}{2} \sum_{m=1}^{k} g_m \binom{2m}{m} \left(\frac{a^2}{4}\right)^m = 1. \tag{39}$$

Finally, the density of eigenvalues is given by

$$\rho(\lambda) = \frac{1}{\pi} P(\lambda) \sqrt{a^2 - \lambda^2} \tag{40}$$

and from there we obtain the string partition function in the large N limit in which we keep only planar surfaces, as

$$\frac{1}{N^2} \log \zeta = \int \rho(\lambda) V(\lambda)\, d\lambda - \int \rho(\lambda)\rho(\lambda') \log |\lambda - \lambda'|\, d\lambda\, d\lambda' \tag{41}$$

with $\rho(\lambda)$ given by (40).

For pure gravity we take $V(\lambda) = \frac{1}{g}(\lambda^2 - \frac{1}{12}\lambda^4)$, and we find that the string partition function is singular at $g_c = 1$ with a singularity proportional to $(1-g)^{5/2}$ (it is easier to compute $\frac{\partial Z}{\partial g}$).

Note that the density of eigenvalues (40) vanishes as a square root at its end points $\pm a$, except if $P(a)$ vanishes, in which case $\rho(\lambda)$ would behave as $(a^2-\lambda^2)^{3/2}$ near the end of the cut. A little bit of algebra shows that if $g = g_c = 1$, then $a = 2$ and indeed $P(a) = 0$.

This suggests, following Kazakov [11], that one can obtain a higher critical behaviour with a density of eigenvalues which would behave as $(a^2-\lambda^2)^{k+1/2}$ near the end point. Clearly this requires that $P(a) = P'(a) = \cdots = P^{(k-1)}(a) = 0$ and $P(a) \neq 0$. Since for an even potential of degree $2k$, P is an even polynomial of degree $(2k-2)$, one can indeed choose the coefficients of the potential so that one can tune the required k-th multicritical point. Let us stress that the k-th multicritical behaviour can be obtained through many potentials of degree at least equal to $2k$; the lowest one of degree exactly equal to $2k$ is unique. For this multicritical of order k a computation of the partition function at that planar level gives

$$\gamma_0 = -\frac{1}{k} \tag{42}$$

which corresponds to (p, q) matter coupled to gravity with $p+q=2k+1$. A more precise identification of operators leads to the assignment $(2, 2k-1)$, i.e.

$$c = 1 - 3\frac{(2k-3)^2}{2k-1} \tag{43}$$

for these multicritical points of one-matrix models. Curiously the geometric fluctuations of the "triangulation" obtained by mixing vertices or order 4, 6, ..., i.e. surfaces made of squares, hexagons, ..., with appropriate weights, induce some matter field coupled to gravity.

Finally we note that the even character of the potential V results in an even density of eigenvalues with two identical singularities at $+a$ and $-a$. If the potential had no parity the support of the eigenvalues distribution would be some non-symmetric interval (b, a). By tuning the coefficients of the potential one can obtain a density $\rho(\lambda)$ which vanishes as $(a-\lambda)^{k+1/2}$ at one end and $(\lambda-b)^{k'+1/2}$ at the other end with $k \neq k'$. An even potential describes thus two identical singularities at $\pm a$, whereas for an arbitrary potential in general one obtains a unique singularity at one end-point, the other one remaining a square root. Keeping in mind this doubling due to the evenness of V, we see that there is no new critical behaviour to be expected from non-even potentials. We shall see now that beyond the simple planar limit the calculations simplify greatly with even potentials.

4. "Double" scaling limit

In this section we want to study the scaling limit appropriate to the study of randomly tiangulated surfaces of arbitrary genera. It consists of a limit in which N goes to infinity, the coupling constant g appraoches the singular point g_c discussed above, and $N^\nu(g-g_c)$ remains fixed and finite, with some exponent ν which depends on the "matter" content [19–21].

4.1. ORTHOGONAL POLYNOMIALS

The unitary invariance of the measure has allowed for a reduction to the integral (31) over N eigenvalues. The Vandermonde determinant of the measure may be written as a determinant of polynomilas instead of monomials. Indeed let us introduce an arbitrary set of polynomials $p_n(\lambda)$, in which the degree of p_n is n:

$$p_n(\lambda) = a_n\lambda_n + \text{lowerdegree} \tag{44}$$

(without loss of generality we can choose a_n positive).

It is straightforward to verify by forming appropriate linear combinations of the rows that

$$\det\begin{pmatrix} 1 & \lambda_1 & \dots & \lambda_1^{N-1} \\ 1 & \lambda_2 & \dots & \lambda_2^{N-1} \\ \dots & \dots & \dots & \dots \\ 1 & \lambda_N & \dots & \lambda_N^{N-1} \end{pmatrix} = \frac{1}{a_0 a_1 \dots a_{N-1}} \det\begin{pmatrix} p_0 & p_1(\lambda_1) & \dots & p_{N-1}(\lambda_1) \\ p_0 & p_1(\lambda_2) & \dots & p_{N-1}(\lambda_2) \\ \dots & \dots & \dots & \dots \\ p_0 & p_1(\lambda_N) & \dots & p_{N-1}(\lambda_N) \end{pmatrix} \tag{45}$$

Expanding the determinant as a sum over permutations, we arrive at

$$\zeta = \int d^N\Lambda\, \Delta^2(\Lambda) \exp\Big[-N\sum_{i=1}^{N} V(\lambda_i)\Big] = \Big(\frac{1}{a_0 a_1 \ldots a_{N-1}}\Big)^2 \sum_{P,Q} (-1)^{|P|+|Q|}$$
$$\times \int d\lambda_1 p_{P1}(\lambda_1) p_{Q1}(\lambda_1) e^{-NV(\lambda_1)} \cdots \int d\lambda_N p_{PN}(\lambda_N) p_{QN}(\lambda_N) e^{-NV(\lambda_N)} \quad (46)$$

Choosing now the, yet unspecified, polynomials p_n to be orthogonal with the measure $\exp(-NV)$:

$$\int d\lambda\, e^{-NV(\lambda)} p_n(\lambda) p_m(\lambda) = \delta_{n,m} \quad (47)$$

we end up with

$$\zeta = \frac{N!}{\big(\prod_{n=0}^{N-1} a_n\big)^2}. \quad (48)$$

We note that since the potential is even $p_n(-\lambda) = (-1)^n p_n(\lambda)$.

4.2. ORTHOGONAL POLYNOMIALS, A REPRESENTATION OF HEISENBERG ALGEBRA

All we need are the coefficients of highest degree of each p_n. The procedure which allows to determine them recursively is based on a simple Heisenberg algebra [22].

We first introduce an infinite matrix $Q_{n,m}$ defined by the relations

$$\lambda p_n(\lambda) = \sum_m Q_{n,m} p_m(\lambda). \quad (49)$$

Comparing the degrees of the two sides of (49) we conclude that $Q_{n,m}$ vanishes unless $m \leq n+1$. Furthermore the matrix Q is real and symmetric since

$$Q_{n,m} = \int d\lambda\, e^{[-NV(\lambda)]} p_n(\lambda) \lambda p_m(\lambda) \quad (50)$$

and, by parity of the p_n's

$$Q_{n,n} = 0. \quad (51)$$

Therefore the matrix Q consists only of two "diagonals" and is characterized conveniently by

$$Q_{n-1,n} = Q_{n,n-1} = \sqrt{R_n}\,. \quad (52)$$

Identifying the coefficients of λ_{n+1} in the two sides of (49) we obtain

$$R_n = \Big(\frac{a_{n-1}}{a_n}\Big)^2. \quad (53)$$

The simple identity

$$a_0 a_1 \cdots a_{N-1} = \frac{a_{N-1}}{a_{N-2}} \Big(\frac{a_{N-2}}{a_{N-3}}\Big)^2 \cdots \Big(\frac{a_1}{a_0}\Big)^{N-1} a_0^N \quad (54)$$

leads to

$$\log(\zeta N!) + 2N \log a_0 = \sum_{n=1}^{N-1} (N-n) \log R_n . \tag{55}$$

Therefore all we need are recursion formulae for the R_n's.

This is achieved by introducing in parallel with the multiplication operator (49) which leads to the matrix Q, a derivative operator. It turns out that it is (slightly) preferable to introduce it on the basis of the functions

$$\varphi_n(\lambda) = \exp\left[-\frac{1}{2} N V(\lambda)\right] p_n(\lambda) \tag{56}$$

which are orthonormal with a flat measure. We define now a matrix P

$$\frac{d}{d\lambda}\varphi_n(\lambda) = \sum_m P_{n,m}\varphi_m(\lambda) \tag{57}$$

The matrix P is a real antisymmetric matrix, since

$$P_{n,m} = \int d\lambda\, \varphi_m(\lambda) \frac{d}{d\lambda}\varphi_n(\lambda) = -\int d\lambda\, \varphi_n(\lambda)\frac{d}{d\lambda}\varphi_m(\lambda). \tag{58}$$

The matrices Q and P may be defined for any set of orthogonal polynomials. However it is only for the cases of measures which are the exponential of a polynomial that one finds the special properties that we will exploit in the scaling limit. Indeed in such cases we shall show that the matrix P consists of a finite number of diagonals, but contrary to Q, for which this property followed simply from its definition (49) for any measure, here it relies on the properties of the measure. From (58) we obtain

$$P_{n,m} = \int d\lambda \exp -NV(\lambda) p_m(\lambda) \left[p_n'(\lambda) - \frac{1}{2} N V'(\lambda) p_n(\lambda)\right]. \tag{59}$$

The first term vanishes for $n \leq m$, and therefore

$$P_{n,m} = -\frac{N}{2}[V'(Q)]_{n,m} \text{ for } m > n \tag{60}$$

(the matrix Q is a representation of the multiplication by λ; hence

$$\int d\lambda\, e^{-NV(\lambda)} p_m(\lambda) \lambda^s p_n(\lambda) = [Q^s]_{n,m},$$

and $\int d\lambda\, e^{-NV(\lambda)} p_m(\lambda) V'(\lambda) p_n(\lambda) = [V'(Q)]_{n,m}$). The relation (60), plus the antisymmetry of P, fixes this matrix completely.

A matrix with a non-zero upper diagonals above the principal diagonal and b lower ones, all others being zero, will be name as being of type $[a, b]$: i.e. M is of type $[a, b]$ if M_{ij} vanishes for $i > j + a$ or $i < j - b$. The matrix Q is $[1, 1]$, thus Q^s of type $[s, s]$, and then from (60) and the antisymmetry, P is of type $[2k-1, 2k-1]$, if $2k$ is the degree of the potential.

Finally from the definitions (49) and (57) follows the commutation relation

$$[Q, P] = 1 \tag{61}$$

which is central to the construction (it follows immediately from the differentiation with respect to λ of $\lambda\varphi_n(\lambda) = \sum_m Q_{n,m}\varphi_m(\lambda)$).

4.3. STRING EQUATIONS FOR ONE-MATRIX MODELS, A FIRST ELEMENTARY APPROACH

We begin with the pure gravity case and follow a gentle trail before going to general multicritical points. The (n, n) matrix element of (61) leads to the identity

$$Q_{n,n+1}P_{n+1,n} - Q_{n-1,n}P_{n,n-1} = \frac{1}{2}$$

i.e. (since $P_{1,0} = 0$)

$$Q_{n,n+1}P_{n+1,n} = \frac{n}{2},$$

or explicitely

$$\sqrt{R_n}[V'(Q)]_{n,n-1} = \frac{n}{N} \tag{62}$$

Taking $V(\lambda) = \frac{1}{g}[\lambda^2 - \frac{1}{12}\lambda^4]$ we have to compute the $(n, n-1)$ matrix element of Q^3, which is a sum of "walks" of three steps on a lattice of integers, from the initial site n to the final site $n-1$:

$$[Q^3]_{n,n-1} = \sqrt{R_n}(R_{n+1} + R_n + R_{n-1})$$

and thus, for this potential, the Heisenberg commutator (61) leads to

$$2R_n - \frac{1}{3}R_n(R_{n+1} + R_n + R_{n-1}) = \frac{ng}{N} \tag{63}$$

the recursion formula which fixes the successive R_n's.

We first set $n = N$ and recover the large N limit by neglecting the variations of R from N to $N \pm 1$:

$$2R_\infty - R_\infty^2 = g, \tag{64}$$

whose solution is

$$R_\infty = 1 - \sqrt{1-g} \tag{65}$$

(it is easy to check that the R_n's must vanish with g). We see that there is a singularity at $g_c = 1$.

We study now a vicinity of size $N^{-\alpha}$ of this singular point and explore a region of size $N^{-\alpha}$ as well for n/N. In other words we define

$$x = N^\alpha(1-g),$$
$$y = N^\alpha\left(1 - \frac{n}{N}\right) \tag{66}$$

with an exponent α which will be determined later. For such values of g and n, R_n which is a function of n, N and g, is a function of ng/N alone, i.e. of $N^{-\alpha}(x+y)$ and thus we can set $g=1$ or $x=0$ without loss of generality. We then make a scaling ansatz, compatible with (63),

$$R_n(g=1) = R(N,y) = 1 + N^{-\beta} f(y) \tag{67}$$

and return to (63). Dropping the sub-leading terms one obtains

$$N^{-2\beta} f^2 + \frac{1}{3} N^{2\alpha-\beta-2} f'' = N^{-\alpha} y \tag{68}$$

from which one fixes

$$\alpha = \frac{4}{5} \text{ and } \beta = \frac{2}{5} \tag{69}$$

with f solution of the "string equation"

$$f^2(y) + \frac{1}{3} f''(y) = y. \tag{70}$$

We now recover the string partition function: the singular part near g_c in the continuum limit is given by

$$Z_{\text{string}} = \frac{1}{N^2} \int_0^{N^{4/5}} dy\, y f(x+y) \tag{71}$$

and the singular part in the vicinity of $x=0$ satisfies indeed the scaling law with $\gamma_0 = -\frac{1}{2}$. The usual string perturbation theory is recovered by letting N go to infinity first, i.e. x or y go to infinity. The planar limit (65) corresponds to the asymptotic behaviour

$$f(y) \simeq -\sqrt{y} \tag{72}$$

for large y, and this fixes uniquely the asymptotic expansion of the solution of (70)—the Painlevé I equation—for large y

$$f(y) = \sum_{h=0}^{\infty} f_h y^{(1-5h)/2} \tag{73}$$

with $f_0 = -1$, $f_1 = \frac{1}{24}$, $f_2 = \frac{49}{1152}$, ... The equation (70) defines an asymptotic series which fixes "pertubatively" all the f_h's, but there does not seem to exist any global non-pertubative solution free of ambiguity. Indeed the matrix model that we have solved had an exponential of the type $\exp\left(\frac{+Ng}{12}\operatorname{Tr} X^4\right)$. Order by order in an expansion in powers of $1/N^2$, there is a circle of analyticity in the g-plane, $|g| < 1$, and therefore all the coefficients f_h are indeed well defined. However clearly the matrix integral does not define anything but an asymptotic expansion in increasing values of the genera.

4.4. A MORE SYSTEMATIC APPROACH TO THE SCALING LIMIT

For higher multicritical points one could proceed in the same way: (62) is still valid and since the potential is now of higher degree we would have to consider longer walks on the same lattice of integers. It is doable but there is a powerful algebraic machinery due to Gel'fand and Dikii [23], as pointed out by Gross and Migdal [21], and further elaborated by Douglas, which allows one to understand the mathematical structure underlying this continuum limit (usually referred to as the "double scaling limit").

The matrix Q being of a type $[1,1]$, flows in the continuum limit into a second order differential operator, whereas the matrix P which is of type $[2k-1, 2k-1]$ will flow, at the k-th multicritical point, into a differential operator of order $(2k-1)$:

$$Q \to 2 + N^{-2/(2k+1)}\left(\frac{d^2}{dy^2} + f(y)\right)$$
$$P \to N^{2/(2k+1)}\left(c_k \frac{d^{2k-1}}{dy^{2k-1}} + \text{lower degree}\right). \tag{74}$$

We should remember that Q and P are respectively even and odd under transposition, an operation which is mapped in the continuum limit to

$$y^T = y, \quad \left(\frac{d}{dy}\right)^T = -\frac{d}{dy}. \tag{75}$$

We now return to the Heisenberg commutator (61); a priori the commutator of a differential operator of degree n with a second operator of degree n' is a differential operator of degree $n + n' - 2$, whereas we know that it should be a constant, i.e. not only an operator of degree zero, but it does not even imply any multiplication by a function of y. This induces many constraints and in particular determines the string equation satisfied by f.

To convince oneself of this fact we return to the $k = 2$ example; the antisymmetry of P fixes the lower order terms up to one unknown function $g(y)$:

$$P = N^{2/5}\left(c_2 \frac{d^3}{dy^3} + g(y)\frac{d}{dy} + \frac{1}{2}g'(y)\right) \tag{76}$$

and the commutator equations lead first to

$$g(y) = \frac{3}{2}c_2 f(y), \tag{77}$$

then to

$$2ff' + \frac{1}{3}f''' = -\frac{4}{3c_2}, \tag{78}$$

the derivative of the string equation up to a normalization. (A tedious calculation yields in fact

$$c_k = (-1)^{k-1}\frac{(k!)^2}{(2k)!}2^{2k-1} \tag{79}$$

and in particular $c_2 = -\frac{4}{3}$ as it should). The general solution is more conveniently expressed through the Gel'fand–Dikii formalism of pseudo-differential operators [23]. One defines the "Laurent" expansion of the operator $(\partial^2 + f(y))^{k-1/2}$ in which ∂ stands for $\frac{d}{dy}$, as an infinite series in powers of ∂ from $2k-1$ to $-\infty$ (with $\partial^{-1} g = \sum_1^\infty (-1)^{k-1} g^{(k-1)} \partial^{-k}$). Then one can define $\left(\partial^2 + f(y)\right)^{k-1/2}\big|_+$ as the non-negative powers of ∂ in this expansion:

$$\left(\partial^2+f(y)\right)^{k-1/2} = \left(\partial^2+f(y)\right)^{k-1/2}\big|_+ + R_k f \partial^{-1} + \partial^{-1} R_k f + \text{lower order in } \partial \quad (80)$$

(The form of the coefficient for ∂^{-1} is fixed by the odd character of $\left(\partial^2+f(y)\right)^{k-1/2}$ under transposition). Since $\partial^2 + f(y)$ commutes with $\left(\partial^2 + f(y)\right)^{k-1/2}$, we obtain from (80) that

$$\left[\partial^2 + f(y), \left(\partial^2 + f(y)\right)^{k-1/2}\big|_+\right] = -4\frac{d}{dy} R_k f. \quad (81)$$

Therefore, if we take $P = N^{2/(2k+1)} c_k \left(\partial^2 + f(y)\right)^{k-1/2}\big|_+$, we satisfy the commutator equation (61) provided $f(y)$ satisfies the string equation

$$-4c_k \frac{d}{dy} R_k f = 1 \quad (82)$$

and one can check that the large y, i.e. the planar limit of the theory, gives the final form for the string equation

$$R_k\{f\} = (-1)^k \frac{(2k)!}{2^{2k+1}(k!)^2}\, y. \quad (83)$$

Tedious, but straightforward calculations based on the definition (80) allow one to compute the functionals $R_k\{f\}$; the lowest ones are:

$$\begin{aligned} R_1\{f\} &= \frac{1}{4} f \\ R_2\{f\} &= \frac{3}{16}\left(f^2 + \frac{1}{3} f''\right). \end{aligned} \quad (84)$$

In the large-N, or large-y, limit one verifies easily that the solution to the string equation behaves as $y^{1/k}$, and following the same steps as for $k = 2$, we obtain that the string susceptibility exponent is given by $\gamma_0 = -1/k$ at the k-th multicritical point.

4.5. CORRELATION FUNCTIONS AND KdV FLOWS

Correlation functions are defined by loops inserted in the random surface. In the discretized random surface the insertion a microscopic loop of length p is given by the expectation value of $\mathrm{Tr}(X^p)$ in a matrix integral: indeed this adds a vertex with p "legs", i.e. a polygon of perimeter p in the dual diagram (remember that the Feynman graph is dual to the triangulated surface). We have thus to deal with expectation values of the type $\langle \mathrm{Tr}(X^{p_1})\,\mathrm{Tr}(X^{p_2})\cdots\mathrm{Tr}(X^{p_n})\rangle$. They can be generated by calculating derivatives of the matrix partition function z in the presence of source terms of the type added to the potential V of the form $\delta V = t_p\mathrm{Tr}(\mathrm{X^p})$. We thus consider the effect of changing

$$V \to V + \delta V \tag{85}$$

in the integral which gave the partition function $\zeta\{V\}$.

This has the effect of modifying the measure and thus the orthogonal polynomials p_n, and correspondingly the matrices Q and P of (49) and (57). If $p \to p_n + \delta p_n$, one can expand δp_n on the basis of the p_n themselves:

$$\delta p_n = \sum_{m=1}^{n} \omega_{n,m} p_m \tag{86}$$

in which ω is an infinitesimal matrix. Returning to the definition (49) of Q and varying V we obtain

$$\lambda \delta p_n(\lambda) = \sum_{m=1}^{n} \delta Q_{n,m} p_m + \sum_{m=1}^{n} Q_{n,m} \delta p_m. \tag{87}$$

The r.h.s. gives $(\delta Q + Q\omega)_{n,m} p_m$, whereas the l.h.s. is equal to $(\omega Q)_{n,m} p_m$. It follows that

$$\delta Q = [\omega, Q]. \tag{88}$$

Similarly the variation of the orthogonality relation of the p_n's leads to the relation

$$\omega + \omega^T - \delta V(Q) = 0, \tag{89}$$

and we thus introduce the antisymmetric matrix Ω

$$\Omega = \omega - \frac{1}{2}\delta V(Q), \tag{90}$$

from which we also have

$$\delta Q = [\Omega, Q]. \tag{91}$$

The antisymmetry of Ω ensures that Q remains symmetric. Similarly, returning to $\varphi_n = p_n \exp\left(-\frac{1}{2}V\right)$, one finds, after variation

$$\delta \varphi_n = \sum \Omega_{n,m} \varphi_m,$$

and varying the definition of P, $\varphi'_n = \sum P_{n,m}\varphi_m$, one finds $\delta\varphi'_n = \sum(\Omega P)_{n,m}\varphi_m = \sum \delta P_{n,m}\varphi_m + \sum(P\Omega)_{n,m}\varphi_m$, i.e.

$$\delta P = [\Omega, P], \tag{92}$$

from which P remains antisymmetry. The Jacobi identity and (91, 92) ensures that the Heisenberg commutation relation (61) is preserved under variation.

Near a multicritical point Q is, up to a normalization, of the form $\partial^2 + f(y)$, and (91) is thus an equation for δf. Let us show that the matrix Ω which governs the flow, is an antisymmetric operator of finite type, provided δV is a polynomial. Indeed, from its definition (86), ω is a lower triangular matrix; thus above the diagonal $\Omega_{n,m} = -\frac{1}{2}[\delta V(Q)]_{n,m}$ for $m > n$, which vanishes if $(m-n)$ is larger than the degree of the polynomial δV. Therefore (91) implies that the commutator with Q of the antisymmetric operator of finite type Ω, is an operator of degree zero. The solution fixes again Ω to be proportional to $Q^{p+1/2}|_+$. Therefore the flow equation induced by the scaling operator corresponding to the insertion of a loop of perimeter p takes (up to normalizations) the form []

$$\frac{\partial f}{\partial t_p} = \frac{d}{dy} R_{p+1}\{f\}. \tag{93}$$

The set of commuting flows induced on Q and P by the various δV are well-known in the mathematical literature under the name of KdV (Korteweg-de Vries) flows [].

5. Matter field and multimatrix models

The one matrix models describe only, at multicriticality, matter fields of central charge (14) coupled to gravity. This does not include interesting models, such as minimal unitary models, for which $c = 1 - \frac{6}{p(p+1)}$ (except for $c = 0$, and -2). For such models of matter coupled to gravity, several matrices are needed.

5.1. AN EXAMPLE: ISING MATTER COUPLED TO GRAVITY

Let us consider for definiteness the minimal unitary case with $c = 1/2$, i.e. Ising spins coupled to gravity [26]. To each cell of the triangulation one associates a "spin" σ, which can be up or down. For a given triangulation T one first compute the Ising partition function

$$Z_T(\beta) = \sum_{\sigma=\pm 1} \exp\left[\beta \sum_{\langle i,j\rangle} (\sigma_i \sigma_j - 1) - B \sum_i \sigma_i\right] \tag{94}$$

with a coupling constant β and an external magnetic field B. The bracket $\langle i,j\rangle$ restricts the interaction to adjacent cells sharing one link. One then compute the full matter-gravity partition function by summing

$$Z(\mu, \lambda, \beta) = \sum_{h=0}^{\infty} e^{-\mu h} \sum_{A=1}^{\infty} e^{-\lambda A} Z_T(\beta). \tag{95}$$

The whole sum (95) may be mapped into an integral over two $N \times N$ Hermitian matrices X_+ and X_-:

$$\begin{aligned}&\zeta(N, g_+, g_-, a)\\ &= \int dX_+ dX_- \exp\left(-N \operatorname{Tr}\left[\frac{1}{2}X_+^2 - aX_+X_- + \frac{1}{2}X_-^2 - g_+ X_+^3 - g_- X^{-3}\right]\right).\end{aligned} \tag{96}$$

Indeed, if one expands $\log\zeta$ in powers of $1/N^2$, g_+ and g_-:

$$\frac{1}{N^2}\log\zeta = \sum_{h,V_+,V_-}\left(\frac{1}{N^2}\right)^h (g_+)^{V_+}(g_-)^{V_-}\, d_{h,V_+,V_-}, \tag{97}$$

the coefficient d_{h,V_+,V_-} is the sum of connected Feynman diagrams of genus h, with V_+ and V_- vertices of type ${X_+}^3$ and ${X_-}^3$ respectively. The area of the discretized manifold is thus equal to $V_+ + V_-$. One particular diagram is made of propagators obtained by inverting the matrix of the quadratic part of the action (96), i.e.

$$\begin{pmatrix} 1 & -a \\ -a & 1 \end{pmatrix}^{-1} = \frac{1}{1-a^2}\begin{pmatrix} 1 & a \\ a & 1 \end{pmatrix}. \tag{98}$$

This leads to the Wick pairings

$$\begin{aligned} \langle X_+X_+\rangle &= \langle X_-X_-\rangle = \frac{1}{1-a^2} \\ \langle X_+X_-\rangle &= \frac{a}{1-a^2}. \end{aligned} \tag{99}$$

Therefore if one associates to a vertex a variable σ which takes the value $+1$ for a vertex ${X_+}^3$ and -1 for ${X_-}^3$, the propagator between two vertices of respective type σ and σ' is simply

$$\frac{a^{(1-\sigma\sigma')/2}}{1-a^2}.$$

Noting that the number of propagators is equal to $\frac{3}{2}(V_+ + V_-)$, we obtain the contribution of one particular diagram as

$$(g_+)^{V_+}(g_-)^{V_-}\left(\frac{1}{1-a^2}\right)^{3/2(V_++V_-)} a^{\sum_{i,j}(1-\sigma_i\sigma_j)/2}$$

and thus

$$d_{h,V_+,V_-} = (g_+)^{V_+}(g_-)^{V_-}\left(\frac{1}{1-a^2}\right)^{3/2(V_++V_-)} \sum_{\{\sigma\}} a^{\sum_{i,j}(1-\sigma_i\sigma_j)/2}.$$

It follows that the string partition function is equal to $N^{-2}\log\zeta$, provided we make the following identification

$$\begin{aligned} e^{-\ mu} &= \frac{1}{N^2} \\ e^{-\lambda-B} &= g_+(1-a^2)^{-3/2} \\ e^{-\lambda+B} &= g_-(1-a^2)^{-3/2} \\ e^{-2\beta} &= a. \end{aligned} \tag{100}$$

5.2. CHAIN OF COUPLED MATRICES

The Ising matter led to a model with two matrices, coupled in a probability distribution of the form

$$P(X_+, X_-) = \frac{1}{\zeta} \exp\left(-N\mathrm{Tr}[\mathrm{V_+(X_+) + V_-(X_-) - aX_+X_-}]\right). \tag{101}$$

More generally, let us consider the problem of an open chain of matrices with nearest neighbour coupling. Consider p matrices X_a, $a = 1, \dots, p$, and the distribution of probability

$$P(X_1 \cdots X_p) = \frac{1}{\zeta} \exp\Big(-N\,\mathrm{Tr}\Big[\sum_{l=1}^{p} V(X_l) + \sum_{l=1}^{p-1} X_l X_{l+1}\Big]\Big). \tag{102}$$

As with one matrix models, the key to the solution comes from the fact that it is possible again here to integrate out the unitary degrees of freedom. Writing

$$X_l = U_l \Lambda_l U_l^\dagger \tag{103}$$

in which the Λ_l's are diagonal, yields first a Jacobian $\prod_{l=1}^{p} \Delta^2(\Lambda_l)$; ($\Delta(\Lambda)$ is the Van der Monde determinant associated to the eigenvalues λ_α of the matrix X). However there is still a dependence in the U_l's through the coupling between matrices in the distribution (102). Consider for instance the term $\mathrm{Tr}(X_1X_2) = \mathrm{Tr}\big[\Lambda_1 U_1^\dagger U_2 \Lambda_2 (U_1^\dagger U_2)^\dagger\big]$; we still have to integrate over the relative unitary transformation $U = U_1^\dagger U_2$. This is done through a remarkable formula due to the mathematician Harish-Chandra [27] and rediscovered in this context by Itzykson and Zuber [28]:

$$\int dU \, \exp N\,\mathrm{Tr}\big[\Lambda_1 U \Lambda_2 U^\dagger\big] = C\,\frac{\det_{i,j=1,\dots,N} \exp\left(N(\lambda_1)_i(\lambda_2)_j\right)}{\Delta(\Lambda_1)\Delta(\Lambda_2)} \tag{104}$$

in which the numerator is the determinant of an $N \times N$ matrix whose (i, j) matrix element is $\exp N(\lambda_1)_i(\lambda_2)_j$, and C is a normalization constant independent of the matrices Λ_1 and Λ_2. It turns out that this formula is one of the cases in which the semi-classical approximation is exact; namely, if we consider the stationnary points of the "action" $\mathrm{Tr}\big[\Lambda_1 U \Lambda_2 U^\dagger\big]$, it is elementary to verify that they consist of the $N!$ permutation matrices P. Expanding U near one of those permutations, namely writing $U = P\exp iH$, and expanding to order H^2, the Gaussian integration over H yields one of the $N!$ terms of the determinant in the numerator of (104). The sum over the P's reconstructs (104). Of course this is not a derivation and it does not seem very easy to follow this route to prove that higher orders in X do cancel; we shall give a simple derivation of this formula in the next section.

We now return to the matrix integral

$$\zeta = \int \prod_{l=1}^{p} d^{N^2} X_l \exp\Big(-N\,\mathrm{Tr}\Big[\sum_{l=1}^{p} V(X_l) + \sum_{l=1}^{p-1} X_l X_{l+1}\Big]\Big), \tag{105}$$

perform the change of variables (103) and integrate over the U_l's. For a given matrix X_l with $1 < l < p$, i.e. not located at one of the two ends of the chain, we collect a factor $\Delta^2(\Lambda_l)$ from the Jacobian, a factor $\Delta^{-1}(\Lambda_l)$ from the integration over the relative unitary transformation of X_l and X_{l+1} (from the denominator of (104)) and a second factor $\Delta^{-1}(\Lambda_l)$ from the integration over the relative unitary transformation of X_l and X_{l-1}. For the two matrices at the end points of the chain, which have only one neighbour, the integration over the unitary degrees of freedom yield only one power $\Delta^{-1}(\Lambda)$. We thus end up with an integral over pN variables (instead of pN^2):

$$\zeta = \int \prod_{l=1}^{p} (d^N \Lambda_l) \Delta(\Lambda_1)\Delta(\Lambda_p) \exp\left[-N \operatorname{Tr} \sum_{l=1}^{p} V_l(\Lambda_l)\right] \prod_{l=1}^{p-1} \det\{\exp N(\lambda_l)_i(\lambda_{l+1})_j\}. \tag{106}$$

The antisymmetry under permutations of $\Delta(\Lambda_1)$ implies that the $N!$ terms of the expansion of the determinant $\det\{\exp N(\lambda_1)_i(\lambda_2)_j\}$ yield identical contributions; we can thus keep only the diagonal term $\exp \sum_{i=1}^{N} (\lambda_1)_i(\lambda_2)_i$ and multiply by $N!$. Similarly, the $N!$ terms resulting of the expansion of the determinant $\det\{\exp N(\lambda_2)_i(\lambda_3)_j\}$ give equal integrals, as can be seen from a simultaneous identical permutation of the $(\lambda_1)_i$'s and $(\lambda_2)_i$'s. Dropping irrelevant multiplicative constants this leads to the final representation

$$\zeta = \int \prod_{l=1}^{p} (d^N \Lambda_l) \Delta(\Lambda_1)\Delta(\Lambda_p) \exp\left(-N \operatorname{Tr}\left[\sum_{l=1}^{p} V_l(\Lambda_l) + \sum_{l=1}^{p-1} \Lambda_l \Lambda_{l+1})\right]\right). \tag{107}$$

Following Mehta [29] and Douglas [24], one can introduce two sets of polynomials $p_n(\lambda_1)$ and $\pi_n(\lambda_p)$, orthogonal with respect to a multilocal measure:

$$\int d\lambda_1 d\lambda_2 \cdots d\lambda_p p_n(\lambda_1)\pi_m(\lambda_p) \exp\left[-\sum_{l=1}^{p} V_l(\lambda_l) + \sum_{l=1}^{p-1} \lambda_l \lambda_{l+1}\right] = \delta_{n,m}. \tag{108}$$

Then, if a_n is the coefficient of λ_1^n in $p_n(\lambda_1)$ and b_n the coefficient of λ_p^n in $\pi_n(\lambda_p)$, it is straightforward to show that, as in the one-matrix integrals,

$$\zeta = \frac{N!}{\prod_0^{N-1} a_n b_n}. \tag{109}$$

Then we can introduce two Heisenberg algebras corresponding to multiplication by λ_1 and derivation with respect to λ_1 of the p_n's, or multiplication by λ_p and derivation with respect to λ_p of the π_n's, i.e. four infinite matrices Q_1, P_1, Q_2 and P_2, with $[Q_i, P_i] = 1$, $i = 1, 2$. Douglas has shown [24], that these four matrices are again of finite type, i.e. consist of a finite number of non-vanishing diagonals, their number depending of the degree of the polynomials V_l. However the matrices Q_i are no longer of $[1, 1]$ type: their type depends now upon the degree of the V's. The double scaling limit now consists for, say Q_1 and P_1, of two differential operators of respective degrees q and p, and the string equation follows from the commutator equation and the relation

$$P_1 = Q_1^{p/q}|_+ \tag{110}$$

corresponding to a critical massless matter field of central charge $c = 1 - 6\frac{(p-q)^2}{pq}$ coupled to gravity. In fact Douglas has also shown [24] that with two matrices only, by tuning appropriately the polynomials V_1 and V_2, one can generate arbitrary (p, q) models.

5.3. ONE-DIMENSIONAL CHAIN OF MATRICES AND FREE FERMIONS

In the discretized approach to the $c = 1$ string theory, a one-dimensional matter field is associated to each dual site of the triangulation, namely to the center of each polygon. The contribution of neighbouring plaquettes to the matter partition function is simply $\exp\left(-\beta(t-t')^2\right)$, if t and t' are the values of the matter field on those two cells. The Gaussian integration over this matter field is elementary, but useless: the determinant is a function of the triangulation and it is not easy to perform the sum over the triangulations. In the matrix model representation we have to introduce one $N \times N$ matrix for any value t of the matter field, i.e. a matrix function $X(t)$. The matrix integral becomes a one-dimensional path integral over the N^2 dynamical degrees of freedom $X_{ab}(t)$. The matrix propagator should reproduce the factor $\exp\left(-\beta(t-t')^2\right)$; one thus demands that

$$\langle X_{ij}(t) X_{kl}(t')\rangle = \frac{1}{N}\delta_{il}\delta_{jk}\; e^{-\beta(t-t')^2}. \tag{111}$$

In Fourier space (with respect to t) the propagator should be proportional to $\exp\left(-p^2/\beta\right)$. The momentum p is one-dimensional; we can thus expect that if one modifies the large momentum behaviour of this propagator, as long as the integrals over the momenta converge, the theory belongs to the same "class of universality". It is more convenient to replace $\exp\left(-p^2/\beta\right)$ by $\frac{1}{1+p^2/\beta}$, which correponds to the replacement of $\exp\left(-\beta(t-t')^2\right)$ by $\exp\left(-\sqrt{\beta}\,|t-t'|\right)$. The corresponding matrix model is then simply

$$\zeta(N, g, T) = \int D^{N^2} X(t) \exp\left(-N \,\mathrm{Tr} \int_0^T dt \left[\frac{1}{2}\dot{X}^2 + \frac{1}{2}X^2 + g_3 X^3 + g_4 X^4 + \cdots\right]\right) \tag{112}$$

which will generate triangles, squares, ...; the power of the expansion in powers of $1/N^2$ will select the genus and the matrix propagator in Fourier space is equal to $\dfrac{1}{N}\delta_{il}\delta_{jk}\dfrac{1}{p^2+1}$. Note that $1/N$ plays the role of $\hbar$. Any connected diagram is proportional to the "period" T and therefore the string partition function is given by

$$Z_{st} = \lim_{T\to\infty} \frac{\log \zeta}{TN^2} \tag{113}$$

and, as before, $1/N^2$ is identified to the string coupling constant and the cosmological constant is related to the g's.

The one-dimensional path integral (112) is nothing but the Feynman–Kac integral for the N^2 dynamical variables $X_{ij}(t)$. We can write instead a Schrödinger equation for a wave function Ψ_{ij}, with the Hamiltonian

$$H = -\frac{1}{2N^2}\nabla^2 + \mathrm{TrV(X)} \tag{114}$$

in which ∇^2 is the Laplacian with respect to the N^2 variables X_{ij} and V is the polynomial

$$V(\lambda) = \frac{1}{2}\lambda^2 + g_3\lambda^3 + g_4\lambda^4. \tag{115}$$

In the large T limit

$$\lim_{T\to\infty}\frac{\log\zeta}{T} = -E_0 \tag{116}$$

in which E_0 is the ground state energy of the Hamiltonian H.

Changing variables as before from the X_{ij}'s to the eigenvalues and unitary degrees of freedom:

$$X(t) = U(t)\Lambda(t)U^\dagger \tag{117}$$

the Laplacian in (114) becomes a second order differential operator with respect to the λ's and the U's [18]. The ground state wave function is a unitary singlet, and thus a function of the N eigenvalues λ_i alone. The Laplacian on the eigenvalues takes the form

$$\nabla^2 = \frac{1}{\Delta(\Lambda)}\sum_{i=1}^{N}\frac{\partial^2}{\partial\lambda_i^2}\Delta(\Lambda), \tag{118}$$

in which $\Delta(\Lambda)$ is the Vandermonde determinant $\Delta(\Lambda) = \prod_{1\le i<j\le N}(\lambda_i - \lambda_j)$.

For two singlet wave function Ψ_1 and Ψ_2 the scalar product

$$\langle\Psi_1|\psi_2\rangle = \int d^{N^2}X\,\overline{\Psi_1(X)}\Psi_2(X) \tag{119}$$

takes the form

$$\langle\Psi_1|\psi_2\rangle = \int d^N\Lambda\,\Delta^2(\Lambda)\overline{\Psi_1(\Lambda)}\Psi_2(\Lambda) \tag{120}$$

and therefore if we introduce the wave function

$$\chi(\Lambda) = \Delta(\Lambda)\Psi(\Lambda) \tag{121}$$

the scalar product becomes

$$\langle\chi_1|\chi_2\rangle = \int d^N\Lambda\,\overline{\chi_1(\Lambda)}\chi_2(\Lambda) \tag{122}$$

and the Schrödinger equation for the ground state reads

$$\Big\{\sum_{i=1}^{N}\Big[-\frac{1}{2N^2}\frac{\partial^2}{\partial\lambda_i^2} + V(\lambda_i)\Big]\Big\}\chi(\Lambda) = E_0\chi(\Lambda). \tag{123}$$

The ground state wave function $\Psi(\Lambda)$ is also invariant under permutation of the λ's (permutations are particular unitary transformations) and therefore $\chi(\Lambda)$ is totally antisymmetric. Hence the Schrödinger equation (123) describes N independent Fermions in the potential well $V(\lambda)$ [18].

Remark. One can derive the same result by a "Lagragian" path integral formulation, rather than this Hamiltonian derivation. Instead of working with a continuum time t, and matrices $X(t)$, one can discretize t and introduce a chain of matrices X_a as in the previous section. Then the representation (107), which relies on the Itzykson–Zuber formula, shows that if we introduced a transfer matrix, the eigenvalues would be independent, and the Vandermonde at the two ends would project onto the antisymmetric eigenstates of this transfer matrix.

Let us return to the discussion of the free fermion problem (123). If ε_a denotes the eigenvalues of the one-body Hamiltonian

$$\Big[-\frac{1}{2N^2}\frac{\partial^2}{\partial\lambda^2} + V(\lambda)\Big]\phi_a(\lambda) = \varepsilon_a\phi_a(\lambda), \tag{124}$$

the ground state energy E_0 is the sum of the N lowest eigenvalues

$$E_0 = \sum_1^N \varepsilon_a. \tag{125}$$

It is convenient to change scale $\lambda \to N^{-1/2}\lambda$ and to define

$$\tilde{V}(\lambda) = NV(N^{-1/2}\lambda). \tag{126}$$

Then if $\tilde{\varepsilon}_a$ are the eigenvalues of $-\frac{1}{2}\frac{\partial^2}{\partial\lambda^2} + \tilde{V}(\lambda)$, one introduces the density of eigenvalues

$$\rho(\tilde{\varepsilon}) = \sum_{a=1}^N \delta(\tilde{\varepsilon} - \tilde{\varepsilon}_a) \tag{127}$$

from which one has

$$-N^2 Z_{st} = \tilde{E}_0 = \int_0^\mu d\tilde{\varepsilon}\,\tilde{\varepsilon}\,\rho(\tilde{\varepsilon}) \tag{128}$$

in which we integrate up to a Fermi level μ defined by the condition that we have N fermions

$$N = \int_0^\mu d\tilde{\varepsilon}\,\rho(\tilde{\varepsilon}) \tag{129}$$

In the large N limit most terms in the sum (125) correspond to a high quantum number, and they are accurately given by a WKB approximation. The leading order is thus obtained from the WKB representation of ρ:

$$\rho_{\mathrm{WKB}} = \int \frac{d\lambda\, dp}{2\pi}\,\delta\Big(\tilde{\varepsilon} - \frac{1}{2}p^2 - \tilde{V}(\lambda)\Big). \tag{130}$$

From there one finds the planar approximation to the $c = 1$ string. The result is a function of the coupling constants which becomes singular when the horizontal line defined by the Fermi energy μ becomes tangent to the top of the potential [30]; namely at this critical point the Fermi energy μ takes the value $\tilde{\varepsilon}_F$ defined by

$$\tilde{V}(\lambda_0) = \tilde{\varepsilon}_F, \qquad \tilde{V}'(\lambda_0) = 0. \tag{131}$$

For instance for a potential without cubic terms, $E_0(g_4)$ is analytic in a circle with a singularity g_c on the real negative axis.

It is easy to verify that ρ_{WKB} is logarithmic in $\tilde{\varepsilon}$ for $\tilde{\varepsilon}$ near $\tilde{\varepsilon}_F$, the singularity coming form the motion near the turning point near the top of the potential. This logarithmic singularity gives, when we perform the integrations (128, 129) a logarithmic singularity in the energy itself. If one computes instead the secong derivative of Z_{st} with respect to the coupling constant one finds that it behaves as $1/\log(g - g_c)$, a logarithmic singularity on top of the simple KPZ scaling laws that we have met up to now. Higher orders in the $1/N^2$ expansion, are easy to handle in the double scaling limit [25]. Indeed the singular part comes from the motion in λ near the vicinity of the top of the potential, which is generically a simple parabola. We have therefore to deal with an inverted harmonic oscillator (cut-off by confining walls at large distance). We leave here the reader to the original articles, and to V. Kazakov's lectures in this volume, for the discussion of the very rich problem of the $c = 1$ string compactified on a circle.

References

1. Wigner, E. (1951) *Proc. Cambridge Philos. Soc.*, **47**, 790.
2. Dyson, F. J. (1970) *Commun. Math. Phys.*, **19**, 235.
3. Dyson, F. J. (1972) *J. Math. Phys.*, **13**, 90.
4. Mehta, M. L. (1991) *Random matrices*, 2nd ed., Academic, New York.
5. Delande, D. and Gay, J. C. (1986) *Phys. Rev. Lett.* , **57**, 2006.
6. Bohigas, O., Giannoni, M. J. and Schmidt, C. (1984) *Phys. Rev. Lett.*, **52**, 1.
7. Andreev, A. V., Agam, O., Simons, B. D. and Altshuler, B. L. (1996) *Nucl. Phys.* B, **482**, 536; `cond-mat/9601001` and `cond-mat/9605204`.
8. Beenakker, C. W. J. (1997) *Rev. Mod. Phys.*, **69**, 731; `cond-mat/9612179`.
9. Prahőfer and Spohn (2000) *Phys. Rev. Lett.*, **84**, 4882; `cond-mat/9912264`.
10. Odlyzko, A. M. (1987) *Math. Comput.*, **48** 273.
11. Kazakov, V. (1989) *Mod. Phys. Lett.* A, **4**, 2125.
12. Kazakov, V. (1986) *Phys. Lett.* A, **119**, 140.
13. David, F. (1992) *Proceedings Les Houches*, **LVII**, 679.
14. Knizhnik, V. C., Polyakov, A. M. and Zamolodchikov, A. B. (1988) *Mod. Phys. Lett.* A, **3**, 819.
15. David, F. (1988) *Mod. Phys. Lett.* A, **3**, 1651.
16. Distler, J. and Kawai, H. (1989) *Nucl. Phys.* B, **231**, 509.
17. Polyakov, A. (1981) *Phys. Lett.* B, **163**, 207.
18. Brézin, E., Itzykson, C., Parisi, G. and Zuber, J.-B. (1978) *Commun. Math. Phys.*, **69**, 35.
19. Brézin, E. and Kazakov, V. (1990) *Phys. Lett.* B, **236**, 144.
20. Douglas, M. and Shenker, S. (1990) *Nucl. Phys.* B, **335**, 635.
21. Gross, D. J. and Migdal, A. A. (1990) *Phys. Rev. Lett.*, **64**, 127.
22. Douglas, M. (1990) *Phys. Lett.* B, **238**, 213.
23. Gel'fand, L. M. and Dikii, L. A. (1975) *Russian Math. Surveys*, **30**, no. 5, 77.
24. Douglas, M. (1990) in O. Alvarez, E. Marinari and P. Windey (Eds.) *Proceedings of the Cargèse meeting "Random surfaces and quantum gravity"*.
25. Brézin, E., Kazakov V. and Zamolodchikov, A. B. (1990) *Nucl. Phys.* B, **338**, 673.
26. Brézin, E., Douglas, M., Kazakov, V. and Shenker, S. (1990) *Phys. Lett.* B, **237**, 43.

27. Harish-Chandra (1956) *Proc. Nat. Acad. Sci.*, **42**, 252.
28. Itzykson, C. and Zuber, J.-B. *J. Math. Phys.*
29. Mehta, M. L. (1981) *Commun. Math. Phys.*, **79**, 327.
30. Kazakov, V. and Migdal, A. A. (1985) *Phys. Lett.* B, **157**, 295.

A CLASS OF THE MULTI-INTERVAL EIGENVALUE DISTRIBUTIONS OF MATRIX MODELS AND RELATED STRUCTURES

V. BUSLAEV (buslaev@mph.phys.spbu.ru)
St.-Petersburg University
St.-Petersburg
Russia

L. PASTUR (pastur@math.jussieu.fr)
University of Wales Swansea
Swansea, United Kingdom
Permanent address:
University Paris 7
Paris
France

Abstract. For any integer $p \geq 1$ we present a class of polynomial potentials of matrix models for which the limiting density of eigenvalues can be found explicitly in elementary functions. The support of the density consists generically from p intervals. We introduce also certain p-periodic real symmetric Jacobi matrices and we give formulas relating the limiting eigenvalue density and the potential of the considered random matrix ensembles with the density of states and the Lyapunov exponent of these Jacobi matrices.

1. Introduction

The paper deals with certain problems motivated by and related to recent studies of the eigenvalue distribution of random matrix ensembles of $n \times n$ Hermitian matrices defined by the probability distribution

$$P(dM) = \frac{1}{Z_n} \exp\left\{-\frac{n}{g} \operatorname{Tr} V(M)\right\} dM, \tag{1}$$

where Z_n is the normalization constant, V/g (called the *potential*) is a real polynomial of degree $2p$ and positive at infinity, and

$$dM = \prod_{j=1}^{n} dM_{jj} \prod_{1\leq j<k\leq n} d\operatorname{Re} M_{jk}\ \operatorname{Im} M_{jk}. \tag{2}$$

V. A. Malyshev and A. M. Vershik (eds.),
Asymptotic Combinatorics with Application to Mathematical Physics, 51–70.

The ensembles of this form are often called the *matrix models*, because they are related to certain models of the Quantum Field Theory (see e.g. [9]). It is clear from (1) that without loss of generality we can write V as follows

$$V(\lambda) = \frac{\lambda^{2p}}{2p} + \sum_{l=1}^{2p-1} V_l \frac{\lambda^l}{2l}. \tag{3}$$

We will use below this normalization of the potential and the scale of the spectral parameter λ, making explicit the dependence of the potential on its overall amplitude $1/g$ that will be of considerable use below.

Denoting by $\lambda_l^{(n)}$, $l = 1, \ldots, n$ eigenvalues of an $n \times n$ Hermitian matrix M we can define its *normalized counting measure* N_n as

$$N_n(\Delta) = \sharp\{\lambda_l^{(n)} \in \Delta\}\, n^{-1} \tag{4}$$

for any Borel set Δ of $\mathbf{R}$.

It was pointed out by Wigner [27] (see also important physical papers [10, 5]) that N_n tends to the non-random limiting measure N as $n \to \infty$ and that N is a (unique) minimizer of the functional

$$\mathcal{E}[f] = -\int_{\mathbf{R}}\int_{\mathbf{R}} \log|\lambda - \mu|\widetilde{N}(d\lambda)\widetilde{N}(d\mu) + \int_{\mathbf{R}} V_{ext}(\lambda)\widetilde{N}(d\lambda), \tag{5}$$

where $V_{ext} = V/g$ and $\widetilde{N}$ is normalized as follows

$$\widetilde{N}(\mathbf{R}) = 1. \tag{6}$$

The variational problem (5)–(6) goes back to Gauss and is called the *minimum energy problem in the external field* V_{ext}. The unit measure N minimizing (5)–(6) is called the *equilibrium measure* in the external field V_{ext} because of its evident electrostatic interpretation as the equilibrium distribution of linear charges on the ideal conductor occupying the axis $\mathbf{R}$ and confined by the external electric field of potential V_{ext}. We stress that the respective variational procedure determine the both, the (compact) support σ of the measure and its form. This should be compared with the variational problem of the theory of logarithmic potential, where the external field is absent but the support σ is given. The respective measure is known as the equilibrium measure of the compact σ. We discuss in Section 3 the relation between these two measures.

The minimum energy problem in the external field (5)–(6) arises in various domains of analysis and its applications (see recent book [23] for a rather complete account of results and references concerning the problem). In particular, the problem is important in recent advances in asymptotic properties of certain classes of orthogonal polynomials [7, 8], and in random matrix theory [4, 11, 22]. In particular, it was shown in [4] that the normalizing counting measure (4) of the ensemble (1)–(2) converges in a certain metrics and in probability as $n \to \infty$ to a non-random measure N, coinciding with this equilibrium measure for any V that is locally C^γ for some $\gamma > 0$ and that grows faster than $2\log|\lambda|$ at infinity.

Most of studies in the random matrix theory treat the cases in which the support of the limiting eigenvalue counting measure is a single interval. Meanwhile, the cases, in which the support of the measure consists of several disjoint intervals are also of considerable interest and display a number of interesting effects that are absent in the single interval case.

In this paper we present a class of many-interval solutions of the minimum energy problem with a polynomial potential of a special form (see Theorem 1 below). The solutions are explicit in the sense that they can be written directly via potential in elementary functions, while generally, even in the case of polynomial potentials, the minimizers of (5)–(6), in particular the edges of their supports, are determined by a rather non-explicit and involved procedure including elliptic functions.

The paper is organized as follows. In Section 2 we present our construction. In Section 3 we discuss certain Jacobi matrices and their asymptotic forms whose spectral properties explain and, in fact, motivate the form of our solutions.

2. A Class of Multi-Interval Eigenvalue Distributions

We begin by formulating basic results of papers [4, 19, 22] on the equilibrium measure N, obtained in the random matrix context, where N coincides with the limiting eigenvalue distribution of the ensemble (1)–(2) (for other results and derivations see [7, 23]). We restrict ourselves in the sequel to the case where V is a polynomial of degree $2p$, $p \geq 1$.

It follows from these papers that in the case of a polynomial potential V, $\deg V = 2p$, the measure N is absolute continuous, its support σ is a compact set of real axis consisting of $q \leq p$ intervals, and its density ρ is real analytic everywhere except its zeroes, in particular, the edges of the support. The density can be found as a unique non-negative and normalized to unity minimizer of the functional (5) in which $\widetilde{N}(d\lambda)$ is replaced by $\widetilde{\rho}(\lambda)d\lambda$, the minimizing procedure determines the both the support σ and the functional form of the density ρ. In fact, the former determines the latter. Indeed, denoting by $-\infty < a_1 < b_1 < a_2 < \cdots < b_q < \infty$ the edges of the support

$$\sigma = \bigcup_{l=1}^{q} [a_l, b_l], \tag{7}$$

of N, we can write that for any interval Δ of the real axis

$$N(\Delta) = \int_{\Delta} \rho(\lambda) d\lambda \tag{8}$$

where

$$\rho(\lambda) = t(\lambda)\ R_+^{1/2}(\lambda), \tag{9}$$

$$t(\lambda) = \frac{1}{\pi g} \int_{\sigma} \frac{V'(\lambda) - V'(\mu)}{\lambda - \mu} \frac{d\mu}{R_+^{1/2}(\mu)}, \tag{10}$$

$t(\lambda)$ is a non-negative on σ polynomial of degree $2p-q-1$,

$$R_+^{1/2}(\lambda) = \operatorname{Im} \sqrt{R(z)}\Big|_{z=\lambda+i0}, \quad R(z) = \prod_{l=1}^{q}(z-a_l)(z-b_l), \tag{11}$$

and the branch of the square root is determined by its asymptotic behavior $\sqrt{R(z)} = z^q(1+O(z^{-1}))$, $z \to \infty$. The necessary conditions on the potential in the case of a q-interval support are

$$\int_\sigma \frac{V'(\lambda)\lambda^a d\lambda}{(R(\lambda))_+^{1/2}} = 2\pi g \delta_{aq}, \; a = 0, 1, \ldots, q. \tag{12}$$

Another useful although implicit expression for the density is

$$\rho(\lambda) = \frac{1}{2\pi g} \left(V'^2(\lambda) - 4Q(\lambda)\right)_+^{1/2}, \tag{13}$$

where

$$Q(\lambda) = Q(\lambda + i0), \; Q(z) = \int_\sigma \frac{V'(z) - V'(\mu)}{z - \mu} \rho(\mu)\, d\mu. \tag{14}$$

The following theorem provides a class of polynomial potentials of the degree $2p$, $p \geq 1$ for which there exists an explicit solution of the minimum energy problem (5)–(6). In other words, the random matrix ensemble (1)–(2) possess an explicit limiting eigenvalue distribution given by the theorem.

Theorem 1. *Let $v(z)$ be a polynomial of degree p with real coefficients, having the properties:*

(*i*) $v(z) = z^p + O(z^{p-1})$, $z \to \infty$;

(*ii*) *there exists $g > 0$ such that all zeroes $(a_1, b_1, \ldots a_p, b_p)$ of the polynomial $v^2(z) - 4g$ are real and simple.*

Assume that there exists a real constant C such that V in (5) *can be written as*

$$V(\lambda) = \frac{v^2(\lambda)}{2p} + C. \tag{15}$$

Then the support of the equilibrium measure corresponding to the potential $V_{ext} = V/g$ in (5) *is*

$$\sigma_g = \{\lambda \in \mathbf{R} : \; v^2(\lambda) \leq 4g\} = \bigcup_{l=1}^{p} [a_l, b_l], \tag{16}$$

and the function

$$\rho_g(\lambda) = \frac{|v'(\lambda)|}{2\pi p g} \left|v^2(\lambda) - 4g\right|^{1/2} \chi_\sigma(\lambda), \tag{17}$$

where $\chi_\sigma(\lambda)$ is the indicator of σ, is the density of the equilibrium measure N_g, i.e. a unique solution *of the minimum energy problem* (5)–(6) *with $V_{ext} = V/g$.*

Proof. According to [23] the function

$$W(\lambda) = U(\lambda) + \frac{1}{2g}V(\lambda), \tag{18}$$

where

$$U(\lambda) = -\int_{\mathbf{R}} \log|\lambda - \mu|\rho(\mu)\, d\mu, \tag{19}$$

satisfies the relations

$$W(\lambda) = F_V, \ \lambda \in \sigma, \tag{20}$$

$$W(\lambda) \geq F_V, \ \lambda \notin \sigma, \tag{21}$$

with a certain constant F_V if and only if ρ is the density of the equilibrium measure. We will use below this form of the variational problem (5)–(6). The constant F_V is called the ***modified Robin constant.*** Thus, it suffices to verify integral relations (20) and (21) for the density (17).

For a given real μ denote by $\log(z-\mu)$ the branch of logarithm analytic in the complex plane with the cut $(-\infty, \mu)$ and positive on the upper edge of the cut. Consider the function

$$w(z) = \frac{1}{2p}(\log g - 1) - \int_{\mathbf{R}} \log(z-\mu)\rho_g(\mu)\, d\mu, \tag{22}$$

where ρ_g is defined in (17). Let us show that in this case the function $w(z)$ has the form

$$w(z) = \frac{1}{p}\left[u(z)\sqrt{u^2(z)-1} - \log\left(u(z) + \sqrt{u^2(z)-1}\right)\right] - \frac{u^2(z)}{p}, \tag{23}$$

where

$$u(z) = \frac{v(z)}{2\sqrt{g}}. \tag{24}$$

Consider the complex plane with the cuts along the support $\sigma_g = \cup_{l=1}^{p}[a_l, b_l]$. By using our definition of the branch of $\sqrt{v^2(z) - 4g}$, we conclude that this function is pure imaginary on σ and has opposite signs on the upper and lower edges of the cuts. This and conditions (i) and (ii) of the theorem imply that the signs of v' and of $\operatorname{Im}\sqrt{v^2(\lambda + i0) - 4g}$ alternate on the intervals $[a_l, b_l]$, $l = 1, \ldots, p$, being both positive on $[a_p, b_p]$. This allows us to write the function ρ_g of (17) as

$$\rho_g(\lambda) = \frac{v'(\lambda)}{2\pi p g}\left(v^2(\lambda) - 4g\right)_+^{1/2}. \tag{25}$$

Further, we have from (22)

$$w'(z) = \int_{\sigma} \frac{\rho_g(\mu)\, d\mu}{\mu - z}, \quad \operatorname{Im} z \neq 0, \tag{26}$$

and according to the above

$$w'(z) = \frac{1}{4i\pi p g}\int_{\cup_{l=1}^{p} C_l} \frac{v'(\zeta)\sqrt{v^2(\zeta) - 4g}}{\zeta - z}\, d\zeta = \frac{1}{i\pi p}\int_{\cup_{l=1}^{p} C_l} \frac{u'(\zeta)\sqrt{u^2(\zeta) - 1}}{\zeta - z}\, d\zeta,$$

where C_l is the contour encircling the cut $[a_l, b_l]$, $l = 1, \dots, p$. The integral in the r.h.s. of this relation will not be changed if we subtract from the numerator the part of its large-ζ expansion containing all the non-negative powers of ζ. Indeed, this part is the polynomial $u'(z)u(z)$, thus analytic in the interior of all the contours C_l, $l = 1, \dots, p$. Since z lies in the exterior of all the contours, respective contour integral is zero by the Cauchy theorem. After this subtraction the numerator will be analytic at infinity, thus, using again the Cauchy theorem, we obtain that

$$w'(z) = \frac{2}{p}\left[u'(z)\sqrt{u^2(z)-1} - u'(z)u(z)\right], \quad \mathrm{Im} z \neq 0. \tag{27}$$

On the other hand, differentiating the r.h.s. of (23), we find the same expression as in the r.h.s. of (27). Thus, the r.h.s. of (23) and of (22) differs at most by a constant. Since, however, the both r.h.s.'s have the same asymptotic form

$$-\log z + \frac{1}{2p}(\log g - 1) + O(1/z), \; z \to \infty,$$

we conclude that they coincide identically. Thus, we have established (23).

Since the function ρ_g of (17) is Hölder continuous (see e.g. (13)), we have from (19), and from (22)

$$\mathrm{Re}\, w(\lambda + i0) = \frac{1}{2p}(\log g - 1) - \int_\sigma \log|\lambda - \mu|\rho_g(\mu)\, d\mu = \frac{1}{2p}(\log g - 1) + U(\lambda), \tag{28}$$

and since V is given by (15), i.e. $V/2g = u^2/p + C/2g$, we can rewrite the form (20)–(21) of the minimum energy problem (5)–(6) as follows

$$\mathrm{Re}\,\Phi(\lambda + i0) = F_V, \; \lambda \in \sigma, \tag{29}$$

$$\mathrm{Re}\,\Phi(\lambda + i0) \geq F_V, \; \lambda \notin \sigma, \tag{30}$$

where

$$\begin{aligned}\Phi(z) &= w(z) + \frac{V}{2g} \\ &= \frac{1}{p}\left[u(z)\sqrt{u^2(z)-1} - \log\left(u(z) + \sqrt{u^2(z)-1}\right)\right] + \frac{C}{2g}\end{aligned} \tag{31}$$

Now, taking into account our definition of $\sqrt{u^2(z)-1}$ and analyzing the values of the r.h.s. of (31) for $z = \lambda + i0$, one can check directly validity of relations (29)–(30). We will give below another proof of these relations that will also be useful later. The proof is based on the theorem of [15], according to which a real polynomial u of degree p such that all zeroes of $u^2 - 1$ are real can always be written in the form

$$u(z) = \cos\theta(z), \tag{32}$$

where $\theta(z)$ is the conformal mapping of the upper half-plane $\mathbf{C}_+ = \{z \in \mathbf{C} : \mathrm{Im}\, z > 0\}$ onto the domain (the comb with p teeth)

$$\{\theta : p_1\pi < \mathrm{Re}\,\theta < p_2\pi, \; \mathrm{Im}\,\theta > 0\} \setminus \bigcup_{p_1 < l < p_2,} \{\theta : \mathrm{Re}\,\theta = l\pi, \; p_1 < l < p_2, \; 0 < \mathrm{Im}\,\theta \leq h_l\}, \tag{33}$$

where $p_1 < p_2$ are integers, $p_2 - p_1 = p$, $0 \le h_l < \infty$ and $\theta(\infty) = \infty$. In fact, $u(z) = \cos\theta(z)$ is a polynomial of degree p if and only if $\theta(z)$ admits values congruent to $\pi/2$ modulo π exactly p times, which is equivalent to the condition that $-\infty < p_1 < p_2 < \infty$, $p_2 - p_1 = p$. Function $\theta(z)$ is analytic in the open upper half-plane $\mathbf{C}_+$ and continuous in the closed upper half-plane $\overline{\mathbf{C}_+}$. When $z = \lambda + i0$ varies from $-\infty$ to ∞, the limiting value $\theta(\lambda + i0)$ runs along the boundary (the comb) of the domain (33) of $\mathbf{C}_+$, so that either $\operatorname{Im}\theta(\lambda + i0) = 0$, if λ belongs to the l-th "band" $[a_l, b_l]$, $l = 1, \dots, p$ of σ_g, or $\operatorname{Re}\theta(\lambda + i0) \equiv 0(\operatorname{mod}\pi)$ and $\operatorname{Im}\theta(\lambda + i0) = \kappa$, $0 \le \kappa \le h_l$, if λ belongs to the l-th "gap" (b_l, a_{l+1}) of σ_g, where we set $a_{p+1} = a_1$ viewing the gap (b_p, a_1) as passing the infinity. By using the terminology of solid state theory we can say that $\theta(z)$ is an analytic continuation of the quasimomentum as a function of energy.

We set $p_1 = 0$, $p_2 = p$, so that $\theta(\lambda + i0)$ varies from $(l-1)\pi$ to $l\pi$, $l = 1, \dots, p$ when λ varies from a_l to b_l in the l-th band. By using (32) and (33) we can rewrite (29)–(30) in the form

$$\operatorname{Re}\Phi(\lambda + i0) = \frac{C}{2g} + \begin{cases} 0, & \text{if } \lambda \in [a_l, b_l], \quad l = 1, \dots, p, \\ g(\kappa_l(\lambda)), & \text{if } \lambda \in [b_l, a_{l+1}], \quad l = 1, \dots, p, \end{cases} \tag{34}$$

where

$$g(\kappa) = \sinh 2\kappa - 2\kappa = 4\int_0^{2\kappa} \cosh^2 t\, dt > 0, \ \kappa > 0.$$

Thus, we obtained (29), thereby we proved that (17) is the density of the equilibrium measure corresponding to the potential $V_{ext} = V/g$. Relation (16) follows from (17). □

Remarks. 1. It is easy to see that an additive constant in the potential (15) does not change neither the random matrix ensemble (1) nor variational problem (5)–(6).

2. We obtain form the above the following expression for the modified Robin constant F_V

$$F_V = \frac{C}{2g}. \tag{35}$$

Also, comparing (25) and (13) we conclude that $Q = v'^2/p^2 g$. The same expression for Q can be obtained from (14) and (17) by computing respective integrals over σ by the contour integration as in the proof of Theorem 1. Analogously, the polynomial t from (10) is $v'/2\pi p g$.

3. According to representation (32)–(33) all zeros of v are simple and real. Indeed, zeros of v are those values of z for which $\theta(z) \equiv \pi/2(\operatorname{mod}\pi)$. But all these points are real and such that $\theta'(z) \ne 0$. Thus, condition (ii) of the theorem can be written the form

$$2\sqrt{g} < v_*, \tag{36}$$

where v_* is the minimum modulus of finite extremums of v. Thus the support of the equilibrium measure consists of maximum number p of disjoint intervals possible for a polynomial potential (3). This situation is obviously generic. On the other hand, the case where the polynomial $v^2 - 4g$ has multiple zeros can be viewed as the limiting one for the above case corresponding to the situation where some of the gaps (b_l, a_{l+1}) shrink to a point and respective adjacent bands touch. Thus, we can write more general condition

$$2\sqrt{g} \le v_*, \tag{37}$$

and under this condition the support of the equilibrium measure will consists of $q < p$ intervals, where $2q$ is the number of simple zeros of the polynomial $v^2 - 4g$. The form of the respective density can be easily found from (17) (see e.g. formula (76) corresponding to the quartic potential $p = 2$). This has to be compared with spectral theory of periodic and almost periodic Jacobi matrices (see e.g. [25, 26]), where the similar situation corresponds to the so-called finite-band potentials. For links with spectral theory of periodic Jacobi matrices see also the next section and [14].

Corollary 2. *Let N_g be the limiting normalized counting measure of the ensemble* (1) *with the potential given by* (15), *i.e. the equilibrium measure corresponding to $V_{ext} = V/g$ whose density is given by* (17). *Then for any interval $\Delta = (a, b)$ $-\infty < a < b < \infty$ of the spectral axis we have*

$$N_g(\Delta) = \frac{1}{\pi p}\left(\theta_+(\lambda) - \frac{\sin 2\theta_+(\lambda)}{2}\right)\Bigg|_{\lambda=a}^{\lambda=b}, \tag{38}$$

where $\theta_+(\lambda) = \mathrm{Re}\ \theta(\lambda + i0)$, and $\theta(z)$ is defined in (32). *In particular, if $[a_l, b_l]$ is a "band" of the p-interval spectrum* (7), *then*

$$N_g([a_l, b_l]) = \frac{1}{\pi p}, \quad l = 1, \ldots, p. \tag{39}$$

Proof. By using formulas (32) and (25) we obtain that

$$\rho_g(\lambda) = \frac{2}{\pi p}\sin^2\theta_+(\lambda)\theta_+'(\lambda). \tag{40}$$

This relation implies (38). Then (38) and (33) lead to (39). □

Remark. It is easy to see from (17) that N_g converges weakly as $g \to 0$ to the atomic measure concentrated at zeros of v and having all atoms equal $1/p$. In view of the form (15) of the potential this fact can be regarded as a detalization of Theorem 1.3 (ii) of recent paper [12], according to which in a general case of a real analytic potential N_g converges as $g \to 0$ to the atomic measure concentrated at the absolute minima of V. On the other hand, by using formulas (9)–(14), it can be shown that in the opposite regime $g \to \infty$ of the potential (15) after rescaling $\lambda g^{-1/2p}$ the measure N_g converges weakly to the measure corresponding to the monomial potential $\lambda^{2p}/2p$ (see (73)) with the density given by (75).

3. Periodic Jacobi Matrices

In this section we restrict ourselves to the case of even potentials, more precisely we will assume that the polynomial v in (15) has the property

$$v(-\lambda) = (-1)^p v(\lambda). \tag{41}$$

Denote by $p_l^{(n)}(\lambda)$, $l = 0, 1, \ldots$ polynomials orthogonal on $\mathbf{R}$ with respect to the weight

$$w_n(\lambda) = e^{-nV(\lambda)/g}. \tag{42}$$

$p_l^{(n)}(\lambda)$ is a polynomial of the degree l. We use the normalization

$$\int_{\mathbf{R}} p_l^{(n)}(\lambda) p_m^{(n)}(\lambda) e^{-nV(\lambda)/g}\, d\lambda = \delta_{lm}, \tag{43}$$

assuming also that the coefficient in front of λ^l in $p_l^{(n)}(\lambda)$ is positive. We also denote

$$\psi_l^{(n)}(\lambda) = e^{-nV(\lambda)/2g} p_l^{(n)}(\lambda), \quad l = 0, \ldots \tag{44}$$

the corresponding system of the orthonormalized functions in $L^2(\mathbf{R})$.

The system $\{p_l^{(n)}(\lambda)\}_{l=0}^{\infty}$ corresponding to an even weight (42) generates the real symmetric semi-infinite symmetric Jacobi matrix $J^{(n)}$ via standard recursion relations

$$\begin{aligned} r_0^{(n)} p_1^{(n)}(\lambda) &= \lambda p_0^{(n)}(\lambda) \\ r_{l-1}^{(n)} p_{l-1}^{(n)}(\lambda) + r_l^{(n)} p_{l+1}^{(n)}(\lambda) &= \lambda p_l^{(n)}(\lambda), \quad l \geq 1, \end{aligned} \tag{45}$$

i.e.

$$J^{(n)} = \begin{pmatrix} 0 & r_0^{(n)} & 0 & \cdots \\ r_0^{(n)} & 0 & r_1^{(n)} & \ddots \\ 0 & r_1^{(n)} & 0 & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{pmatrix}. \tag{46}$$

Our considerations in this section will be based on the following limiting property of the coefficients $r_{n-1}^{(n)}$.

Theorem 3. *For any potential of the form* (15), (41) *there exists a p-periodic sequence* $\{r_m,\ r_{m+p} = r_m\}_{m \in \mathbf{Z}}$, *such that if* $n = n_1 p + m$, $1 \leq m \leq p$, *then*

$$\lim_{n \to \infty} r_{n-1}^{(n)} = r_m. \tag{47}$$

Proof. The assertion is a simple corollary of the recent strong results by Deift et al. [8] on the asymptotics of orthogonal polynomials (43), corresponding to a real analytic (not necessarily of the form (3)) potential V in the weight (42). Set

$$N_g(\lambda) = \int_\lambda^\infty \rho_g(\lambda') d\lambda', \tag{48}$$

and

$$\alpha_l = N_g(a_l),\ \alpha = \{\alpha_l\}_{l=2}^p \in \mathbf{R}^{p-1}, \tag{49}$$

i.e. α_l is the charge of the union $[a_l, b_l] \cup \cdots \cup [a_p, b_p]$ of the bands of σ_g with respect to the distribution $N_g(d\lambda) = \rho_g(\lambda)d\lambda$. Then, according to [8], there exists a function $F(x, \lambda, g)$, $x \in \mathbf{T}^{p-1}, \lambda \in \mathbf{R},\ g \in \mathbf{R}_+$, continuous in all its variables, 1-periodic in each variable x_l, $l = 1, \ldots, p-1$, and such that

$$r_{n-1}^{(n)}(g) = F(n\alpha, \lambda, g) + O(1/n^\kappa), \quad n \to \infty \tag{50}$$

for some $\kappa > 0$. The function F can be expressed via the Riemann theta-function constructed in a canonical way from the Riemann surface of genus $p-1$ associated with $\sqrt{v^2(z) - 4g}$ and obtained by gluing together the two copies of the complex plane along $\mathbf{R} \setminus \sigma$.

According to (39) we have $\alpha_l = (p+1-l)/p, l = 2, \ldots, p$. This relation and (50) imply (47). □

Remark. We will need below the limiting form of the coefficients $r_{l-1}^{(n)}$, where l is not necessarily n. This form can be deduced from (47) and from the dependence of the coefficients on the amplitude g. We make explicit this dependence writing $r_{l-1}^{(n)}(g)$ and $r_{l-1}(g)$ in (47). We have obviously

$$r_{l-1}^{(n)}(g) = r_{l-1}^{(l)}\left(\frac{l}{n}g\right).$$

Thus, if

$$n \to \infty,\ l \to \infty,\ l/n \to \xi,\ l = l_1 p + m,\ 0 \le m < p, \tag{51}$$

then we obtain from (47)

$$\lim r_l^{(n)}(g) = r_m(\xi g). \tag{52}$$

where the symbol lim... denotes here and below the limiting procedure (51).

Denote now by $J(g)$ the infinite in both directions real symmetric Jacobi matrix, generated by the p-periodic sequence $\{r_m(g)\}_{m \in \mathbf{Z}}$. Recall that for any real symmetric Jacobi matrix J with periodic coefficients r_m, $m \in \mathbf{Z}$, one can define the integrated density of states. Namely, denote by J_{m_1,m_2} the restriction of J to the finite interval $[m_1, m_2+1]$, $m_1 \le m_2$, i.e. the $(m_2 - m_1 + 2) \times (m_2 - m_1 + 2)$ matrix

$$J_{m_1,m_2} = \begin{pmatrix} 0 & r_{m_1} & 0 & \cdots & 0 \\ r_{m_1} & 0 & r_{m_1+1} & \ddots & \vdots \\ 0 & r_{m_1+1} & 0 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & r_{m_2} \\ 0 & \cdots & 0 & r_{m_2} & 0 \end{pmatrix}, \tag{53}$$

and let ν_{m_1,m_2} be the normalized counting measure of this matrix, i.e. if $\mu_l^{(m_1,m_2)}$, $l = 1, \ldots, m_2 - m_1 + 2$ are eigenvalues of J_{m_1,m_2} counted according to their multiplicities, then for any interval $\Delta \in \mathbf{R}$

$$\nu_{m_1,m_2}(\Delta) = \sharp\{\mu_l^{(m_1,m_2)} \in \Delta\}(m_2 - m_1 + 2)^{-1}. \tag{54}$$

It can be shown that measures ν_{m_1,m_2} converge weakly as $m_2 - m_1 \to \infty$ to a non-negative and normalized to unity measure ν, known as the *Integrated Density of States* (IDS) of J:

$$\lim_{m_2-m_1\to\infty} \nu_{m_1,m_2} = \nu \tag{55}$$

(see e.g. [21], where this fact is proved in a general case of matrix operators with ergodic coefficients).

In the case of the matrix $J(g)$ defined by the periodic sequence $\{r_m(g)\}$ of (47) the integrated density of states depends on g and will be denoted ν_g. The following assertion relates ν_g to the limiting eigenvalue counting measure N_g of the random matrix ensemble (1)–(2) i.e., in view of results of [4], to the equilibrium measure of the energy functional (5).

Theorem 4. *Let N_g be the limiting eigenvalue counting measure of the matrix ensemble* (1)–(2) *and ν_g is the integrated density of states of the Jacobi matrix $J(g)$ whose coefficients are defined in* (51). *Then*

$$N_g = g^{-1} \int_0^g \nu_\xi \, d\xi. \tag{56}$$

Proof. (a scheme) We will use the Stieltjes transform of all involved measures and simple properties of the resolvent of the self-adjoint operator $J^{(n)} = \{J_{kl}^{(n)}\}_{k,l=0}^{\infty}$ defined in $l_2(\mathbf{Z}_+)$ by the matrix (46):

$$J_{kl}^{(n)} = r_k^{(n)} \delta_{k+1,l} + r_{k-1}^{(n)} \delta_{k-1,l}, \quad k,l = 0,\ldots, \quad r_{-1}^{(n)} = 0. \tag{57}$$

Recall that the Stieltjes transform $f_\nu(z)$ of a non-negative unit measure ν is defined as

$$f_\nu(z) = \int_{\mathbf{R}} \frac{\nu(d\lambda)}{\lambda - z}, \quad \operatorname{Im} z \neq 0\ . \tag{58}$$

f_ν is an analytic function for $\operatorname{Im} z \neq 0$,

$$\operatorname{Im} f_\nu(z) \cdot \operatorname{Im} x > 0\ , \quad \operatorname{Im} z \neq 0\ , \tag{59}$$

$$\overline{f(z)} = f(\overline{z}), \quad \sup_{\eta \geq 1} \eta |f_\nu(i\eta)| = 1, \tag{60}$$

and any function possessing these properties can be represented in the form (58) [1]. $f_\nu(z)$ determines uniquely the measure: if $\Delta = (a,\, b)$ and a and b are continuity points of ν, then

$$\nu(\Delta) = \lim_{\varepsilon\to+0} \frac{1}{\pi} \int_a^b \operatorname{Im} f_\nu(\lambda + i0)\, d\lambda. \tag{61}$$

Besides, if $\{\nu_n\}_{n=1}^{\infty}$ is a sequence of measures that converges weakly to a measure ν, then

$$\lim_{n\to\infty} f_{\nu_n}(z) = f_\nu(z) \tag{62}$$

uniformly on compact sets of $\mathbf{C}\backslash\mathbf{R}$. The converse statement is also true.

Further, if $\mathcal{E}^{(n)} = \{\mathcal{E}_{kl}^{(n)}\}_{k,l=0}^{\infty}$ is the resolution of the identity of the self-adjoint operator $J^{(n)}$ defined by (46), i.e. by the semi-infinite Jacobi matrix (46), then [1]

$$\mathcal{E}_{kl}^{(n)}(d\lambda) = \psi_k^{(n)}(\lambda)\psi_l^{(n)}(\lambda)\,d\lambda \equiv e_{kl}^{(n)}(\lambda)\,d\lambda, \quad k,\, l = 0, 1, \ldots, \tag{63}$$

where $\{\psi_l^{(n)}(\lambda)\}_{l=0}^{\infty}$ are defined in (42)–(44). On the other hand, according to [16] the expectation $\overline{N}_n = E\{N_n\}$ of the normalized counting measure (4) with respect to the probability distribution (1)–(2) is the absolutely continuous measure whose density is

$$\rho_n(\lambda) = n^{-1}\sum_{l=0}^{n-1}[\psi_l^{(n)}(\lambda)]^2. \tag{64}$$

In view of (63) we can rewrite (64) as

$$\rho_n(\lambda) = n^{-1}\sum_{l=0}^{n-1} e_{ll}^{(n)}(\lambda)\ . \tag{65}$$

This relation and the spectral theorem for self-adjoint operators imply that the Stieltjes transform $f^{(n)}(z)$ of the measure $\rho_n(\lambda)\,d\lambda$ is

$$f^{(n)}(z) = n^{-1}\sum_{l=0}^{n-1} G_{ll}^{(n)}(z)\ , \tag{66}$$

where $G^{(n)}(z) = (J^{(n)} - z)^{-1}$ is the resolvent of $J^{(n)}$. By using (57) and the resolvent identity we can prove that $f^{(n)}(z)$ differs by $O(r_{n-1}^{(n)}(n(|\operatorname{Im} z|)^{-1})$ from the Stieltjes transform $f_{\nu_n}(z)$ of the normalized eigenvalue counting measure ν_n of the Jacobi matrix $J_n^{(n)} = J_{0,n-1}$ defined by the coefficients $r_0^{(n)}, \ldots, r_{n-1}^{(n)}$ (cf. (53) and by (54)). According to [4] the coefficients $r_l^{(n)}, l = 0, 1, \ldots$ are uniformly bounded in l and in n

$$\sup_{l,n} r_l^{(n)} \le C < \infty.$$

Thus

$$\lim_{n\to\infty} |f_{\nu_n}(z) - f^{(n)}(z)| = 0, \ \operatorname{Im} z \ne 0.$$

Set now $n = n_1 n_2$, divide the "interval" $[0, n]$ on n_1 "subintervals" of the length n_2 each, and show by similar arguments that $f_{\nu_n}(z)$ differs by $O((n|\operatorname{Im} z|^{-1}))$ from the arithmetic mean of n_1 Stieltjes transforms of the eigenvalue counting measures of the $n_2 \times n_2$ Jacobi matrices defined by $r_{sn_2+t}^{(n)}$, $t = 1, \ldots, n_2 - 1$ for any $s = 0, 1, \ldots, n_1 - 1$. Furthermore we choose n_1 to be so large as to guarantee the replacement, in view of (52), of corresponding coefficients $r^{(n)}$'s within the s-th block by the periodic coefficients $r_t(sg/n_1)$ defined in (52). If n_2 is also large enough then in view of (47) the contribution of each block is close enough to the Stieltjes transform $f(\{r_t(sg/n_1)\}_{t\in\mathbf{Z}}, z)$ of the infinite periodic Jacobi matrix defined by the coefficients $\{r_t(sg/n_1)\}_{t\in\mathbf{Z}}$. This means that uniformly in z, $|\operatorname{Im} z| \geq \eta_0 > 0$ we have

$$\lim_{n\to\infty} f^{(n)}(z) = g^{-1}\int_0^g f(\{r_m(\xi g)\}, z)\, d\xi,$$

where $f(\{r_m(\xi g)\}, z)$ denotes the Stieltjes transform of the integrated density of states of the periodic Jacobi matrix defined by the coefficients $\{r_m(\xi g)\}_{m\in\mathbf{Z}}$. This formula and the continuity of the one to one correspondence between unit nonnegative measures and their Stieltjes transforms leads to (56) □

According to (8)–(10) and to [21] both measures N_g and ν_g have densities, which we denote ρ_g and d_g respectively. Then we obtain from (56)

Corollary 5. *The densities ρ_g and d_g of the measures N_g and ν_g are related as follows:*

$$\rho_g(\lambda) = g^{-1}\int_0^g d_\xi(\lambda) d\xi. \tag{67}$$

Remarks. 1. Formulas (56) and (67) illustrate the "slow varying" (adiabatic) in g character of the coefficients $r_l^{(m)}$ of the second order finite-difference operator (57). Concerning similar formulas for a variety of finite-difference and differential operators see [21].

2. The analogue of (56) for the one-interval case but not necessary for potentials of form (15) was given in [19] (see also [13] for its another proof).

Examples. We consider now simplest cases of potentials discussed above.

1). The case $p = 1$ of a potentials of the form (3) is simple and general, because any such polynomial of the second degree can be written in the form (15) with $v(\lambda) = \lambda + V_1$:

$$V(\lambda) = \frac{1}{2}(\lambda + V_1)^2 - \frac{V_1^2}{2}.$$

In this case ensemble (1) becomes the well known Gaussian Unitary Ensemble [16] after the shift: $\lambda + V_1 \to \lambda$. In particular, respective orthogonal polynomials are

$$p_l^{(n)}(\lambda) = (n/2g)^{1/4}(2^l l!\sqrt{\pi})^{1/2} H_l(\lambda\sqrt{n/2g}),$$

where H_l are the Hermite polynomials, and $r_{l-1}^{(n)} = \sqrt{gl/n}$ (see e.g. [24]). Thus, in this case the limiting procedure (51) leads to the Jacobi matrix with constant (1-periodic) coefficients $r(g) = \sqrt{g}$, the equilibrium measure is the Wigner semicircle law [16], whose support is the interval $|\lambda| \leq 2\sqrt{g}$, and whose density is

$$\rho_W(\lambda) = (2\pi g)^{-1}\sqrt{4g - \lambda^2}\chi_{[-2\sqrt{g}, 2\sqrt{g}]}(\lambda). \tag{68}$$

We see that the class of equilibrium measures given in Theorem 1 shares with the Wigner law the property of the simplest dependence on the amplitude of the external potential and differs from the Wigner law by a more complex dependence on the spectral parameter via the polynomial v of degree $p \geq 1$ specified by conditions (i)–(ii) of Theorem 1, in particular by a p-interval support.

2). Consider now an arbitrary polynomial potential leading to a single-interval density of states. In particular, this is always the case if the potential is convex [4, 11]. According to [2, 8, 11], if the support of the equilibrium measure (the limiting eigenvalue measure of the random matrix ensemble (1)) consists of a single interval, then the limiting procedure (51) leads to constant in m coefficients. The density of states of the Jacobi matrix with constant coefficient r is $[\pi\sqrt{4r^2-\lambda^2}]^{-1}\chi_{[-2r,2r]}(\lambda)$. Thus we have from (67) and from (83) that in a general single-interval case [19]

$$\rho_g(\lambda) = \frac{1}{\pi g}\int_0^g \frac{d\xi}{(\lambda^2 - 4r^2(\xi))_+^{1/2}}. \tag{69}$$

The coefficient $r(g)$ can be found from the limiting form of the relation

$$r_{l-1}^{(n)} V'(J^{(n)}(g))_{l,l-1} = \frac{l}{n}g, \quad l = 1, \ldots \tag{70}$$

The relation can be easily obtained from the relations (42)–(43) by integration by parts. It has been known in the theory of orthogonal polynomials [17] and was rediscovered and widely used in the Quantum Field Theory under the name (pre)string equation (see e.g.([9]). Passing to the limit $n \to \infty$ in (70) we obtain the equation for $r(g)$

$$r(g)V'(J(g))_{01} = g, \tag{71}$$

where $J(g)$ is now the Jacobi matrix with the constant coefficients $r(g)$. By using spectral representation of $J(g)$ we can rewrite the last equation as

$$\int_{-2r(g)}^{2r(g)} \frac{V'(\lambda)\lambda\, d\lambda}{\sqrt{4r^2(g) - \lambda^2}} = 2\pi g. \tag{72}$$

A simple sufficient condition to have a single-interval equilibrium measure is convexity of the external potential [4, 11]. In this case the l.h.s. of this relation is monotone in r and the equation is uniquely solvable. It is easy to see that the equation coincides with the one-interval case (q=1) of (12). It can also be shown that (69) is the one-interval case (q=1) of formulas (9)–(11). Consider, for example the simple case of the monomial potential

$$V(\lambda) = \frac{\lambda^{2p}}{2pg}. \tag{73}$$

It follows from (72) that

$$r(g) = g^{1/2p}r(1),\ r(1) = \frac{1}{2}(\pi/C_p)^{1/2p}, \quad C_p = \int_0^1 \frac{t^{2p}dt}{\sqrt{1-t^2}}, \tag{74}$$

These relations and formula (69) lead to the following expression for the density of the limiting eigenvalue distribution corresponding to the potential (73)

$$\rho(\lambda) = \frac{1}{2r(g)}h\left(\frac{\lambda}{2r(g)}\right), \quad h(x) = \frac{2p}{\pi}\int_{|x|}^1 \frac{t^{2p-1}dt}{\sqrt{t^2-x^2}}. \tag{75}$$

These formulas were obtained in [18] by the method of orthogonal polynomials for $V = |\lambda|^{\alpha}/\alpha$, where $\alpha \geq 3$ but is not necessarily an even integer. They also follow from [4] for any $\alpha > 1$.

3). The case $p = 2$ in (15) is also general for an even potentials because any even polynomial (3) of the degree 4 can be written in form (15) with $v(\lambda) = \lambda^2 + V_2$, and if $V_2 < -2\sqrt{g}$. the polynomial $v^2 - 4g$ has 4 simple zeros $\pm a, \pm b$, where $a = \sqrt{-V_2 - 2\sqrt{g}}$, and $b = \sqrt{-V_2 + 2\sqrt{g}}$. According to the theorem we have in this case

$$\rho_g(\lambda) = \frac{8|\lambda|}{\pi(b^2 - a^2)^2}\sqrt{(b^2 - \lambda^2)(\lambda^2 - a^2)}\chi_\sigma(\lambda),$$

where $\sigma = [-b, -a] \cup [a, b]$. and $0 < a < b < \infty$. This is the generic case. If $V_2 = -2\sqrt{g}$, then $a = 0$ and the two intervals of spectrum merge into the interval $[-b, b]$, $b^2 = 4\sqrt{g}$ and the respective density is

$$\rho_g(\lambda) = \frac{8\lambda^2}{\pi b^4}\sqrt{(b^2 - \lambda^2)}\chi_{[-b,b]}(\lambda) \tag{76}$$

Here we can again use the result of [2], according to which if in the case of an even potential (not necessarily of the form (15)) the support of the equilibrium measure consists of two intervals, then the limiting procedure (1) leads to 2-periodic sequence $\{r_m(g)\}_{m\in\mathbf{Z}}$. The similar but more precise result for an even quartic polynomial was also proved in [3] as a part of the derivation of asymptotic formulas for the orthogonal polynomials corresponding to the weight (42) with this potential. However, there exist (non-even) polynomials of degree 4 that lead to non-periodic Jacobi matrices. Namely, by using [8, 12] it can be shown that if coefficient V_3 in (3) is not zero, then generically the number α_2 from (49) is irrational and the coefficient $r_{n-1}^{(n)}$ is quasiperiodic in n, i.e. there exists a continuous 1-periodic function $F(x)$, $x \in \mathbf{R}$ such that $r_{n-1}^{(n)} = F(\alpha_2 n) + o(1)$, $n \to \infty$. This leads to the quasiperiodic Jacobi matrix and to certain new effects [14].

Recall now the notion of the Lyapunov function that plays an important role in the spectral theory of periodic Jacobi matrices (see e.g. [25, 26]). For any p-periodic matrix J defined by coefficients $\{r_m,\ r_m = r_{m+p}\}_{m\in\mathbf{Z}}$ we can consider the second order finite-difference equation

$$r_{m-1}y_{m-1} + r_m y_{m+1} = zy_m,\ z \in \mathbf{C}, \tag{77}$$

and introduce its two linearly independent solutions $s_m(z)$ and $c_m(z)$ uniquely determined by the initial conditions

$$s_0(z) = c_1(z) = 0,\ s_1(z) = c_0(z) = 0.$$

The polynomial in z of degree p

$$\Delta(z) = s_{p+1}(z) + c_p(z) \tag{78}$$

is called the *Lyapunov function* (or the *Hill discriminant*) of the p-periodic matrix J. The spectrum $\sigma(J)$ of J is

$$\sigma(J) = \{\lambda \in \mathbf{R} : \Delta^2(\lambda) \leq 4\}, \tag{79}$$

and the density of states $d(\lambda)$ of J is

$$d(\lambda) = -\frac{\Delta'(\lambda)}{\pi p\,(\Delta^2(\lambda) - 4)_+^{1/2}}, \tag{80}$$

where $(\dots)_+^{1/2}$ is defined in (11).

We will prove now

Theorem 6. *The Lyapunov function $\Delta(g,\lambda)$ of the periodic Jacobi matrix $J(g)$, defined by the limiting p-periodic coefficients* (47) *and the polynomial $v(\lambda)$ from the representation* (15) *are related as follows*

$$\Delta(\lambda, g) = \sqrt{g}v(\lambda). \tag{81}$$

Proof. According to (67) we have

$$\frac{\partial}{\partial g}\left(g\rho_g(\lambda)\right) = d_g(\lambda). \tag{82}$$

This relation, and formulas (25), and (80) yield the equality

$$\frac{v'(\lambda)}{(v^2(\lambda) - 4g)_+^{1/2}} = \frac{\Delta'(\lambda)}{(\Delta^2(\lambda) - 4)_+^{1/2}}$$

valid for all $\lambda \in \sigma_g$. Since $v(\lambda)$ and $\Delta(\lambda)$ are polynomials of degree p in λ, we obtain (81). □

Remarks. 1. Relations (80), and (81) imply that the density of states d_g of the p-periodic Jacobi matrix $J(g)$ is

$$d_g(\lambda) = -\frac{v'(\lambda)}{\pi p}\left(v^2(\lambda) - 4g\right)_+^{-1/2} \tag{83}$$

By using (24) and (32), we can also rewrite (83) as (cf. (40))

$$d_g(\lambda) = \frac{1}{\pi p}\theta_+'(\lambda), \quad \theta_+(\lambda) = \theta(\lambda + i0), \tag{84}$$

thus the Integrated Density of States ν_g of the matrix $J(g)$ is (cf (38))

$$\nu_g(\Delta) = \frac{1}{\pi p}\theta_+(\lambda)\Big|_{\lambda=a}^{\lambda=b}, \quad \Delta = (a,b). \tag{85}$$

In particular, if $[a_l, b_l]$, $l = 1, \dots, p$ is a band of the spectrum of $J(g)$, then

$$\nu_g([a_l, b_l]) = \frac{1}{\pi p}, \quad l = 1, \dots, p. \tag{86}$$

These formulas illustrate the fact that $\theta_+(\lambda)$ is the quasimomentum as the function energy in the extended band scheme of spectral theory and solid state theory. In particular, the inverse of $\theta_+(\lambda)$ on the l-th band is closely related to the band function $\lambda_l(k)$ (see formula (89) below).

2. Formula (83) allows us to give another interpretation of the formulas (56) and (67). Namely, consider the functional of electrostatic energy (5) with $V=0$ but with a given support σ of the unit measure. This yields the standard variational problem of the theory of logarithmic potential. Its solution is known as the equilibrium (or Robin) distribution of the compact σ_g [23]. By repeating almost literally the proof of Theorem 1 we find that the density d_g given in (83) solves this problem if the support is given by (16). Comparing (25) and (83) we conclude that the equilibrium measure corresponding to the external field $V_{ext}=V/g$ in (5), where V is given by (15) and the equilibrium distribution of the compact σ_g given by (16) are related by formulas (56) and (82). In this potential theoretic context the formula was derived and analyzed in paper [6] under very general conditions (see also [13] for related results).

By using this general version, it can be proved that the equality

$$N_g([a_l,b_l])=\nu_g([a_l,b_l]),\quad l=1,\dots,p, \tag{87}$$

resulting from (39), and from (86), is valid in general case of an arbitrary polynomial potential, leading generically to a quasiperiodic Jacobi matrix $J(g)$ (see [14] for its definition and certain properties).

We will use now formula (81) to explain the form (25) of the density of the limiting counting measure of random matrix ensemble (1)–(2) with potential (15) (the solution of the minimum energy problem with this external field).

To any p-periodic Jacobi matrix J we can associate the $p\times p$ "Bloch" matrix $J(k)$, obtained by reducing J to the linear manifold of Bloch–Floquet solutions $\psi_l=e^{ikl/p}u_l$, where $u_{l+p}=u_l$, $l\in\mathbf{Z}$, and $k\in[-\pi,\pi]$ is known as the *quasimomentum*. The eigenvalues $\lambda_l(k)$, $l=1,\dots,p$ of $J(k)$ are the band functions of J. They determine the spectrum of J and since they are the roots of the equation

$$A(\lambda,k)\equiv\det\left(J(k)-\lambda\right), \tag{88}$$

the density of states $d(\lambda)$ of the matrix J can be written as

$$d(\lambda)=\frac{1}{2\pi p}\sum_{l=1}^{p}\int_{-\pi}^{\pi}\delta(\lambda-\lambda_l(k))\,dk=\frac{1}{2\pi p}\sum_{l=1}^{p}\int_{-\pi}^{\pi}\delta(A(\lambda,k))A'_\lambda(\lambda,k)\,dk. \tag{89}$$

On the other hand, it can be shown either by computing $A(\lambda,k)$ directly or by using the monodromy (transfer) matrix technique that

$$A(\lambda,k)=\Omega(z)-2s\cos k,\quad s=\prod_{m=1}^{p}r_m, \tag{90}$$

where $\Omega(z)$ is a polynomial of degree p independent of k. By using this form of $A(\lambda,k)$ we obtain from (89)

$$d(\lambda)=-\frac{\Omega'(z)}{\pi p}\left(\Omega^2(z)-4s^2\right)_+^{-1/2} \tag{91}$$

Comparing this formula with (80) we get the relation

$$\Delta(z) = \Omega(z)/s \tag{92}$$

that is well known in spectral theory.

In the case of the p-periodic Jacobi matrix $J(g)$ defined by coefficients (47), s and Ω depend on g. This dependence can be studied by using the relation (70). Namely, by using (47) we can rewrite (70) as the system of p equations

$$r_{m-1}V'(J(g))_{m,m-1} = g, \quad m = 1, \dots, p \tag{93}$$

for the coefficients r_m, $m = 1, \dots, p$ of the p-periodic matrix $J(g)$. Analyzing the system we can find that at least in the case when all r_m, $m = 1, \dots, m$ are pair-wise different we obtain that

$$s^2 = \prod_{m=1}^{p} r_m^2 = g, \tag{94}$$

and that $\Omega(z)$ is independent of g. Thus in this case the dependence of s^2 and of Ω on g is simple and (92) becomes

$$\Delta(\lambda, g) = \Omega(\lambda)/\sqrt{g}, \tag{95}$$

Now, by using this formula, (80), (67), and identifying $\Omega(\lambda)$ with $v(\lambda)$ we obtain our form (17) of the of the equilibrium measure. Knowing the measure we can compute corresponding form (23)–(24) of the function w defined by (22). This form and (18), (20), and (28) lead evidently to (15).

One more link of results of the Section 1 with spectral theory of periodic Jacobi matrices, can be given by using the Thouless formula. The formula is valid in a general case of Jacobi matrices with ergodic (in particular periodic) coefficients [21]. Recall first the definition of the Lyapunov exponent. Given a solution $y_l(\lambda, \alpha)$ of the finite-difference equation (77), satisfying the conditions

$$y_0 = \sin\alpha, \quad y_1 = \cos\alpha, \quad \alpha \in [0, \pi[,$$

define the *Lyapunov exponent* $\gamma(\lambda)$ of the equation as

$$\gamma(\lambda) = \lim_{l\to\infty} (2l)^{-1} \sup_{\alpha\in[0,\pi[} \log[\, y_l^2(\lambda,\alpha) + y_{l+1}^2(\lambda,\alpha)\,].$$

The Lyapunov exponent γ and the Integrated Density of States ν of a respective Jacobi matrix with ergodic coefficients are related by the Thouless formula [21]:

$$\gamma(\lambda) = -\lim_{n\to\infty} n^{-1} \sum_{l=1}^{n} \log r_l + \int_{\sigma(J)} \log|\lambda - \mu|\, \nu(d\mu), \tag{96}$$

where $\sigma(J)$ is the spectrum of the respective Jacobi matrix. In the particular case of a p-periodic Jacobi matrices the first term in the r.h.s. of (96) can be written as

$$p^{-1} \sum_{m=1}^{p} \log r_m = (2p)^{-1} \log \prod_{m=1}^{p} r_m^2.$$

In view of (94), the last expression is equal to $-(2p)^{-1} \log g$ and we obtain the respective relation between the Lyapunov exponent $\gamma_g(\lambda)$ and the density of states $d_g(\lambda)$ of the periodic Jacobi matrix $J(g)$:

$$\gamma_g(\lambda) = -(2p)^{-1} \log g + \int_{\sigma(J(g))} \log |\lambda - \mu| \, d_g(\mu) \, d\mu. \tag{97}$$

It is clear from this formula that

$$\gamma_g(\lambda) = -\operatorname{Re} w_g(\lambda + i0), \tag{98}$$

where (cf (22))

$$w_g(z) = (2p)^{-1} \log g - \int_{\sigma(J(g))} \log(z - \mu) \, d_g(d\mu). \tag{99}$$

Applying to this relation the operation

$$g^{-1} \int_0^g \ldots d\xi,$$

we obtain in view of (67) that

$$g^{-1} \int_0^g w_\xi \, d\xi = -(2p)^{-1}(\log g - 1) + \int_{\sigma_g} \log |\lambda - \mu| \rho_g(\mu) \, d\mu,$$

where σ_g is defined in (16) and $\sigma_g = \bigcup_{\xi=0}^{g} \sigma(J(\xi))$. Thus we have the relation

$$w(z) = g^{-1} \int_0^g w_\xi \, d\xi.$$

In particular, setting here $z = \lambda + i0$ and using (18), (28) and (98) we obtain a relation between the Lyapunov exponent of the p-periodic matrix $J(g)$ and the potential (15) of the random matrix ensemble (1)–(2)

$$\int_0^g \gamma_\xi(\lambda) d\xi = -\frac{1}{2p} V(\lambda) + \text{const}, \ \lambda \in \sigma_g. \tag{100}$$

This formula was obtained in [6] in the variational context.

References

1. Akhiezer, N. (1965) *Classical Moment Problem*, Haffner, N.Y.
2. Albeverio, S., Pastur, L. and Shcherbina, M. (1997) On asymptotic properties of certain orthogonal polynomials, *Math. Phys., Anal., and Geom.*, **4**, 263–277.
3. Bleher, P. and Its, A. (1999) Semiclassical asymptotics of orthogonal polynomials, Riemann–Hilbert problem, and universality in the matrix model, *Ann. of Math.*, **150**, 185–266.
4. Boutet de Monvel, A., Pastur, L. and Shcherbina, M. (1995) On the statistical mechanics approach to the random matrix theory: the integrated density of states, *J. Stat. Phys.*, **79**, 585–611.
5. Brezin, E., Itzykson, C., Parisi, G. and Zuber, J.-B. (1997) Planar diagrams, *Comm. Math. Phys.*, **59**, 759–782.
6. Buyarov, V. S. and Rakhmanov, E. A. (1999) Families of equilibrium measures in an external field on the real axis, *Sbornik: Mathematics*, **190**, 791–802.
7. Deift, P., Kriecherbauer, T. and McLaughlin, K. (1998) New results on the equilibrium measure for logarithmic potentials in the presence of external field, *J. Approx. Theory*, **95**, 388–475.
8. Deift, P., Kriecherbauer, T., McLaughlin, K., Venakides, S. and Zhou, X. (1999) Uniform asymptotics for polynomials orthogonal with respect to varying exponential weights and applications to universality questions in random matrix theory, *Comm. Pure Appl. Math.*, **52**, 1335–1425.
9. Di Francesco, P., Ginsparg, P. and Zinn-Justin, J. (1995) 2D gravity and random matrices, *Phys. Rep.*, **254**, 1–133.
10. Dyson, F. (1962) Statistical theory of the energy level spectra of complex systems, *J. Math. Phys.*, **3**, 140–175.
11. Johansson, K. (1998) On fluctuations of eigenvalues of random Hermitian matrices, *Duke Math. J.* **91**, 151–204.
12. Kuijlaars, A. B. J. and McLaughlin, K. (2000) Generic behaviour of the density of states in random matrix theory and equilibrium problem of real analytic external field, *Comm. Pure Appl. Math.*, **53**, 736–785.
13. Kuijlaars, A. B. J. and Van Assche, W. (1999) The asymptotic zero distribution of orthogonal polynomials with varying recurrence coefficients, *J. Approx. Theory*, **99**, 167–197.
14. Lenoble, O. and Pastur, L. (2001) On the asymptotic behavior of correlators of multi-cut matrix model, *J. Phys. A: Math. Gen.*, **34**, L409–L415.
15. Marchenko, V. and Ostrovski, I. (1975) A characterization of the spectrum of the Hill operator, *Math. USSR Sb.*, **25**, 493–554.
16. Mehta, L. (1991) *Random Matrices*, AP, N.Y.
17. Nevai, P. (1986) Géza Freud, orthogonal polynomials, and Christoffel functions, *J. Approx. Theory*, **48**, 3–117.
18. Pastur L. (1992) On the universality of the level spacing distribution for some ensembles of random matrices, *Letters in Math. Phys.*, **5**, 259–265.
19. Pastur, L. (1996) Spectral and probabilistic aspects of matrix models, in A. Boutet de Monvel and V. Marchenko (Eds.), *Algebraic and Geometric Methods in Mathematical Physics*, Kluwer, Dodrecht, 1996, 207–242.
20. Pastur, L. (2000) Random Matrices as paradigm, in A. Grigoryan, T. Kibble and B. Zegarlinski (Eds.) *Mathematical Physics 2000*, Imparial College Press, London, 216–266.
21. Pastur, L. and Figotin, A. (1992) *Spectra of Random and Almost Periodic Operators*, Springer, Berlin.
22. Pastur, L. and Shcherbina M. (1997) Universality of the local eigenvalue statistics for a class of unitary invariant matrix ensembles , *J. Stat. Phys.*, **86**, 109–147.
23. Saff, E. B. and Totik, V. (1997) *Logarithmic Potentials with External Fields*, Springer–Verlag, N.-Y.
24. Szego, G. (1975) *Orthogonal Polynomials*, AMS, Providence.
25. Teschl, G. (2000) *Jacobi Operators and Completely Integrable Systems*, AMS, Providence.
26. Toda, M. *Theory of Non-linear Lattices*, Springer–Verlag, Berlin, 1989
27. Wigner, E. (1959) Statistical properties of real symmetric matrices with many dimensions, in *Proceedings of 4-th Cand. Math. Congress*, Toronto, 174–185.

COMBINATORICS AND PROBABILITY OF MAPS

V. A. MALYSHEV
I. N. R. I. A.

Abstract. We give an introductory pedagogical review of rigorous mathematical results concerning combinatorial enumeration and probability distributions for maps on compact orientable surfaces, with an emphasize on applications to two-dimensional quantum gravity.

1. Introduction

We give an introductory pedagogical review of rigorous mathematical results concerning combinatorial enumeration and probability distributions for maps on compact orientable surfaces. The main goal of this paper is to present the main ideas concerning central unsolved mathematical problems in the field related to quantum gravity and strings. I could not include, for example, billions of results concerning various classes of plane maps, there are very nice results among them but most of them just confirm that those classes of maps belong to the same universality class. The reader will find such results in more special reviews.

We give also a rigorous description of some terminology, related to discrete quantum gravity. Bibliographical reviews on quantum gravity see [2–4]. Earlier mathematical ideas see in [1, 16], some fresh mathematical exposition of some parts of discrete quantum gravity see in [34]. We do not touch an enormous number of relations of maps with other fields of mathematics like moduli of curves [8], Galois theory see [5, 6], topology [28].

After the definition of map itself it is important to know main examples of subclasses maps, they are defined via some topological restrictions on maps. Next we give an introduction to Tutte's enumeration theory, using partly different techniques. After this we present elements of dynamical triangulation calculus. A short exposition of random matrix approach is given as well.

Gibbs families are defined as a far going generalization of Gibbs distributions. For easiest subclass of maps, called planar Lorentzian models, we present more or less complete results. The last section is devoted to the most difficult problem, maps with matter fields on them. Rigorous results can only be obtained now with cluster expansion techniques.

I would like to thank Prof. E. Brezin, V. Kazakov, V. Liskovets, A. Okounkov, A. Vershik for valuable discussions and information.

V. A. Malyshev and A. M. Vershik (eds.),
Asymptotic Combinatorics with Application to Mathematical Physics, 71–95.

2. Maps are defined

Consider a compact closed orientable smooth surface S_ρ of genus ρ, that is a sphere with ρ handles. It is useful but not necessary to have a differentiable structure (which is known to be unique) on it. Denote also $S_{\rho,k}$ the surface S_ρ with k holes (that is k disks deleted), thus $S_{\rho,0} = S_\rho$. The boundary of $S_{\rho,k}$ consists of k smooth circles. A smooth map on $S_{\rho,k}$ is a triple $(S_{\rho,k}, G, \phi)$, where G is a connected graph, considered as a one-dimensional complex. G may have loops and multiple edges. Let $V(G), L(G)$ are the sets of vertices and (open) edges of G. ϕ is an embedding of G into $S_{\rho,k}$ such that the following conditions are satisfied:

1. for each edge $l \in L(G)$ $\phi(l)$ is a smooth curve on $S_{\rho,k}$;
2. the connected components (called cells or faces) of the complement $S_{\rho,k} \setminus \phi(G)$ are homeomorphic to an (open) disk. It follows that if $k > 0$ then the boundary is contained in $\phi(L(G) \cup V(G))$.

Two smooth maps $(S_{\rho,k}, G, \phi), (S'_{\rho,k}, G', \phi')$ are called equivalent if there is a one-to-one homeomorphism $f : S_{\rho,k} \to S'_{\rho,k}$, respecting orientation and such that $f : \phi(V(G)) \to \phi'(V(G'))$ and $f : \phi(L(G)) \to \phi'(L(G'))$ are one-to-one. Combinatorial map is an equivalence class of smooth maps. Further on we consider only combinatorial maps and call them maps for shortness.

Combinatorial definition of maps

There is a pure combinatorial definition of maps, that does not use surfaces at all. It is based on the notion of a ribbon (or ordered, or fat) graph: graph with a cyclic order of edge-ends at each vertex. Edge-end (or leg) is a pair (v, l) where v is a vertex and l is one of its incident "half-edges".

The combinatorial definition starts with a triple (E, ω, P), where E is a finite set with even number of elements and two permutations ω and P. It is assumed that all cycles of ω have length 2 and that the group generated by ω, P acts transitively on E. For each map $(S_{\rho,k}, G, \phi)$ one can canonically define a triple (E, ω, P) so that E is the set of edge ends of G, ω interchanges two edge-ends of each edge, and for each vertex v P is a cyclic permutation of the set of all edges incident to v corresponding to the clockwise order for a given orientation of $S_{\rho,k}$. Inverse construction is given by the following theorem. In combinatorics community it is refered to Edmonds [7], but algebraists indicate that it goes back to Hamilton.

Theorem 1. *For any triple (E, ω, P) there is a unique map T where the factor set E/ω is the set of edges, the factor set E/P is the set of vertices (with a clockwise order of incident edges), the factor set $E/(\omega P)$ is the set of faces, denoted by $F(T)$.*

The construction of the map proceeds by induction: on each step one face from $E/(\omega P)$ is appended by identifying corresponding edges.

Fat (ribbon) graph can be obtained from an ordinary graph by replacing each edge with fat edge, thus an edge will have two distinct sides. If vertex v has $l(v)$

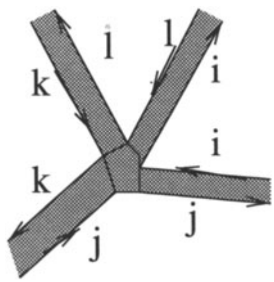


Figure 1. Vertex of a ribbon graph

incident edges then $2l(v)$ sides will also have cyclic order, and in a neighbourhood of v they will look as in the Figure 1. This combinatorial definition was used also in algebraic problems, for example, concerning moduli space of curves, see [8].

Classes of maps

Denote $\mathcal{A}$ the class of all maps. There are many subclasses of $\mathcal{A}$, defined by some restrictions on maps. We give a few number of important examples.

1. Maps of genus ρ, that is all maps with fixed ρ and any k. In particular planar maps have $\rho = 0$. See a long list of subclasses of planar maps in [9, 10].

2. We define triangulations (called quasi-triangulations in [11]) by the following restrictions: the boundary of each cell consists exactly of three edges and the map is nonseparable, thus multiple edges are allowed but no loops. For example, let $\mathcal{A}_{0,1}(N, m)$ be the class of all triangulations of $S_{0,1}$, that is of the disk, with N triangles and m edges on the boundary of the disk. Note that all combinatorial triangulations can be obtained by taking N copies of a triangle and identifying some pairs of their edges, that is why their number is finite for fixed N. Note also that the class of simplicial complexes is a subclass of triangulations in this terminology.

3. Rooted triangulations $\mathcal{A}_{0,1}^{0}(N, m)$ of $S_{0,1}$ that is in each triangulation one boundary edge is distinguished with one of its vertices, this edge is called the root. Equivalence relation is modified correspondingly so that the mappings f respect the roots.

4. One face maps on S_ρ, that is all $T \in \mathcal{A}$ with $|F(T)| = 1$.

5. Slice-triangulations. We define slice-triangulations (in physical literature they are called Lorentzian models) of the 2-dim cylinder $S_{0,2}$. Consider triangulations T of the cylinder $S^1 \times [M, N]$, where S^1 is a circle, $M < N$ are integers. Assume the following properties of T: each triangle belongs to some strip $S^1 \times [j, j+1]\,, j = M, \ldots, N-1$, and has all vertices and exactly one edge on the boundary $(S^1 \times \{j\}) \cup (S^1 \times \{j+1\})$ of the strip $S^1 \times [j, j+1]$. Let

$k_j = k_j(T)$ be the number of edges on $S^1 \times \{j\}$. We assume $k_j \geq 1$. Then the number of triangles $F = F(T)$ of T is equal to

$$F = 2 \sum_{j=M+1}^{N-1} k_j + k_M + k_N. \tag{1}$$

3. Enumeration for fixed genus

3.1. GENUS ZERO

Starting from 1962 Tutte publishes a series of papers where he solves the enumeration (censoring) problem for various classes of planar maps. He invents a new method of solving such problems which was developped in hundreds of papers. The heart of this method consists of two parts: deleting of an edge to get recurrent equation and a method to solve the resulting quadratic functional equation, containing two unknown generating functions: from two and one variables correspondingly. I use this occasion to note that some years later quite independently I developped (in connection with the Riemann–Hilbert problem for two complex variables) a new method for solving linear functional equation with three unknown functions: one function of two variables and two functions of one variable. These methods are quite different but have one common point—projection of the equation on some algebraic curve.

Let $C_0(N, m)$ be the number of triangulations in $\mathcal{A}_{0,1}^0(N, m)$, that is with N triangles and m edges on the boundary. It is not difficult (see [13]) to prove apriori bounds: there exist $0 < C_1 < C_2 < \infty$ such that for all N and m

$$C_1^N < C_0(N, m) < C_2^N$$

There are more general results concerning such exponential apriori bounds for the number of triangulations of a manifold for dimensions more than 2, see [17].

More difficult is to find exact asymptotics. One of the theorems by Tutte can be formulated as follows

Theorem 2. *If $N \to \infty$ and m is fixed then*

$$C_0(N, m) \sim cN^{\alpha}\gamma^N$$

where $\gamma = \sqrt{\frac{27}{2}}, \alpha = -\frac{5}{2}, c = c(m)$.

Proof. The following recurrent equations

$$C_0(N, m) = C_0(N - 1, m + 1) + \sum_{N_1+N_2=N-1, m_1+m_2=m+1} C_0(N_1, m_1)C_0(N_2, m_2)$$

can be obtained by deleting the rooted edge. After deletion of the rooted edge there is a rule to define a new rooted edge for each of the resulting maps. This is

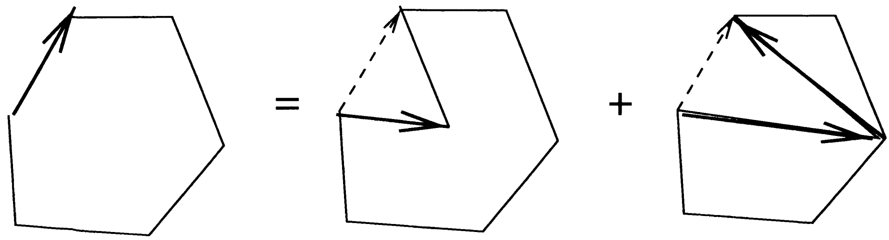


Figure 2. Recurrent equation

shown on Figure 2. There are two possibilities which correspond to the linear and quadratic terms correspondingly.

It is very convenient to assume the following conditions which will be the boundary conditions for the systems of equations below

$$C_0(N,0) = C_0(N,1) = 0, C_0(0,m) = \delta_{m,2}, C_0(1,m) = \delta_{m,3}$$

Only the case $N = 0$, $m = 2$ needs comment: this corresponds to a degenerate disk, an edge with two vertices.

Multiplying (7) on $x^N y^m$ and summing $\sum_{N=0}^{\infty}\sum_{m=2}^{\infty}$ we get the following equation in a small neighbourhood $\Omega \in C^2$ of $x = y = 0$

$$F(x,y) = F(x,y)xy^{-1} + F^2(x,y)xy^{-1} + y^2 - xyF_2(x) \tag{2}$$

where we introduced the generating functions

$$F(x,y) = \sum_{N=0}^{\infty}\sum_{m=2}^{\infty} C_0(N,m)x^N y^m, \quad F_m(x) = \sum_{N=0}^{\infty} C_0(N,m)x^N.$$

F and F_2 are unknown functions.

Here we solve the functional equation (2). We rewrite it in the following form

$$(2xF(x,y) + x - y)^2 = 4x^2y^2F_2(x) + (x-y)^2 - 4xy^3 \tag{3}$$

and denote D its righthand side. Consider the analytic set $\{(x,y) : 2xF+x-y = 0\}$ in a small neighbourhood of $x = y = 0$. Note that it is not empty, $(0,0)$ belongs to this set and it defines a function $y(x) = x + O(x^2)$ in a neighbourhood of $x = 0$. In particular, it will be shown that $y(x)$ and $F_2(x)$ are algebraic functions. Because of the square in the righthand side of (3) we have two equations valid at the points of this analytic set

$$D = 0, \quad \frac{\partial D}{\partial y} = 0$$

or

$$4x^2y^2F_2(x) + (x-y)^2 - 4xy^3 = 0 \tag{4}$$

$$8x^2yF_2(x) - 2(x-y) - 12xy^2 = 0$$

from where one can exclude the function $F_2(x)$ by multiplying second equation (4) on $\frac{y}{2}$ and subtracting it from the first equation. Then

$$y = x + 2y^3 \tag{5}$$

or

$$y = \frac{x}{1-2y^2} \tag{6}$$

By the theorem on implicit functions this equation gives the unique function $y(x)$, analytic for small x with $y(0) = 0$. It is evident from (6) that the convergence radius of $y(x)$ is finite. Note that $y(x)$ is odd and $F_2(x)$ is even, because for any triangulation $N - m$ is even.

We continue to rededuce here Tutte results in a different way. $y(x)$ is an algebraic function satisfying the equation $y^3 + py + q = 0$ with $p = -\frac{1}{2}, q = \frac{x}{2}$. The polynomial $f(y) = y^3 + py + q$ can have multiple roots only when $f = f'_y = 0$, which gives $x_\pm = \pm\sqrt{\frac{2}{27}}$. These roots are double roots because $f''_y \neq 0$ at these points. For $x_+ = \sqrt{\frac{2}{27}}$ we have $y_+ = y(x_+) = \frac{1}{\sqrt{6}}$, that can be seen from $f'_y = 3y^2 - \frac{1}{2} = 0$ and $f = 0$. From (6) it also follows that $x(-y) = -x(y)$ and thus $y(x)$ is odd. It follows that $y(x)$ has both $x_\pm = \pm\sqrt{\frac{2}{27}}$ as its singular points.

From (4) we know $F_2(x)$ explicitely, after that $F(x, y)$ is explicit from equation (3). The unique branch $y(x)$, defined by equation (6), is related to the unique branch of $F_2(x)$ by the equation

$$F_2 = \frac{(1 - 3y^2(x))}{(1 - 2y^2(x))^2} = x^{-2}y^2(1 - 3y^2)$$

that is obtained by substituting $x = y - 2y^3$ to the first equation (6).

We know that $F_2(x)$ has positive coefficients, that is why $x = \sqrt{\frac{2}{27}}$ should be among its first singularities. Then $x = -\sqrt{\frac{2}{27}}$ should also be a singularity of both $y(x)$ and $F_2(x)$.

The principal part of the singularity at the double root x_+ is $y(x) = A(x - x_+)^{d+\frac{1}{2}}$ for some integer d. As $y_+ = y(x_+)$ is finite then $d \geq 0$. At the same time $y'(x) = \frac{1}{1-6y^2(x)}$ that is ∞ for $x = x_+$. It follows that $d = 0$. For F_2 we have the same type of singularity $A(x - x_+)^{d+\frac{1}{2}}$ but here $d = 1$ as $F_2(x_+)$ and $F'_2(x_+)$ are finite but $F''_2(x_+)$ is infinite.

The theorem follows from this. We proved also that the generating functions are algebraic. One can check also that

$$F(x, y) = \frac{y - x}{2x} - \frac{(y - y(x))\sqrt{\frac{x^2}{y^2(x)} - 4xy}}{2x}.$$

$\square$

Automorphism groups. How from the results for rooted triangulations to get results for unrooted? Everything here is based on the following principle. Let

$\mathcal{A}_{01}(N,m)$ be the class of unrooted triangulations of the disk with N triangles and m edges on the boundary, let $|\mathcal{A}_{01}(N,m)| = C(N,m)$ and let $C^{nontr}(N,m)$ be the number of such maps with nontrivial automorphism group.

Theorem 3. *For almost all maps in the classes $\mathcal{A}_{01}(N)$ and $\mathcal{A}_0(N)$ the automorphism group is trivial, more exactly, for fixed m*

$$\frac{C^{nontr}(N,m)}{C(N,m)} \underset{N\to\infty}{\longrightarrow} 0$$

There are many proofs of this theorem for many classes of maps, see [9]. This is convenient to do in two steps.

Lemma 4. *If m is fixed and $N \to \infty$ then*

$$C(N,m) \sim \frac{1}{m} C_0(N,m)$$

Proof. Let us enumerate the edges of the boundary $1, 2, \ldots, m$ in the cyclic order, starting from the root edge. Automorphism ϕ of the triangulation of the disk is uniquely defined if the image $j = \phi(1)$ of the edge 1 is fixed. In fact, then the triangle adjacent to the edge 1 has to be mapped by ϕ onto the triangle adjacent to j, and so on by connectedness.

Consider the strip of width 1, adjacent to the boundary, that is the set of triangles of three types: those which have a common edge with the boundary (type 1), two common adjacent edges (type 2) and those, having with the boundary only a common vertex (type 0). Thus the strip, that is the sequence of triangles can be identified with the word $\alpha = x_1 \ldots x_n$, $n > m$, where $x_i = 0, 1, 2$ represent the triangle types. Consider the set $W(m, n_0, n_1, n_2)$ of words with given m and numbers n_i of symbols $i = 0, 1, 2$. An automorphism of a disk gives a cyclic automorphism of the word α. The sets $W(m, n_0, n_1, n_2)$ are invariant however. Note that

$$m = n_1 + 2n_2, \quad n_0 \geq n_1 + n_2$$

and the length of other boundary of the strip equals $m' = n_0$. Thus, if there is no nontrivial cyclic automorphism of the word, then there are no automorphisms of the whole triangulation of the disk. It is easy to see, that for a given sequence $n_0(m)$ and as $m \to \infty$ the set of words from $\cup_{n_1,n_2} W(m, n_0, n_1, n_2)$, having nontrivial cyclic automorphisms, is asymptotically zero compared with the number of words in $\cup_{n_1,n_2} W(m, n_0, n_1, n_2)$. $\square$

Let now $C(N)$ be the number of triangulation of the sphere with N triangles.

Lemma 5. *If $N \to \infty$ then*

$$C(N) \sim \frac{1}{3N} C_0(N,3).$$

Proof of this intiotively clear assertion can be obtained along similar lines.

3.2. ARBITRARY GENUS

Let $C(N,\rho,0)$ be the number of (unrooted) triangulations of S_ρ, and $C_0(N,\rho, k+1; m, m_1,\dots,m_k)$, $m, m_1,\dots,m_k \geq 2$, be the number of triangulations of $S_{\rho,k+1}$, $k = 0,1,2,\dots$, where one boundary (with m edges) has a rooted edge, and where the remaining k boundaries have $m_1,\dots,m_k$ edges correspondingly.

Theorem 6. *Then for fixed $\rho, k, m_1,\dots,m_k$ as $N \to \infty$*

$$C(N,\rho,0) \sim f(\rho)N^{a\rho+b}\gamma^N, \quad a = \frac{5}{2},\ b = -\frac{5}{2} - 1,\ \gamma = \sqrt{\frac{27}{2}}$$
$$C_0(N,\rho,k;m,m_1,\dots,m_k) \sim f(\rho,k,m,m_1,\dots,m_k)N^{a\rho+b+1+k}\gamma^N.$$

There are many similar results on enumerating such maps but not in terms of the number of triangles, see [14, 15]. This theorem can be proved by using Tutte's idea of deleting the rooted edge. We shall see what are the resulting recurrent equations. Start with some triangulation with parameters $N, \rho, k+1, m, m_1,\dots,m_k$.

The operation of deleting the rooted edge can produce more possibilities than in case $\rho = 0, k = 1$:

1. neither ρ nor k change, only $m \to m+1$, $N \to N-1$;
2. the hole merges with another hole, thus ρ is not changed, k becomes less by 1;
3. the hole cuts a handle, thus $\rho \to \rho - 1$, $k \to k+1$;
4. the hole cuts the surface itself thus producing two surfaces with parameters ρ_i, k_i, $i = 1,2$, such that $\rho_1 + \rho_2 = \rho$, $k_1 + k_2 = k+1$.

For example, we get a closed system of equations for $\rho = 0$, ρ cannot increase. Moreover, for $\rho = 0$ the parameter k cannot increase, moreover equations for $C_0(N,0,k;m,m_1,\dots,m_k)$ are linear assuming we know $C_0(N,0,j;m,m_1,\dots,m_j)$ for $0 \leq j < k$. Thus, equations for the generating functions are nonlinear only on the first step with $\rho = 0$, $k = 0$. Afterwards (for $\rho = 0$, $k > 1$ and for $\rho > 0$) the equations for the generating functions become linear, however very bulky. One can find treatment of similar problems in [14, 15] and references therein.

Universality classes. We considered only the case of triangulations. There are many results (see [9]) showing that γ (and also c) strongly depends on the class of maps, on the contrary α (called the critical exponent) does not. However, γ does not seems not to depend on $\rho, k, m, m_1,\dots,m_k$. An example of another universality class is the class of slice-triangulations, see below.

4. What is two-dimensional gravity

Metric structure on the surface related to a given triangulation is defined once it is defined for each closed cell so that the lengths of the edges are compatible. There are two basic approaches for defining such metric structure. In the Dynamical Triangulation approach all triangles are identical, lengths of edges are equal a and the metrics inside triangles is the standard euclidean metrics. In the Quantum Regge Calculus they are random.

We will use further on the Dynamical Triangulation approach. One can show then that the differentiable structure exists such that the metrics is sufficiently smooth everywhere except for the vertices, where the number q_v of triangles adjacent to the vertex v is different from 6. One can prove this by induction putting pairs of adjacent triangles on the plane.

We can show that the curvature is zero everywhere except for the vertices v with $q_v \neq 6$, in those vertices the curvature becomes discontinuous. In fact, the curvature is measured using the parallel transport (Levi–Civita connection). That is the curvature should be proportional to the difference of the angles for initial and transported vectors of a vector (lying in the plane of the triangle) along a small closed path. If the path lies inside a triangle then the angle is zero as on the euclidean plane, if the path encircles a point on some edge then it is zero (by unfolding the two half planes separated by this edge). Only paths around vertices may give nonzero difference. Around the vertex v the angle between the initial and the transported vector is $\varepsilon_v = 2\pi - \sum_f \varphi_{fv} = \frac{\pi}{3}(6 - q_v)$, where φ_{fv} is the angle of the simplex f at vertex v. Note that

$$2\pi V - \sum_v \varepsilon_v = \sum_v \sum_f \varphi_{fv} = \sum_f \sum_v \varphi_{fv} = \pi F$$

Using the Euler formula

$$\chi = 2 - 2\rho = V - L + F$$

one can get from this the Gauss–Bonnet formula

$$2 \sum_v \varepsilon_v = 4\pi\chi$$

for triangulations, using $L = \frac{3F}{2}$.

Einstein–Hilbert action on the smooth manifold for the so called pure gravity (that is without matter fields) is

$$\int (c_1 R + c_2)\sqrt{g}\, dx$$

where R is the gaussian curvature, g-metrics. By Gauss–Bonnet formula for smooth surfaces $\int R\sqrt{g}\, dx = 4\pi\chi$. Thus the discrete action should be (up to a constant) $\lambda\rho + \mu N$, where ρ is the genus and N is the number of triangles.

When the genus is fixed, that is the surface itself is fixed (but not its metrics), the first term becomes superfluous, and pure gravity (without matter fields) is defined by the following grand canonical partition function

$$Z = \sum_T \exp(-\mu N(T)) = \sum_N C(N) \exp(-\mu N)$$

with the corresponding canonical partition function

$$Z_N = \sum_{T:|F(T)|=N} \exp(-\mu N) = C(N) \exp(-\mu N)$$

There is a critical point $\mu_{cr} = \log \gamma$. If $\mu > \mu_{cr}$ then $Z < \infty$, and if $\mu < \mu_{cr}$ then $Z = \infty$. In the critical point the terms of the series have power decrease (or increase), defined by the critical exponent α. This critical point does not depend on ρ.

However, if the surface has a boundary then the situation is different, there are some extra degrees of freedom (see more about this in the section 7.2). If the surface has a boundary then Euler formula is

$$\chi = V - L + F + k$$

where k is the number of components of the boundary. Gauss–Bonnet formula has the form

$$2\sum_v \varepsilon_v = 4\pi(\chi - k + \partial L)$$

where ∂L is the number of edges (or vertices) on the boundary, if for the vertices v of the boundary we define ε_v by $\pi - \varepsilon_v = \frac{\pi}{6} q_v$.

4.1. SOME CENTRAL PROBLEMS

The situations when ρ is fixed and when ρ is random are physically quite different. The first one corresponds in physics to string diagrams, the second—to the so called spin foam, when there are local fluctuations of the topology. In fact, in this case ρ is of the order N.

Spin foam. If the genus is not fixed then the grand canonical partition function would be

$$Z = \sum_T \exp(-\mu N(T) - \lambda \rho(T))$$

where T runs over all triangulations. However one can easily show that this series is divergent for any μ, λ. In fact, the number of maps with N triangles irregardless to genus has a factorial growth, that cannot be compensated by exponential factors. Thus it is natural to consider the canonical partition function

$$Z_N = \sum_{T:|F(T)|=N} \exp(-\lambda \rho(T))$$

instead of the grand canonical. Unfortunately, there are no results concerning this canonical partition function—this is one of the central unsolved problems. However, there are another interesting possibilities to choose a canonical ensemble. For example,

$$Z_N = \sum_{T:L(T)=N} \exp(-\lambda\rho(T) - \mu F(T)).$$

For this case we shall get some results in the next section.

Strings. Assume now that to each triangle f of $F(T)$ there corresponds a spin σ_f, taking its values in the space S with some positive measure μ_0 on it. For some symmetric function $\Phi : S \times S \to R$ define the following partition function

$$Z_N = \sum_{T:N(T)=N} \prod_{f\in|F(T)|} \int d\mu_0(\sigma_f) \exp(-\lambda\rho(T) - \beta \sum_{<f,f'>} \Phi(\sigma_f, \sigma_{f'})).$$

The case $S = R^d$ and $\Phi(\sigma, \sigma') = (\sigma - \sigma')^2$ corresponds to Polyakov action for strings with the target space-time of dimension d. There are no results (even on the physical level) concerning calculating this partition function even for $\rho = 0$, random matrix models do not work here. The latter circumstance is quite surprising because another physical approache (Hamiltonian quantization) to free quantum boson strings work quite well.

Z_N gives an example of probability distributions which we call Gibbs families below. We shall say some word about the general theory of such probability distributions.

Continuum limit. It is reasonable to assume that under some scaling limit random discrete spaces converge in the Hausdorf–Gromov distance to a smooth manifold. Even for simple examples it is a formidable problem to prove such a convergence.

Conventional string theory approach. An alternative approach to two-dimensional gravity, see [29], is similar to the conventional continuum path integral approach, and its more refined BRST procedure. String theory approach to treat the continuum analog of Z_N consists heuristically in the following remark. Due to invariance of the action with respect to the diffeomorphism group G_1 and the group G_2 of Weyl transformations, the path integral can be factored on the integrals over G_1, G_2 and some residual factor. The residual factor can be reduced to the integral over moduli of complex algebraic curves. It seems difficult however to make this illuminating staff well-defined. Relation between the two approaches were discussed in [8].

5. Random genus.

Consider the class $\mathcal{A}^0$ of all rooted maps of S_ρ irregardless to ρ. Let $F_{b,p}(\rho)$ be the number of rooted (one edge is specified together with its direction) maps with

$p+1$ vertex and $b+p$ edges of the closed compact surface of genus ρ. We give a simple asymptotic formula for canonical partition function

$$Z_N(y) = \sum_{b+p=N} \sum_{\rho=0}^{\infty} F_{b,p}(\rho) y^p$$

with any complex y.

The operation of deleting a rooted edge gives the following recurrent equations, see [19], for the numbers $F_{b,p} = \sum_\rho F_{b,p}(\rho)$

$$F_{b,p} = \sum_{j=0}^{p-1} \sum_{k=0}^{b} F_{k,j} F_{b-k,p-j-1} + (2(b+p)-1) F_{b-1,p}, \quad b,p \geq 1. \tag{7}$$

In fact these equations hold also if either $b=0$ or $p=0$, except the case $b=p=0$, if we put

$$F_{0,0} = 1, \quad F_{-1,p} = F_{b,-1} = 0$$

We should say more about the cases $b=0$ and $p=0$. The case $b=p=0$ corresponds to an imbedding of one vertex into the sphere. The cases $p \geq 1$, $b=0$ correspond to trees, imbedded to the sphere, $F_{0,p} = \frac{(2p)!}{p!(p+1)!}$. Numbers $F_{b,p}$ with $p=0, b \geq 1$ are equal to $(2b-1)(2b-3)\ldots = \frac{(2b)!}{b!2^b}$, that is to the number of partitions of $\{1,2,\ldots,2b\}$ onto pairs.

Put

$$b_k = b_k(y) = \sum_{b+p=k} F_{b,p} y^p, b_{-1} = 0, b_0 = 1$$

for $y \geq 0$. The main result is the following theorem.

Theorem 7. *For any complex $y > 0$ and as $k \to \infty$*

$$b_k(y) \sim f(y)\, k!\, 2^k k^{y-\frac{1}{2}}$$

where $c(y) > 0$ is a constant analytic in y.

This result has physical interpretation. Consider the probability measure on the set of general rooted maps with $b+p=N$ (canonical ensemble) with the following partition function

$$Z_N(y) = \sum_{b,p:b+p=N} F_{b,p} \exp(-\lambda b) = \sum_{b,p:b+p=N} F_{b,p} y^p$$

with $y = \exp(-\lambda)$. Then we have the following asymptotics for the partition function

$$Z_N(y) \sim c(y)\, N!\, \exp((\ln 2 - \lambda)N) N^{y-\frac{1}{2}}$$

One can, using $-2\rho - F - m - 2 = -b$, rewrite this in other terms up to a constant factor

$$Z_N(y) = \sum_{T:L(T)=N} \exp(-2\lambda\rho(T) - \mu F(T))$$

only for $\mu = \lambda$. It follows that the critical point $\lambda_{cr} = \ln 2$ and the critical exponent is $\exp(-\lambda) - \frac{1}{2}$.

6. Random matrix techniques

For mathematical initialization to this beautiful methods we refer the reader to [24], complete but more physical presentations one can find in many physical papers, see for example [21].

6.1. FIXED GENUS

One of the central models of random matrix theory is the following probability distribution μ on the set of selfadjoint $n \times n$-matrices $\phi = (\phi_{ij})$ with the density

$$\frac{d\mu}{d\nu} = Z^{-1} \exp(-tr(\frac{\phi^2}{2h}) - tr(V))$$

where $V = \sum a_k \phi^k$ is a polynomial of ϕ, bounded from below, ν is the Lebesgue measure on the real n^2-dimensional space of vectors $(\phi_{ii}, \operatorname{Re}\phi_{ij}, \operatorname{Im}\phi_{ij}, i < j)$. If $V = 0$, then the measure $\mu = \mu_0$ is gaussian with covariances $\langle \phi_{ij}, \phi^*_{kl} \rangle = \langle \phi_{ij}, \phi_{lk} \rangle = h\delta_{ik}\delta_{jl}$. The density of μ with respect to μ_0 is equal to

$$\frac{d\mu}{d\mu_0} = Z_0^{-1} \exp(-tr(V)).$$

For the existence of μ it is necessary that the degree p of the polynomial V were even and the coefficient a_p were positive. For this case there is a deep theory of such models, see [25].

Fundamental connection (originated by Hooft) between matrix models and counting of maps on surfaces is given by the formal series in semiinvariants or in diagrams (see for example [20]) (we neglect $\log Z_0$)

$$\log Z = \sum_{k=1}^{\infty} \frac{(-1)^k}{k!} {<}tr(V), \ldots, tr(V){>} = \sum_{k=1}^{\infty} \frac{(-1)^k}{k!} \sum_{D_k} I(D_k)$$

where ${<}tr(V), \ldots, tr(V){>}$ is the semiinvariant of order k of the random variable $tr(V)$ with respect to the gaussian measure μ_0, $\sum_{D_k}$ is the sum over all connected diagrams D_k with k vertices and $L = L(D_k)$ edges.

Assume, for example, $V = a\phi^4$, $a > 0$. Then each diagram has enumerated vertices $1, \ldots, k$, to each vertex corresponds the random variable $tr(\phi^4)$. Moreover, each vertex has enumerated edge-ends (legs) $1, 2, 3, 4$, corresponding to all factors in $\phi_{ij}\phi_{jk}\phi_{kl}\phi_{li}$, one should further sum over all indices i, j, k, l. By coupling legs we $L = 2k$ edges. The leg, for example corresponding to ϕ_{ij}, can be imagined as a strip (ribbon) (defining thus a ribbon or fat graph), the sides of the leg have indices i and j correspondingly. In a neighborhood of the vertex the ribbons are placed on the surface as on the Figure 1. It is assumed that coupling of ribbon legs is such that the coupled sides have the same indices.

As any index appears even number of times in each vertex, then for a given edge l with index i, there is a unique closed connected path in the diagram consisting of sides with index i, and passing through l. These paths are called index loops. Summing over indices we will get then the factor n^N where $N = N(D_k)$ is the number of index loops in the diagram. Finally we get the following formal series

$$\sum_k \frac{(-a)^k}{k!} \sum_{D_k} h^{2k} n^{N(D_k)} = \sum_k (-4ah^2)^k \sum_{E_k} n^{N(E_k)} = \sum_{k,N} (-4ah^2)^k n^N M(k,N)$$

where in the second sum the summation is over all graphs with unordered set of vertices, and the set of legs of each vertex is only cyclically ordered. While changing from D to E we introduced the factor $\frac{k!4^k}{A(E)}$, where $A(E)$ is the power of the automorphism group of E. We omit $A(E)$, assuming that for "almost all" (as $k \to \infty$) graphs $A(E) = 1$.

In the third sum $M(k, N)$ is the number of maps with k vertices and N faces. Here a map means a map from the class of 4-regular maps, that is each vertex has degree 4. By Edmonds theorem (see section 2) for any graph E there exists a unique (up to combinatorial equivalence) embedding $f(E)$ of this graph to some oriented compact closed surface S_ρ for some genus ρ and such that each index loop bounds an open subdomain of S_ρ homeomorphic to a disk. This map has k vertices, $2k$ edges and N faces. By Euler formula $k = N + 2\rho - 2$ we get the following formal expansion, putting $h = 1$, $a = \frac{b}{4n}$

$$\log Z = \sum_{N,\rho} (-b)^{N+2\rho-2} n^{-2\rho+2} C_\rho(N)$$

where $C_\rho(N)$ is the number of maps with N faces and having genus ρ. It follows that for example for $\rho = 0$, the sum of terms corresponding to maps with N faces and having genus ρ can be formally obtained as

$$\lim_{n\to\infty} \frac{\log Z}{n^2} = \sum_N (-b)^{N-2} C_0(N).$$

There is a beautiful techniques to extract $C_0(N)$ from this, however I do not know any completely rigorous treatment.

There are generalization of the random matrix model with Q matrices M_q, where the action is

$$Tr(\sum_{q=1}^{Q} V(M_q) + \sum_{q=1}^{Q-1} M_q M_{q+1})$$

and its continuum analog, see review [23]. Such generalizations allow to do calculations for some models with spin.

One face maps. There is another direction in the combinatorics of maps, now more related to algebra. One considers triangulations with one (see [26]) or fixed finite number of cells (see [8, 28]). Maps with $F = 1$ for S_0 coincide with trees imbedded in S_0, that is with plane trees. Their number is given by Catalan numbers. Generally, their number was calculated in [26, 27]. This is related to string theory approach to two dimensional gravity, see section 4.1.

7. Gibbs Families

Here we introduce very shortly a unified general framework for discrete random spaces, a generalization of the now classical theory of Gibbs fields (see [30, 31]), which we call Gibbs families ("Gibbs fields on Gibbs graphs").

We define graph with a local structure. Let $\mathcal{F}_d$ be the set of all finite graphs with diameter d.

Definition 8. Let the function $\sigma(\gamma)$ on $\mathcal{F}_d$ be given with values in some set S. Local structure (of diameter d) on the graph G is given by the set of values $\{\sigma(\gamma)\}$, where γ runs all regular subgraphs γ of diameter d of the graph G.

Examples. – Graphs G with some function $\sigma(v)$ on the set $V(G)$ of vertices (spin graphs) correspond to a local structure with $d = 0$.

– Gauge fields on graphs: for each vertex and for each edge values from a group R are defined, in this case one can take $d = 1$.

– Simplicial complex is completely defined the following function on complete regular subgraphs (of its one-dimensional skeleton) of diameter 1: it takes value 1, iff this subgraph defines a simplex of the corresponding dimension, and 0 otherwise. Here one can also take $d = 1$.

– Penrose quantum networks [33, 32]: in a finite graph, each vertex of which has degree 3, to each edge l some integer $p_l = 2s_l$ is prescribed, where halfintegers s_l are interpreted as degrees of irreducible representations of $SU(2)$. Moreover, in each vertex the following condition is assumed satisfied: the sum of p_i, for all three incident edges, is even and any p_i does not exceed the sum of two other values. Then for the tensor product of three representations there exists a unique (up to a factor) invariant element, which is prescribed to this vertex.

We fix some function $s(\gamma)$, defining a local structure. Let $\mathcal{F}^{(s)}$ be the set of all finite graphs with the local structure. Potential is a function $\Phi : \mathcal{F}^{(s)} \to R \cup \{\infty\}$. The energy of a finite graph $\Gamma \in \mathcal{F}^{(s)}$ is defined as

$$H(\Gamma) = \sum_{\gamma \subset \Gamma} \Phi(\gamma)$$

where the sum is over all regular subgraphs $\gamma \in \mathcal{F}^{(s)}$ of Γ. For any subclass $\mathcal{F}_0^{(s)} \subset \mathcal{F}^{(s)}$ the Gibbs family with potential Φ is the following probability distribution on $\mathcal{F}_0^{(s)}$

$$\mu(\Gamma) = Z_{\mathcal{F}_0}^{-1} \exp(-\beta H(\Gamma)), \Gamma \in \mathcal{F}_0^{(s)}$$

$$Z_{\mathcal{F}_0^{(s)}} = \sum_{\Gamma \in \mathcal{F}_0^{(s)}} \exp(-\beta H(\Gamma))$$

it is should be assumed that $Z_{\mathcal{F}_0^{(s)}} \neq 0, \infty$. Another possibility is not to introduce a subclass $\mathcal{F}_0^{(s)} \subset \mathcal{F}^{(s)}$ by hand but to choose "hard core" (that is taking values ∞) potential which distinguishes exactly the subclass $\mathcal{F}_0^{(s)}$, see more details in [34].

For some increasing sequences $\mathcal{F}_1^{(s)} \subset \mathcal{F}_2^{(s)} \subset \cdots \subset \mathcal{F}_n^{(s)} \subset \cdots \subset \mathcal{F}^{(s)}$ of finite graphs with the local structure defined by $s(\cdot)$ one can naturally define weak limits of the Gibbs families with given Φ, β. See a detailed exposition of this theory in [34]: for example, analog of DLR-condition.

7.1. CORRELATION FUNCTIONS FOR MAPS

For maps without a local structure correlation functions define frequencies of "small" subgraphs in the map. We consider only the probabilities of graphs containing the rooted edge. Then using the analogs of the theorem 3 one can get rid of specifying rooted edge. Let us consider the class $\mathcal{A}_{01}^0(N,m)$ of rooted triangulations T of the disk with N triangles and m edges on the boundary. Then the potential $\Phi \equiv 0$ gives equal probabilties for all maps in $\mathcal{A}_{01}^0(N,m)$. Fix some map Γ of the disk and let $p^N(\Gamma)$ be the probability that a neighborhood of the rooted edge is isomorphic to the map Γ.

Theorem 9. *For any* Γ *there exists the limit* $\lim p^N(\Gamma) = \pi(\Gamma)$.

The *proof* is easy. Let Γ have $n(\Gamma)$ triangles, $m(\Gamma)$ edges on its own boundary, among them $m_{in}(\Gamma)$ are internal edges of T. Delete all triangles of Γ from T. This can produce one or more maps of the disk. Each of these maps has at least one boundary edge in common with some of $m_{in}(\Gamma)$ edges. The first one (in the clockwise order) we define to be rooted for that map. However, it can be proven using explicit formulae for $C_0(N,m)$ that the probability, that this deletion produces more than one map, tends to one as $N \to \infty$, this is a corollary of $N^{-\frac{5}{2}}$ factor. Then the theorem follows again from the explicit formulae for $C_0(N,m)$. $\square$

In the combinatorics papers there are many more refined results concerning the distribution of subgraphs, see [9].

7.2. PHASE TRANSITION FOR PLANAR TRIANGULATIONS

We consider triangulations $T \in \cup_m \mathcal{A}_{0,1}^0(N,m)$. The canonical distribution on this set of triangulations is defined by the probability $P_{0,N}(T)$ of triangulation T. One can introduce $P_{0,N}$ in four equivalent (easy to verify) ways:

– by the interaction proportional to the general number of edges, that is

$$P_{0,N}(T) = Z_N^{-1} \exp(-\mu_1 L(T))$$

where $L(T)$ is the number of edges of T, including all boundary edges;

– by the interaction proportional to the sum of the degrees of vertices

$$P_{0,N}(T) = Z_{0,N}^{-1} \exp(-\frac{\mu_1}{2} \sum_{v \in V(T)} \deg v).$$

Thus

$$Z_N = \sum_m \sum_{T \in \mathcal{A}^0_{0,1}(N,m)} \exp(-\frac{\mu_1}{2} \sum_{v \in V(T)} \deg v).$$

Using a duscrete analog of Gauss–Bonnet theorem (see above), one can show that this is a discrete analog of Einstein–Hilbert action. Note that this is a particular case (with parameters $t_q = t$) of the interaction , considered in [22]

$$\prod_{q>2} t_q^{n(q,T)}$$

where $n(q,T)$ is the number of vertices of degree q.

- by the interaction proportional to the number of boundary edges.
- By the Gibbs family with fictitious spins σ_v in the vertices v of the triangulation, taking values in any compact set, and assume that the potential of the Gibbs family is

$$\Phi(\sigma_v, \sigma_{v'}) \equiv 1$$

for any two neighboring vertices v, v'.

We will observe phase transitions with respect to parameter μ_1. It has the critical point $\mu_{1,cr} = \log 12$. Let $\beta_0 = \beta_0(\mu_1)$ be such that

$$\frac{(1 + \frac{4\beta_0}{3(1-\beta_0)})}{(1 + \frac{2\beta_0}{1-\beta_0})^2} \exp(-\mu_1 + \log 12) = 1$$

Theorem 10. *The free energy* $\lim_N \frac{1}{N} \log Z_{0,N} = F$ *is equal to* $-\frac{3}{2}\mu_1 + c$, $c = 3\sqrt{\frac{3}{2}}$, *if* $\mu_1 > \mu_{1,cr}$, *and to*

$$-\frac{3}{2}\mu_1 + c + \beta_0(-\mu_1 + \log 12) + \int_0^{\beta_0} \log \frac{(1 + \frac{4\beta}{3(1-\beta)})}{(1 + \frac{2\beta}{1-\beta})^2} \, d\beta$$

if $\mu_1 < \mu_{1,cr}$.

Note that if $\mu_1 \to \mu_{1,cr}$ then $\beta_0 \to 0$.

Let $m(N)$ be the random length of the boundary when N is fixed. Its probability can be written as, using $|L(T)| = \frac{3N}{2} + \frac{m}{2}$,

$$P_{0,N}(m(N) = m) = \Theta_{0,N}^{-1} \exp(-\mu_1 \frac{m}{2}) C_0(N,m), \Theta_{0,N} = \sum_m \exp(-\mu_1 \frac{m}{2}) C_0(N,m)$$

Theorem 11. *There are 3 phases, where the distribution of* $m(N)$ *has quite different asymptotical behaviour:*

– *Subcritical region, that is* $12 \exp(-\mu_1) < 1$. *Here* $m(N) = O(1)$, *more exactly the distribution of* $m(N)$ *has a limit* $\lim_N P_N(m(N) = m) = p_m$ *for fixed* m *as* $N \to \infty$. *Thus the hole becomes neglectable with respect to* N.

– *Supercritical region (elongated phase), that is* $12\exp(-\mu_1) > 1$. *Here the boundary length is of order* $O(N)$. *More exactly there exists* $\varepsilon > 0$ *such that* $\lim P_{0,N}(\frac{m_N}{N} > \varepsilon) = 1$.
– *In the critical point, that is when* $12\exp(-\mu_1) = 1$, *the boundary length is of order* $\sqrt{N}$. *The exact statement is that the distribution of* $\frac{m_N}{\sqrt{N}}$ *converges in probability.*

Let us remove now the coordinate system from the boundary, that is we consider the class $\cup_m \mathcal{A}_{0,1}(N,m)$ of unrooted triangulations. The free energy remains the same. Only in the critical point the distribution of the length changes—stronger fluctuations appear.

Theorem 12. *In the critical point without coordinate system the boundary length is of order* N^α *for any* $0 < \alpha < \frac{1}{2}$. *The exact statement is that the distribution of* $\frac{\log m_N}{\log \sqrt{N}}$ *converges to the uniform distribution on the unit interval, that is* $P_{0,N}(\frac{\alpha}{2} \leq \frac{\log m_N}{\log \sqrt{N}} \leq \frac{\beta}{2}) \to \beta - \alpha$ *for all* $0 \leq \alpha < \beta \leq 1$.

7.3. SOME STOCHASTIC OPERATORS

One can construct probability distributions on $S_{\rho,k}$ generalizing the distribution of the previous section. For example, take $S_{0,2}$ and define a stochastic kernel on the set $\{2,3,\ldots\}$

$$S(m \to m_1) = \lim_{N\to\infty} S_N(m \to m_1), \quad S_N(m \to m_1) = \frac{Z_N(m,m_1)}{\sum_m Z_N(m,m_1)}$$

if the limit exists, where the conditional (the lengths of the boundaries are fixed to m and m_1) partition function $Z_N(m,m_1)$ is defined as

$$Z_N(m,m_1) = C(N,m,m_1)\exp(-\mu_1(m+m_1))$$

where $C(N,m,m_1)$ is the number of unrooted triangulations T of $S_{0,2}$ with N triangles and boundary lengths m and m_1. We study these operators in a forthcoming paper.

Similar stochastic kernels can be defined for systems with spins. Consider, for example, triangulations of $S_{\rho,k}$ with spin σ_v in the vertices taking values in the finite set E. Assume that a function $f : E \times E \to R$ is given, which defines a nearest-neighbor potential Φ so that $\Phi(\sigma_v, \sigma_{v'}) = f(\sigma_v, \sigma_{v'})$ if at least one of two vertices is inside the triangulation and by $\Phi(\sigma_v, \sigma_{v'}) = \frac{1}{2} f(\sigma_v, \sigma_{v'})$, if both are on the boundary. One could say equivalently that a local structure is given such that it knows whether the 1-neighborhood of v is a disk or a "half-disk". Let $L(E)$ is the linear space having elements of E as a basis.

Such operators can have relation to topological field theory (TFT) and scattering matrices for string models. In TFT, for example, one considers an abstract set of such operators

$$S_{\rho,k}(L(E)^{\otimes m_1} \otimes \cdots \otimes L(E)^{\otimes m_p} \to L(E)^{\otimes m_{p+1}} \otimes \cdots \otimes L(E)^{\otimes m_k})$$

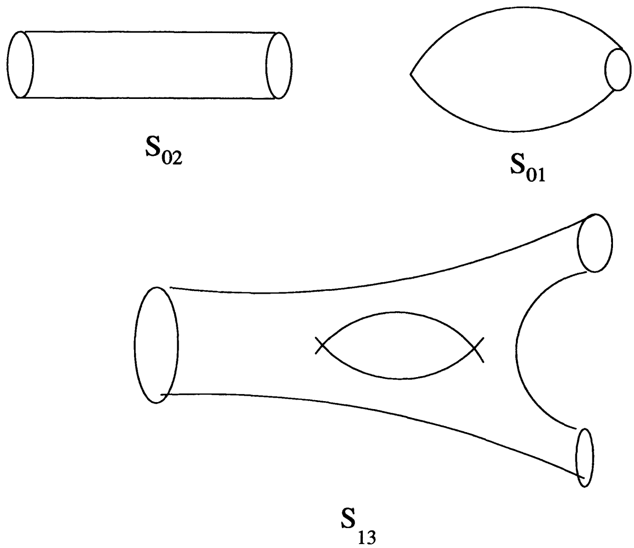


Figure 3. Surfaces with boundaries

satisfying some axioms concerning composition of these operators. These axioms can be given in different ways: that of Atiyah [35, 36] using functorial approach and the (not well-defined) approach using functional integrals. As far as I know now only simple examples of TFT, which are direct sums of the system without spin, see the last chapter of [18].

8. Lorentzian Models

Lorentzian models appeared in physical papers [37–39], where several approaches were suggested. Here we present rigorous formulations and results, see proofs in [40].

Above we defined slice-triangulations of the cylinder. They are easier to handle and complete control is possible.

Assume that $k_M = k$ and $k_N = l$ are fixed. Introduce Gibbs measure on the (countable) set $\mathcal{A}_{[M,N]}(k, l)$ of all such triangulations

$$\mu_{[M,N],k,l}(T) = Z^{-1}_{[M,N]} \exp(-\mu F(T)). \tag{8}$$

Define correlation functions for random variables $k_j, j \in [M+1, N-1]$, taking values from $\mathcal{N} = \{1, 2, \dots\}$, and for finite subsets $J \subset [M+1, N-1]$

$$\mu_{[M,N],k,l}(k_j = n_j, j \in J). \tag{9}$$

The behaviour of the system depends on to which region μ belongs: define subcritical, critical or supercritical region if $2\exp(-\mu) < 1$, $2\exp(-\mu) = 1$ and $2\exp(-\mu) > 1$ correspondingly.

In the subcritical region we get a limiting probability measure on $\mathcal{N}^Z$.

Theorem 13. *If* $2\exp(-\mu) < 1$ *then the limiting correlation functions*

$$\lim_{N\to\infty} \mu_{[-N,N],k,l}(k_j = k_j^0, j \in J)$$

exist for any finite subset $J \subset Z$ *and any vector* $(k_j^0, j \in J)$. *The measure defined by these correlation functions is a stationary ergodic Markov chain on* $\mathcal{N}$ *with the following stationary probabilities*

$$\pi(n) = \left(1 - (\lambda_2(s))^2\right)^2 n \left(\lambda_2(s)\right)^{2(n-1)},$$

where

$$\lambda_2(s) = \frac{1-\sqrt{1-s^2}}{s}, \quad s = 2\exp(-\mu).$$

Theorem 14. *In the subcritical case the asymptotics of the partition function is*

$$Z_{[-N,N]}(k,l) \sim \left(1 - (\lambda_2(s))^2\right)^2 \left(\lambda_2(s)\right)^{k+l-2-4N}.$$

Theorem 15. *In the critical case the asymptotics of the partition function is*

$$Z_{[-N,N]}(k,l) \sim \frac{1}{4}N^{-2}, \quad \sum_{l=1}^{\infty} Z_{[-N,N]}(k,l) \sim \frac{1}{2}N^{-1}.$$

For any J *and any* k_j^0 *the correlation functions*

$$\lim_{N\to\infty} \mu_{[-N,N],k,l}(k_j = k_j^0, j \in J) = 0.$$

In the critical case one can study the continuous limit $N \to \infty$. Take the interval $[0, N]$. We take the length of a horizontal edge equal to 1. At time αN, $\alpha < 1$, let $k_{[\alpha N]} = k_{[\alpha N]}(k,l)$ be the volume (the number of edges) in our one-dimensional Universe. Define the macrolength as $\frac{k_{[\alpha N]}}{N}$ and the macrotime as α.

Theorem 16. *For any positive* k, l

$$\lim_{N\to\infty} \frac{Ek_{[\alpha N]}}{N} = 2\alpha(1-\alpha), \tag{10}$$

and limit does not depend on k *and* l. *If* l *is not fixed then*

$$\lim_{N\to\infty} \frac{Ek'_{[\alpha N]}}{N} = \alpha(2-\alpha).$$

This shows that in our scaling limit the continuous two-dimensional Universe looks like the surface of a paraboloid of revolution, with two (or one, if l is not fixed) singular points. This can be interpreted as the expansion of this Universe.

Proposition 17. *In the supercritical case the finite volume partition function* $Z_{[0,N]}$ *exists only if*

$$\mu > \ln\left(2\cos\frac{\pi}{N+1}\right).$$

As $N \to \infty$ *this region, where the partition function exists, becomes empty.*

9. Gravity with matter fields

Here we shortly describe central ideas of the new cluster expansion techniques, which allows to treat combinatorial and probability problems for maps with spins. In statistical physics cluster expansions proved to a powerful tools, sometimes a unique tool allowing complete control over the problem. Disadvantages of this method is the necessity of a small or large parameter in the interaction, and also sufficient difficulty. We are interested here the stability problem of the critical exponent in the asymptotics of the partition function, however other questions, as correlation functions, can be treated as well. We consider only planar triangulations with matter fields with compact set of values in high temperature region. We shall see that for Gibbs families, to which maps belong, cluster expansions have many new features unknown for statistical physics and quantum field theory. The main new feature is that the empty space has an entropy, to take into account this entropy Tutte's method should be incorporated into the expansion.

We again consider the class $\mathcal{A}_{0,1}^{0}(N,m)$ and denote T^* the dual graph of the triangulation T, its vertices $v \in V(T^*)$ correspond to triangles of T, edges $l \in L(T^*)$—to pairs of adjacent triangles. All vertices of T^* have degree 3 except vertices corresponding to the triangles (there are not more than $m = |B(T)|$ of such triangles), incident to at least one boundary edge, where $B(T)$ is the set of boundary edges of T.

In each triangle of T, or in each vertex v of the dual graph T^*, there is a spin σ_v with values in the set E, this set is assumed finite for simplicity.

Partition function for the canonical ensemble (with fixed number $N \geq 0$ of triangles and fixed number $m \geq 2$ of boundary edges) is defined as

$$Z(N,m) = Z_{\beta}(N,m) = \sum_{T:|F(T)|=N,|B(T)|=m} Z(T)$$

where the partition function $Z(T)$ for a given triangulation $T \in \mathcal{A}_{0,1}^{0}(N,m)$ is

$$Z(T) = |E|^{-N} \sum_{\{\sigma_v : v \in V(T^*)\}} \exp(-\beta \sum_{<v,v'>} \Phi(\sigma_v, \sigma_{v'})), N = |F(T)| = |V(T^*)|$$

where $<v,v'>$ means a pair of nearest neighbor vertices (that is of adjacent triangles) $v, v' \in V(T^*)$, $\Phi(s,s')$ is a symmetric real function on $S \times S$, $\beta > 0$—inverse temperature.

To be concrete we shall consider the ensemble with boundary conditions empty on the internal boundary, that is there are no spins on the triangles of $F(T)$ adjacent to the boundary of the disk, thus no interaction with these triangles.

We prove that in some cases the partition function has canonical asymptotics. This means, there is a constant $c = c(\Phi, \beta)$ such that for fixed m, β, Φ

$$Z(N,m) \sim \phi(m, \Phi, \beta) N^{-\frac{5}{2}} c^N.$$

The critical exponent $\alpha = -\frac{5}{2}$ is also called canonical. For example, we have the following result.

Theorem 18. *Let*

$$k = \sum_{\sigma,\sigma'} [\exp(-\beta\Phi(\sigma,\sigma')) - 1] < 0.$$

Then for β sufficiently small $Z(N,m)$ has canonical asymptotics.

Theorem 19. *If $\Phi \leq 0$ is not identically constant, then for β sufficiently small the asymptotics is not canonical.*

Example: scaling transformation
A simple example is the constant nearest-neighbor interaction $\Psi_\mu(\sigma,\sigma') \equiv \mu$. For non-rooted triangulations the term Ψ gives an overall factor $\exp(-\beta\mu L^*) = \exp(-\frac{3}{2}\beta\mu N)$. Appending Ψ_μ to some interaction Φ results in a scaling transformation of the generating functions (see below). Otherwise speaking, appending such interaction changes only the constant $c(\beta)$ in the asymptotics, and does not change the canonical exponent.

9.1. CLUSTER EXPANSION TECHNIQUES

To present the main ideas we consider a simpler model—maps with random impurities. A map T with impurities is a map where some triangles are colored.

Note the distance between triangles in dynamical triangulation models is the distance between the corresponding vertices in the dual graph, that is the length (number of edges) of the shortest path between them in T^*. A set Δ of triangles is called connected if between each pair of triangles $t, s \in \Delta$ there is a path, belonging to Δ, in which any pair of consecutive triangles are on the distance not greater than $d = 1$.

For each set Δ define the external boundary $\partial_e\Delta$ as the set of triangles on distance 1 from Δ. A cluster (more exactly, a T-cluster for a given triangulation T) of colored triangles is a maximal connected subset of the closure $cl(V^*_{col}(T)) = V^*_{col}(T) \cup \partial_e(V^*_{col}(T))$ of the set $V^*_{col}(T)$ of colored triangles.

We define now the hierarchy of T-clusters for a given triangulation T. For any set $V \subset F(T)$ the complement $F(T) \setminus V$ consists of the two parts: exterior part $Ext(F(T) \setminus V)$, consisting of all triangles of $F(T) \setminus V$, which can be connected with the boundary by connected path, belonging to $F(T) \setminus V$, and the interior part $Int(F(T) \setminus V)$, containing all other triangles.

Let V be one of the T-clusters. Then the interior part of its complement $F(T) \setminus V$ consists of some number r of connected components $V_1, \ldots, V_r$.

For given T a set $V \subset F(T) = V(T^*)$ of triangles is called simple if it is connected and its interior part is empty. We say that T-cluster has level 1 if it is simple.

We define clusters of level $n > 1$ by induction: T-cluster V has level n if n is the minimal number such that in its interior part there are only clusters of level less than n. Thus the T-clusters form a forest (a set of connected trees), where clusters are vertices of this forest. Two vertices of the tree are connected by an edge if one of the corresponding T-clusters is in the interior part of the other one, and their levels differ by 1.

A random cluster model on maps assumes the following properties for any T:

1. Cluster V has a weight $k(V) > 0$, so that the partition function of the random cluster model is
$$Z^{(1)}(N,m) = \sum_{T:F(T)=N} \prod_{V\subset F(T)} k(V)$$
where the product is over all T-clusters.

2. The cluster estimate holds
$$k(V) \leq \beta^{|V|}.$$

3. Isomorphic clusters have equal weight.

4. All clusters are simple.

Nonempty T-cluster V is called complete if it contains all triangles of T. It is obviously simple and thus Σ consists only of this cluster. The complete T-cluster V is obviously unique and we define

$$k(T) = k(V).$$

Then the cluster generating function is defined as

$$W(x,y) = W_1(x,y) = \sum_{N=3}^{\infty} \sum_{m=2}^{\infty} W_{N,m}^{(1)} x^N y^m, W_{N,m}^{(1)} = \sum_{T:T\in\mathcal{A}_{01}^0(N,m)} K(T).$$

Then the equation for the generating functions

$$U_1(x,y) = \sum_{N=0}^{\infty} \sum_{m=2}^{\infty} Z^{(1)}(N,m) x^N y^m$$

is for $n = 1$

$$U_1(x,y) = U_1(x,y)xy^{-1} + U_1^2(x,y)xy^{-1} + y^2 + W_1(x,y) - xyS(x) \tag{11}$$

with

$$S_1(x) = \sum_{N=0}^{\infty} Z^{(1)}(N,2)x^N.$$

Proof. We have recurrent equations, similar to the case $\beta = 0$

$$Z^{(1)}(N,m) = Z^{(1)}(N-1,m+1) + \delta_{N,0}\delta_{m,2} + W_{N,m}^{(1)} + \sum_{N_1+N_2=N-1, m_1+m_2=m+1} Z^{(1)}(N_1,m_1)Z^{(1)}(N_2,m_2) \tag{12}$$

for $m \geq 2, N \geq 0$ and

$$Z^{(1)}(-1,m)) = Z^{(1)}(N,0)) = Z^{(1)}(N,1)) = 0.$$

Tutte's method is applicable to this equation, because W is analytic in a domain of radius which increases when β decreases.

If we do not assume the simplicity of clusters (property 4) then one uses an inductive procedure in n. For example, to get the partition function $Z^{(2)}(N,m)$, taking into account only maps having no clusters of level more than 2 but containing at least one cluster of level 2, we use already obtained level 1 generating function for $Z^{(1)}(N,m)$ as the function $W(x,y) = W_2(x,y)$ for the equation to find the generating function for $Z^{(2)}(N,m)$, etc. On each step the equations look like equation (11).

We omitted also the first step of the expansion, necessary to get the random cluster model from our high temperature model. On this step one uses a (standard, statistical physics type [42, 20]) cluster expansion for each fixed T, followed by the resummation. There are however some finer points, for example, the weights of clusters are not positive anymore.

See details in [41]. I hope that this method can be extended to more general situations: low temperature, unbounded spins etc. However, it demands additional efforts.

I also did not mention here two kind of physical results: methods of conformal field theory to get critical exponents and random matrix techniques to do calculations for Ising model on maps. There should be deep connections with cluster expansion. Isosystolic inequalities approach to genus independence of γ for models with matter fields see in [43].

References

1. Fernandez, R., Frolich, J. and Sokal A. (1992) *Random Walks, Critical Phenomena, and Triviality in Quantum Field Theory*, Springer–Verlag.
2. Gibbs, Ph. (1995) The Small Scale Structure of Space-Time: A Bibliographical Review, Preprint, `hep-th/9506171`.
3. Wallace, D. (2000) The Quantization of Gravity—an introduction, Preprint, `gr-qc/0004005`.
4. Rovelli C. (1998) Strings, loops and others: a critical survey of the present approaches to quantum gravity, Preprint, `gr-qc/9803024`.
5. Schnepps, L., Ed. (1994) *The Grothendieck theory of dessins d'enfants*, London Math. Soc. Lect. Note Series, v. 200, Cambridge Univ. Press.
6. Schnepps, L. and Lochak, P., Eds. (1997) *Geometric Galois Actions*, London Math. Soc. Lect. Note Series, v. 242, 243, Cambridge Univ. Press.
7. Edmonds, J. (1960) A combinatorial representation for polyhedral surfaces, *Notices Amer. Math. Soc.*, **7**, 646.
8. Kontsevich, M. (1992) Intersection Theory on the Moduli Spaces of Curves and the Matrix Airy Function, *Com. Math. Phys.*, **147**, no. 1, 1–23.
9. Liskovets, V. A. (1996) Some asymptotical estimates for planar maps. *Combinatorics, Probability and Computing*, **5**, 131–138.
10. Liu Yanpei, (1999) *Enumerative theory of Maps*, Kluwer.
11. Goulden, I. and Jackson D. (1983) *Combinatorial Enumeration*, Wiley.
12. Tutte, W. (1973) The Enumerative Theory of Planar Maps, in J. Srivastava et al. (Eds.), *A Survey of Combinatorial Theory*, North Holland.

13. Malyshev, V. (1999) Probability around the quantum gravity, *Russian Math. Reviews*, **54**, no. 4, 3–46.
14. Bender, E., Canfield, E. and Richmond, L. (1993) The asymptotic number of rooted maps on a surface. II. Enumeration by vertices and faces. *J. Comb. Theory* A, **63**, 318–329.
15. Gao Zhi-Cheng (1992) The asymptotic number of rooted 2-connected triangular maps on a surface, *J. Comb. Theory* B, **54**, no. 1, 102–112.
16. Frohlich, J. (1990) Regge calculus and discretized gravitational functional integral, Preprint, Zurich.
17. Ambjorn, J., Carfora, M. and Marzuoli, A. (1997) The geometry of dynamical triangulations, 1997; `http://xxx.lanl.gov/hep-th/9612069`.
18. Ambjorn, J., Durhuus, B. and Jonsson, T. (1997) *Quantum geometry*, Cambridge.
19. Walsh, T. and Lehman, A. (1972) Counting Rooted Maps by Genus, parts 1, 2, 3, *J. Comb. Theory* B, **13**, 192–218; (1972) **13**, 122–141; (1975) **18**, 222–259.
20. Malyshev, V. and Minlos, R.. (1989) *Gibbs Random Fields*, Kluwer.
21. Brezin, E., Itzykson, C., Parisi, G. and Zuber, J. (1978) Planar Diagrams, *Comm. Math. Phys.*, **59**, 35–47.
22. Kazakov, V., Staudacher, M. and Wynter, Th. (1996) Exact solution of discrete two-dimensional R^2 gravity, *Nuclear Physics* B **471**, 309–333.
23. Kazakov, V. A. (2000) Solvable Matrix Models. Talk delivered at Workshop "Matrix Models and Painleve equations", 1999, Berkeley; `hep-th/0003064`.
24. Zvonkin, A. (1997) Matrix Integrals and Map Enumeration: An Accessible Introduction, *Mathl. Comput. Modelling*, **26**, no. 8–10, 281–304.
25. Pastur, L. A. (1996) Spectral and Probabilistic Aspects of Random Matrix Models, in A. Boutet de Monvel and V.A. Marchenko (Eds.), Algebraic and Geometric methods in Mathematical Physics, Kluwer, 205–242.
26. Harer, J. and Zagier, D. (1986) The Euler characteristic of the moduli space of curves, *Invent. Math.*, **85**, 457–485.
27. Itzykson, C. and Zuber, J.-B. (1990) *Comm. Math. Phys.*, **134**, 197–207.
28. Okounkov, A. (2001) Gromov–Witten theory, Hurwitz numbers, and Matrix Models. I; `math.AG/0101147`.
29. Witten, E. (1991) Two dimensional gravity and intersection theory on moduli space, *Surveys in Diff. Geometry*, **1**, 243–310.
30. Dobrushin, R. (1968) The description of random field by means of conditional probabilities and conditions of its regularity, *Theory of Probability and Appl.*, **13**, 197–224.
31. Lanford, O. and Ruelle, D. (1969) Observables at infinity and states with short range correlations in statistical mechanics. *Comm. Math. Phys.*, **13**, 174–215.
32. Baez, J. (1996) Spin network states in gauge theory, *Adv. Math.*, **117**, 253–272.
33. Penrose, R. (1971) Angular momentum: an approach to combinatorial space-time, in T. Bastin (Ed.), *Quantum Theory and Beoynd*, Cambridge Univ. Press.
34. Malyshev, V. (2001) Gibbs and Quantum Discrete Spaces, *Russian Math. Reviews*, **56**, no. 5., 117–172.
35. Atiyah, M. (1991) *The geometry and physics of knots* Cambridge Univ. Press, Cambridge.
36. Dubrovin, B., Frenkel, E. and Previato, E. (1993) *Integrable Systems and Quantum Groups*, Springer.
37. Ambjorn, J., Jurkiewich, J. and Loll, R. (2000) Lorentzian and Euclidean quantum gravity—analytical and numerical results, `http://xxx.lanl.gov/hep-th/0001124`.
38. Ambjorn, J. and Loll, R. (1998) Non-perturbative Lorentzian Quantum Gravity, Causality and Topology Change, *Nucl. Phys.* B, **536**, 407.
39. Di Francesco, P., Guitter, E. and Kristjansen, C. (2000) Integrable 2D Lorentzian Gravity and Random Walks. *Nucl. Phys.* B, **567**, 515–553.
40. Malyshev, V., Yambartsev, A. and Zamyatin, A. (2001) Two-dimensional Lorentzian models, *Moscow Math. Journal*, **1**, no. 3, 439–456.
41. Malyshev, V. (in press) Dynamical triangulation models with matter fields: high temperature region, *Comm. Math. Phys.*
42. Glimm, J. and Jaffe, A. (1981) *Quantum Physics*.
43. Rivasseau, V. (1991) Isosystolic inequalities and the topological expansion for random surfaces and matrix models. *Comm. Math. Phys.*, **139**, 183–200.

THE COMBINATORICS OF ALTERNATING TANGLES: FROM THEORY TO COMPUTERIZED ENUMERATION

J. L. JACOBSEN and P. ZINN-JUSTIN
Laboratoire de Physique Théorique et Modèles Statistiques
Université Paris-Sud, Bâtiment 100
91405 Orsay Cedex, France

Abstract. We study the enumeration of alternating links and tangles, considered up to topological (flype) equivalences. A weight n is given to each connected component, and in particular the limit $n \to 0$ yields information about (alternating) knots. Using a finite renormalization scheme for an associated matrix model, we first reduce the task to that of enumerating planar tetravalent diagrams with two types of vertices (self-intersections and tangencies), where now the subtle issue of topological equivalences has been eliminated. The number of such diagrams with p vertices scales as 12^p for $p \to \infty$. We next show how to efficiently enumerate these diagrams (in time $\sim 2.7^p$) by using a transfer matrix method. We give results for various generating functions up to 21 crossings. We then comment on their large-order asymptotic behavior.

1. Introduction

The mathematical description of knots and related combinatorial objects is a subject with a history of several centuries. Giving an account of the progress made over the years is beyond the scope of the present paper, and we refer to [1] for an introduction to the existing literature (from the 19th century onwards). An old idea, first published by Ozanam [2] in 1694, is to treat a knot as a planar diagram with over- and underpasses. This trick, however simple it may appear, is essential to the present work, since it allows us to effectively reduce the dimension of the problem from three to two, and in particular to apply the powerful tools of matrix models in the planar limit. The classification of such diagrams with respect to topological equivalences calls for a particular geometrical framework (tagged *geometria situs* by Leibniz [3]), as was probably first pointed out by Vandermonde [4] in 1771.

For technical reasons, we shall here constrain our attention to the subset of *alternating knots* (and various generalizations such as links and tangles), which by definition possess a representation as a diagram in which over- and underpasses alternate. The first efforts in this domain date from the late 19th century through the pioneering works of Tait, Kirkman and Little. Tait originally thought that all knots were alternating, but Little made it evident (though no proof was known at that time) that non-alternating knots start appearing at order $p = 8$ intersections.

V. A. Malyshev and A. M. Vershik (eds.),
Asymptotic Combinatorics with Application to Mathematical Physics, 97–112.

It is now believed that alternating knots are asymptotically subdominant (the corresponding statement has been proven rigorously in the case of links [12]). However, the enumeration of alternating knots remains a very appealing mathematical problem, and to this date no explicit formula giving the number of knots at order p is known.

The subtle question of factoring out topological equivalences for alternating knots is facilitated by two remarkable theorems which were first conjectured by Tait, but only proved rigorously quite recently. First, alternating diagrams can be reduced to diagrams having minimal crossing number by iteratively eliminating irrelevant crossings through the move shown in Fig. 3a [22]. Second, two such *reduced* alternating diagrams are topologically equivalent if and only if they are connected by a sequence of simple transformations (see Fig. 4) known as *flypes* [10].

In the first part of the paper we shall show that counting alternating knots modulo the flype equivalence is equivalent to computing the generating function for a certain class of planar tetravalent diagrams that are illustrated in Fig. 2. These diagrams possess two types of vertices (self-intersections and tangencies), which are weighted separately. The proof is inspired by the renormalization procedure originally invented within the context of quantum field theory (QFT), and here applied to the random matrix model that generates the above-mentioned planar diagrams. In particular we show that factoring out topological equivalences is tantamount to including various counterterms in the effective action of the matrix model. Although we have found it appropriate to state the argument using the language of physics, it should be stressed that it eventually involves nothing but formal manipulations of generating functions, and thus is completely rigorous.

Unfortunately we have not found a way of analytically computing the generating function of the diagrams depicted in Fig. 2 for general n, although a number of cases ($n = 1$ [12], $n = 2$ [18], $n = \infty$ [25] and $n = -2$ [29]) can be solved exactly. In the second part of the paper we therefore proceed to describe how the integer coefficients of the generating function can be found, up to any desired order p, using a computer. The number of diagrams to be enumerated scales as 12^p for $p \gg 1$, but our algorithm does the counting in a most expeditious fashion, only using a time $\sim 2.7^p$. Once again, we have inspired ourselves from theoretical physics, reformulating the counting process as the action of a *transfer matrix*, which is an object normally used to describe the discrete time evolution of a quantum system. Mathematicians may think of the transfer matrix as a linear operator that acts on a set of basis states consisting of appropriately defined sub-diagrams.

From the description given below it will be evident that the applicability of our transfer matrix method extends far beyond the context of the present paper. Indeed, it can be generalized to the computation of generating functions for all sorts of planar diagrams, provided that the latter have at least one *external leg*. In particular, we are presently unable to generate results for *closed* knots (usually known in the mathematics literature simply as knots, as opposed to *tangles* which possess external legs).

We finally state our results, here given up to a maximal order of $p = 21$, and we analyze the asymptotic behavior of the various generating functions. The number of objects at order $p \gg 1$ is generally supposed to have the asymptotic behavior

$c\mu^p p^{-\alpha}$, where, in the language of statistical physics, μ is a critical temperature and α a critical exponent. Amazingly, the theory of two-dimensional quantum gravity provides a conjecture for α, based solely on the O(n) symmetry of the underlying matrix model, whereas the leading scaling exponent μ can only be accounted for in the exactly solvable cases. Comparing the conjecture for $\alpha(n)$ with our numerical results, we find good agreement in its expected range of validity ($0 \leq n \leq 2$).

2. The matrix model and its renormalization

The objects we want to consider are tangles with $2k$ "external legs", that is roughly speaking the data of k intervals embedded in a ball B and whose endpoints are given distinct points on the boundary ∂B, plus an arbitrary number of (unoriented) circles embedded in B, all intertwined, and considered up to orientation preserving homeomorphisms of B that reduce to the identity on ∂B. Tangles with 4 external legs will be simply called tangles. As mentioned in the introduction, we represent these objects using diagrams, that is regular projections on a plane, and restrict ourselves to alternating diagrams for which under- and over-passes alternate as one follows any connected component.

A natural framework is to introduce a *matrix model* which generates in its Feynman diagram expansion these link diagrams. We shall define it now.

2.1. DEFINITION OF THE $O(n)$ MATRIX MODEL

As in [19, 18, 29], we start with the following matrix integral over $N \times N$ hermitean matrices

$$Z^{(N)}(n,g) = \int \prod_{a=1}^{n} \mathrm{d}M_a \, \mathrm{e}^{N \operatorname{tr}\left(-\frac{1}{2}\sum_{a=1}^{n} M_a^2 + \frac{g}{4}\sum_{a,b=1}^{n} M_a M_b M_a M_b\right)} \tag{1}$$

where n is a positive integer.

Expanding in power series in g generates Feynman diagrams with double edges ("fat graphs") drawn in n colors in such a way that colors cross each other at the vertices. By taking the large N limit one selects the planar diagrams, which can be redrawn as alternating link diagrams, cf. Fig. 1. More precisely, the large N

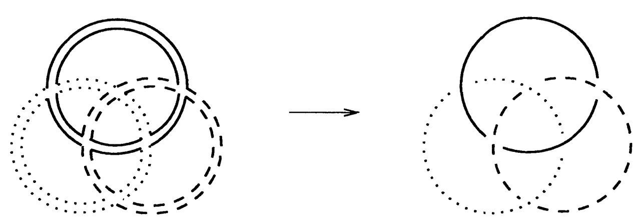


Figure 1. A planar Feynman diagram of (1) and the corresponding alternating link diagram

"free energy"

$$F(n,g) = \lim_{N\to\infty} \frac{\log Z^{(N)}(n,g)}{N^2} \tag{2}$$

is a double generating function of the number $f_{k;p}$ of alternating link diagrams with k connected components and p crossings (weighted by the inverse of their symmetry factor, and with mirror images identified):

$$F(n,g) = \sum_{k=1}^{\infty}\sum_{p=1}^{\infty} f_{k;p}\, n^k g^p \tag{3}$$

If one is interested in counting objects with a weight of 1, one cannot consider the free energy which corresponds to closed diagrams, but instead correlation functions of the model which generate diagrams with external legs, that is tangle diagrams. Typically, we shall be interested in the two-point function

$$G(n,g) \equiv \lim_{N\to\infty} \left\langle \frac{1}{N} \operatorname{tr} M_a^2 \right\rangle \tag{4}$$

where the measure on the matrices M_a is given by Eq. 1 and a is any fixed index, which generates tangle diagrams with two external legs; and the connected four-point functions

$$\Gamma_1(n,g) = \lim_{N\to\infty} \left\langle \frac{1}{N} \operatorname{tr}(M_a M_b)^2 \right\rangle \tag{5a}$$

$$\Gamma_2(n,g) = \lim_{N\to\infty} \left\langle \frac{1}{N} \operatorname{tr}(M_a^2 M_b^2) \right\rangle - G(n,g)^2 \tag{5b}$$

where a and b are two distinct indices, which generate tangle diagrams with four external legs of type 1 and 2 (see Fig. 2).

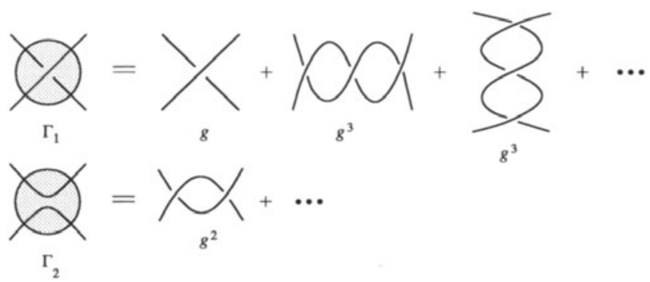


Figure 2. Tangles of types 1 and 2

2.2. RENORMALIZATION OF THE $O(n)$ MODEL

The model presented above counts tangle *diagrams*, but not tangles. There is a redundancy in the counting since to a given tangle/link corresponds many diagrams. In order to properly count tangles, this redundancy must be removed. In the case of alternating diagrams one can distinguish two steps. First one must find a way to select only *reduced* diagrams which contain no irrelevant crossings (Fig. 3a); such diagrams will have minimum number of crossings. It turns out to be convenient to introduce at this point a closely related notion: a link is said to be *prime* if it cannot be decomposed into two pieces in the way depicted on Fig. 3b. It is clear that at the level of diagrams, forbidding decompositions of the type of Fig. 3b automatically implies that the diagram is reduced; and we shall therefore restrict ourselves to prime links and tangles. There may still be several reduced diagrams

Figure 3. a) An irrelevant crossing. b) A non-prime link

corresponding to the same link: according to the flyping conjecture, proved in [10], two such diagrams are related by a finite sequence of flypes, see Fig. 4.

Figure 4. A flype

To summarize, there are two problems: a) the diagrams generated by applying Feynman rules are not necessarily reduced or prime; b) several reduced diagrams may correspond to the same knot due to the flyping equivalence. Following the study of [29], we note (cf. Figs. 3 and 4) that this "overcounting" is local in the diagrams in the sense that problem a) is related to the existence of sub-diagrams with 2 external legs, whereas problem b) is related to a certain class of sub-diagrams with 4 external legs. Clearly such graphs can be cancelled by the inclusion of appropriate *counterterms* in the action. We are therefore led to the conclusion that we must *renormalize* the quadratic and quartic interactions of (1). Renormalization theory tell us that we should include in the action from the start every term compatible with the symmetries of the model, since they will be generated dynamically by the renormalization. A key observation is that, while there is only one quadratic $O(n)$-invariant term, there are *two* quartic $O(n)$-invariant terms, which leads to a generalized model with 3 coupling constants in the action:

$$Z^{(N)}(n,t,g_1,g_2) = \int \prod_{a=1}^{n} \mathrm{d}M_a \, \mathrm{e}^{N\,\mathrm{tr}\left(-\frac{t}{2}\sum_{a=1}^{n} M_a^2 + \frac{g_1}{4}\sum_{a,b=1}^{n}(M_aM_b)^2 + \frac{g_2}{2}\sum_{a,b=1}^{n} M_a^2M_b^2\right)} \tag{6}$$

The Feynman rules of this model now allow loops of different colors to "avoid" each other, which one can imagine as tangencies (Fig. 5). We define again the

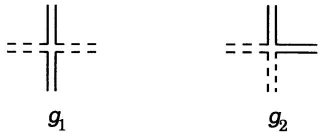


Figure 5. Vertices of the generalized $O(n)$ matrix model.

correlation functions $G(n,t,g_1,g_2)$ and $\Gamma_i(n,t,g_1,g_2)$ (Eqs. (4) and (5)), and want to extract from them the counting of colored alternating tangles with external legs.

The idea is to find the expressions of $t(g)$, $g_1(g)$ and $g_2(g)$ as a function of the renormalized coupling constant g, in such a way that the overcounting is suppressed and the correlation functions are generating series in g (and n) of the number of colored tangles.

The derivation of the renormalization equations was performed in [29]. Consider first the removal of irrelevant crossings and non-prime links. It is clear that one must remove all two-legged subdiagrams, that is impose

$$G(n,t,g_1,g_2) = 1 \tag{7}$$

This implicitly fixes $t(g)$.

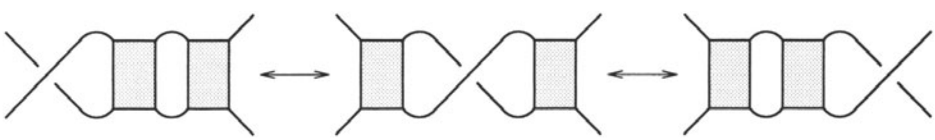


Figure 6. Breaking a flype into two elementary flypes.

Next, consider the flyping equivalence. A flype can be made of several "elementary" flypes (Fig. 6), an elementary flype being by definition one that cannot be decomposed any more in this way. In the terminology of QFT, these elementary flypes are one simple vertex connected by two edges to a non-trivial two-particle irreducible in the horizontal channel tangle diagram (H-2PI). We need to introduce auxiliary generating functions $H_1'(g)$, $H_2'(g)$ and $V_2'(g)$ for non-trivial H-2PI tangles of type 1, of type 2 and of type 2 rotated by $\pi/2$ respectively. By considering all possible insertions of elementary flypes as tangle sub-diagrams of a diagram, we find (Fig. 7) that the renormalization of g_1 and g_2 is given by:

$$g_1(g) = g(1 - 2H_2'(g)) \tag{8a}$$
$$g_2(g) = -g(H_1'(g) + V_2'(g)) \tag{8b}$$

All that is left is to find the expressions of the auxiliary generating functions in terms of known quantities. They are easily obtained by decomposing the four-point

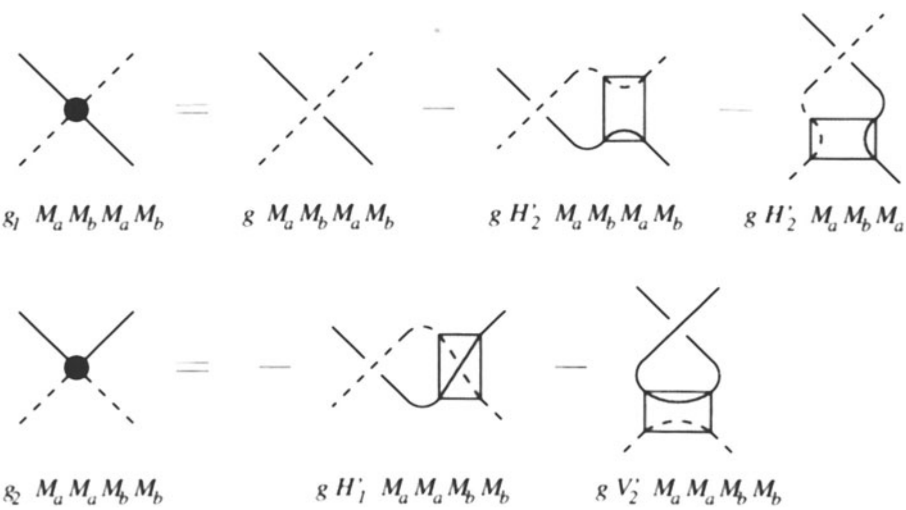


Figure 7. Counterterms needed to cancel flypes.

functions in the horizontal and vertical channels [18]:

$$H_2' \pm H_1' = 1 - \frac{1}{(1 \mp g)(1 + \Gamma_2 \pm \Gamma_1)} \tag{9a}$$

$$H_2' + nV_2' + H_1' = 1 - \frac{1}{(1 - g)(1 + (n+1)\Gamma_2 + \Gamma_1)} \tag{9b}$$

The three renormalization equations (7) and (8), supplemented by the expressions (9), are enough to fix $t(g)$, $g_1(g)$ and $g_2(g)$. Solving them gives access to the Γ_i, which are the generating series of the numbers of prime alternating tangles of type i. However, we can go further. By computing other correlation functions in the model and composing them with the solutions $t(g)$, $g_1(g)$, $g_2(g)$ of the equations above, one can extract the generating functions of the number of alternating tangles with an arbitrary number of external legs. An example will be given below.

3. An algorithm for the counting of decorated planar diagrams

Unfortunately the matrix model described in section 2.1 cannot be solved for a generic value of n (see however [17, 18, 29] for solutions of particular values). We therefore turn to the description of an algorithm that allows to enumerate the required diagrams to a given order p of crossings. This is one of several similar algorithms described in [24, 25]. The numbers of diagrams thus computed are the input needed for the combinatorial treatment of section 2.2.

The principle of the algorithm is to iterate a transfer matrix which "builds" the diagrams slice by slice: starting from an initial state consisting of all external legs (two in the case under consideration), the system is time evolved through the addition of p intersections, until an empty final state is obtained.

We shall first concentrate on the enumeration of (prime, alternating) tangle *diagrams*, which correspond to the "unrenormalized" model of section 2. Adding tangencies, which is needed to take into account the flyping equivalence, will be discussed in Section 3.2, since it is an elementary extension of the algorithm.

3.1. THE SINGLE-STEP ALGORITHM

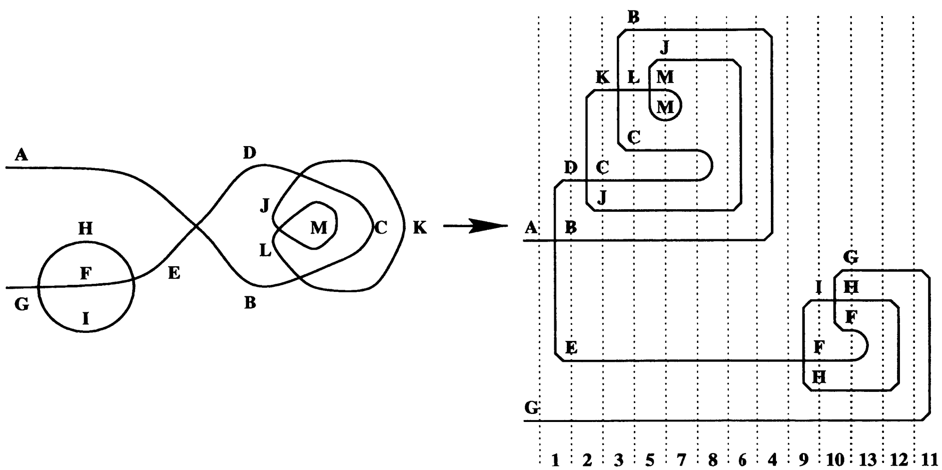


Figure 8. Working principle of the single-step algorithm.
a) A two-legged knot diagram with $p = 6$ intersections and $k = 3$ connected components. The edges are labelled from A to M.
b) The same diagram in the time-slice representation. For reasons of clarity, the time slices are not drawn in chronological order.

The algorithm we want to describe, called "single-step algorithm" in [25], is best explained by an example. Let us consider the tangle diagram shown in Fig. 8. Let us start from an initial state given by the two external legs (edges A and G). We want to "evolve" the various open lines by following them and adding the new crossings they meet.

Moving along either of the edges A or G, a new line segment (DE resp. HI) is encountered. The question then arises which of these to process first. We resolve this ambiguity by stipulating that ***in any given state, we evolve the line which at that instant is uppermost.*** At time $t = 1$, the edge A thus becomes B, and the new line segment DE is added. The edge D is now the new top line. Analogously, at the instants $t = 2$ and $t = 3$, the top line (D resp. K) crosses a new line segment, which is then added to the current state. We can formalize this by stating the transformation rule shown in Fig. 9.1.

At time $t = 3$, the new top line carries the label B, which was however already produced by the transformation acting at $t = 1$. We therefore proceed, at $t = 4$, to the identification of the two "copies" of B, joining them through an arch. This is an example of the general transformation rule shown in Fig. 9.2. The addition of an arch means that the lines intermediate between the two instances of B (at positions $p = 1$ and q on Fig. 9.2) can henceforth not communicate with the lines at the exterior of the arch. These "trapped lines" must therefore eventually evolve to the empty state (vacuum), independently of the rest of the diagram. Since the current top line must always be treated first, the evolution of a possible set of "trapped lines" must take place at a later time. Thus, on Fig. 8, we cannot always *draw* the time-slices in chronological order. The existence of a number of trapped lines is visualized on Fig. 9.2 by a *delimiter* (shown as a gray rectangle), which separated the remaining lines into two *blocks*. Lines in different blocks cannot communicate, and so the transformation 2) only applies when p and q belong to the same block.

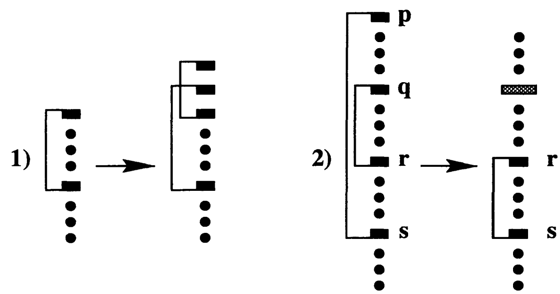


Figure 9. The two types of transformations in the single-step algorithm.
1) Addition of a new line segment.
2) Identification of the top line (at position p) with another line (at position q), accompanied with the creation of a new block.

The purpose of the transfer matrix is not only to count the total number of tangle diagrams, but to do so for any fixed number of connected components. In particular, when performing a type 2 transformation, we need to know whether the points p and q were already connected though an arbitrary number of edges at an *earlier* time. On Fig. 9 we have represented this information by a number of lines on the left, connecting the points at a given instant into pairs. It may thus happen that on Fig. 9.2, $r = p$ and $s = q$. In this case, the type 2 transformation marks the completion of one connected component in the tangle diagram.

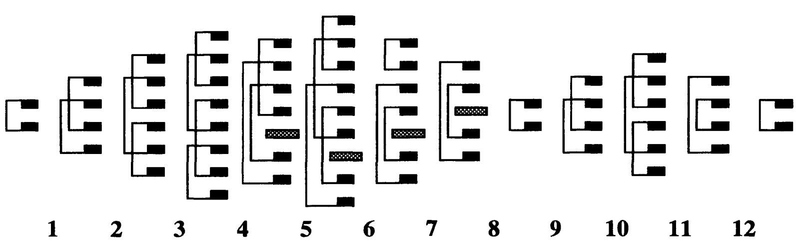


Figure 10. Intermediate states produced by applying the single-step algorithm to the tangle diagram shown in Fig. 8.

We are now ready to define the set of *states* on which the transfer matrix acts. A state is defined by an even number of points (represented on Fig. 9 as black rectangles), connected into pairs by means of the edges encountered at previous times. In addition, the points are divided into $\ell + 1$ blocks by means of $\ell \geq 0$ delimiters. On Fig. 10 we show the set of intermediate states corresponding to the time-slice representation of the tangle of Fig. 8.

Finally, we must define the transfer matrix T which counts *all* tangle diagrams with $\tilde{p}$ vertices and k connected components. Its entries T_{ab}, where a and b are two basis states of the kind just defined, are 0 unless b is a descendant of a. An allowed state b is a descendant of a if it can be obtained via a transformation of one of the two types shown on Fig. 9 (for an arbitrary even $q \geq 2$ belonging to the same block as $p = 1$), followed by an arbitrary number of simplifications. T_{ab} is then the sum over all transformations from a to b of the corresponding weight: 1 or n depending on whether one closes a connected component or not. The simplications are elimination of superfluous delimiters, and reduction of states to a "normal form" by using various symmetries of the transfer matrix [25].

3.2. TANGENCIES

Until now we have been discussing the enumeration of tangle diagrams in which every vertex represents a crossing. However, to account for the flype equivalence we need to enumerate more general diagrams with p_1 intersections and p_2 tangencies, as discussed in Section 2. Adding a tangency rather than an intersection is obtained by modifying the transformation if Fig. 9.1, so that the two points added at time $t + 1$ are both immediately above (resp. immediately below) the uppermost point at time t. any given time, specifies how many tangency transformations were used prior to that instant. The desired diagrams are then generated by sequences of p_1 crossing transformations (type 1), p_2 trangency transformations, and $p_1 + p_2 + 1$ transformations of type 2, so that no intermediate state is empty.

4. Numerical results and analysis

The implementation of the above algorithm provides an efficient numerical way to count the diagrams needed for our purposes. We present here some of the data obtained in [24, 25].

First, we give one example of the "raw data" obtained from the algorithm. It corresponds to counting a certain set of diagrams without taking into account topological equivalences: in the present case, *connected* two-legged alternating diagrams (called two-legged knot diagrams in [24]), see Tab. I. The 2PI diagrams are precisely the reduced diagrams of prime tangles. The importance of going to orders as high as possible comes from the issue of the asymptotic behavior which will be discussed below. Let us also give two examples of renormalized data. Tab. II shows the number of four-legged prime alternating tangles of the two possible types (Fig. 2) with arbitrary number of connected components, while Tab. III shows the number of six-legged tangles, which exist in five types (Fig. 11).

TABLE I. Total number G of two-legged diagrams, as well as the subset Σ_2 of 2PI diagrams.

p	G	Σ_2
0	1	0
1	2	2
2	8	0
3	42	2
4	260	4
5	1796	12
6	13396	60
7	105706	226
8	870772	1076
9	7420836	5156
10	65004584	24984
11	582521748	128548
12	5320936416	663040
13	49402687392	3514968
14	465189744448	18918792
15	4434492302426	103123906
16	42731740126228	569877652
17	415736458808868	3180066004
18	4079436831493480	17921451960
19	40338413922226212	101842206548
20	401652846850965808	583109887600
21	4024556509468827432	3361640932872

Finally, let us discuss the expected asymptotic behavior of these series. Let us consider an unrenormalized generating series for tangle diagrams, say $G(n,g) = \sum_{p=1}^{\infty} a_p(n) g^p$. Then it is expected that

$$a_p(n) \overset{p\to\infty}{\approx} \mathrm{e}^{\hat{s}(n)p} p^{-\alpha(n)} \tag{10}$$

The conjecture holds for any correlation function in the model; for the free energy $F(n,g) = \sum_{p=1}^{\infty} f_p(n) g^p$, that is the generating series for link diagrams (the symmetry factors being most likely irrelevant for the asymptotic behavior), it is

$$f_p(n) \overset{p\to\infty}{\approx} \mathrm{e}^{\hat{s}(n)p} p^{-\alpha(n)-1} \tag{11}$$

The bulk entropy $\hat{s}(n)$ can be easily extracted with a good accuracy from the numerical data. However it is *non universal*, which implies that it is not preserved by the renormalization of the model. On the other hand, the exponent $\alpha(n)$ is universal, and presumably preserved by the renormalization (unless an unusual phase transition takes place in the model). Since the raw data are "cleaner" than the renormalized data, they are better suited to the extraction of $\alpha(n)$. A conjecture made in [19], based in particular on the KPZ relation [28], is that for n analytically continued to $0 \le n \le 2$, one has

$$\alpha(n = -2\cos(\pi\nu)) = 1 + 1/\nu \qquad 1/2 \le \nu \le 1 \tag{12}$$

TABLE II. Table of the number of prime alternating tangles with 4 external legs.

	Γ_1							Γ_2						
$p \backslash k$	0	1	2	3	4	5	6	0	1	2	3	4	5	6
1	1							0						
2	0							1						
3	2							1						
4	2							3	1					
5	6	3						9	1					
6	30	2						21	11	1				
7	62	40	2					101	32	1				
8	382	106	2					346	153	24	1			
9	1338	548	83	2				1576	747	68	1			
10	6216	2968	194	2				7040	3162	562	43	1		
11	29656	11966	2160	124	2			31556	17188	2671	121	1		
12	131316	71422	9554	316	2			153916	80490	15295	1484	69	1	
13	669138	328376	58985	5189	184	2		724758	425381	87865	6991	194	1	
14	3156172	1796974	347038	22454	478	2		3610768	2176099	471620	52231	3280	103	1
15	16032652	9298054	1864884	193658	10428	260	2	17853814	11376072	2768255	308697	15431	290	1

TABLE III. Table of the number of prime alternating tangles with 6 external legs.

	Ξ_1				Ξ_2				Ξ_3					Ξ_4				Ξ_5			
$p \backslash k$	0	1	2	3	0	1	2	3	0	1	2	3	4	0	1	2	3	0	1	2	3
2	0				1				0					0				0			
3	2				0				2					0				0			
4	0				7				2					4				3			
5	18				6				16	2				8				9			
6	18				53	8			42	2				42	7			41	7		
7	156	24			154	6			171	44	2			156	14			168	21		
8	516	18			609	181	6		748	114	2			608	153	10		663	165	12	
9	2016	598	18		2956	422	6		2877	858	81	2		2850	586	20		3072	740	36	
10	10608	1428	18		11203	3498	318	6	14037	3752	213	2		11918	3445	364	13	13347	3966	438	18
11	40428	12318	1062	18	57664	15330	738	6	61028	19757	2511	131	2	57602	17558	1406	26	63393	20994	2040	54

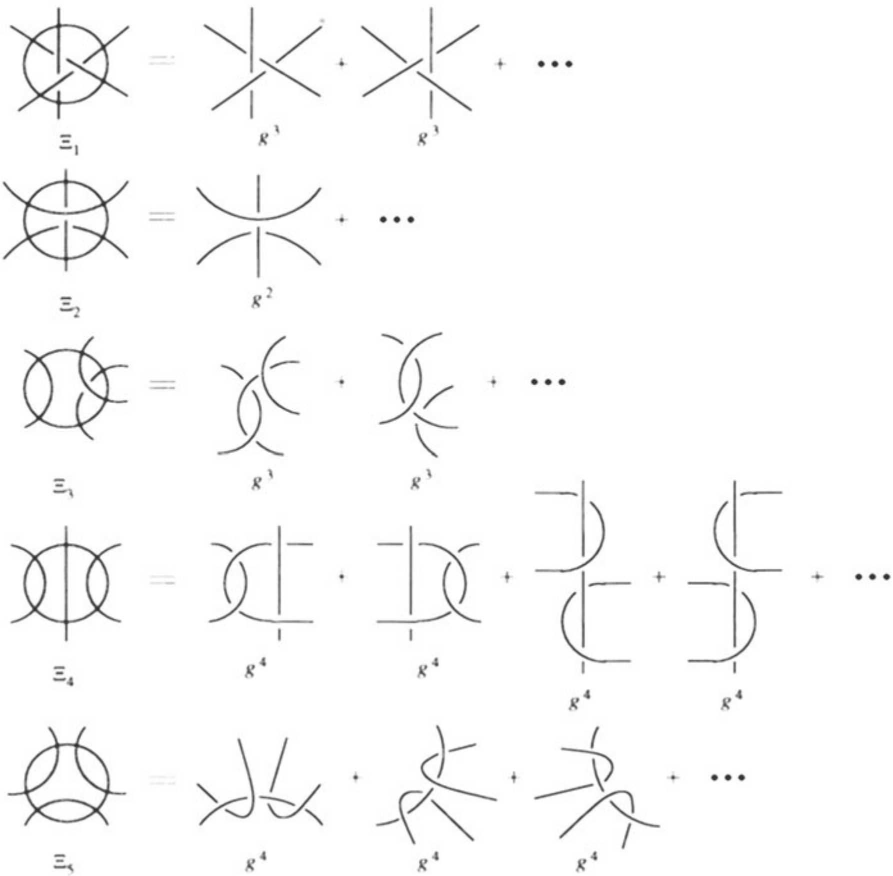


Figure 11. The five types of tangles with 6 external legs.

Using Tab. I one can try to evaluate $\alpha = \alpha(n = 0)$, the exponent of tangles with minimum number of components, which also gives the exponent of knots according to Eq. (11). The conjecture 12 yields $\alpha = 3$. Unfortunately, with the current data, it is difficult to estimate α without any knowledge of the subleading corrections. Indeed, a direct fit gives $\alpha \approx 2.76$ but with very weak accuracy. In [24], it was suggested to fit the data with $a_p = e^{\hat{s}p} p^{-\alpha}(a \log p + b + o(1))$, the result being in good agreement with the conjecture:

$$e^{\hat{s}} = 11.42 \pm 0.01 \quad \alpha = 2.98 \pm 0.04 \quad a = 0.03 \pm 0.01 \quad b = 0.1 \pm 0.03 \tag{13}$$

However there is little evidence for such a logarithmic correction, and the issue remains open. A more detailed discussion of $\alpha(n)$ for all n can be found in [25].

References

1. Hoste, J., Thistlethwaite, M. and Weeks, J. (1998) The First 1,701,936 Knots, *The Mathematical Intelligencer* **20**, 33–48.
2. Ozanam, J. (1694) *Récréations mathématiques et physiques* Paris, 2 vols.; (1778) Revised by Montucla, Paris, 4 vols.; (1803) English translation by Hutton, London, 4 vols. We here cite p. 222, vol. 4 of the 1725 edition. We do not know whether the idea was already manifest in the books by C. Bachet (1612 and 1624), and by C. Mydorge (1630).
3. von Leibniz, G. W. (1771), cited in *Histoire de l'Académie royale des Sciences de Paris* pour l'année, p. 55.
4. Vandermonde, A.-T. (1774) *Remarques sur les problèmes de situation*, Mém. de l'Ac. des Sc. de Paris pour l'année 1771, Paris, p. 566.
5. Arefeva, I. Ya. and Volovich, I. V. (1998) Knots and Matrix Models, *Infinite Dim. Anal. Quantum Prob.*, **1**; `hep-th/9706146`.
6. Brézin, E., Itzykson, C., Parisi G. and Zuber, J.-B. (1978) Planar Diagrams, *Commun. Math. Phys.*, **59**, 35–51.
7. Bessis, D., Itzykson, C. and Zuber, J.-B. (1980) Quantum Field Theory Techniques in Graphical Enumeration, *Adv. Appl. Math.*, **1**, 109–157.
8. Di Francesco, P., Ginsparg, P., and Zinn-Justin, J. (1995) 2D Gravity and Random Matrices, *Phys. Rep.*, **254**, 1–133.
9. 't Hooft, G. (1974) A Planar Diagram Theory for Strong Interactions, *Nucl. Phys.* B **72**, 461–473.
10. Menasco, W. W. and Thistlethwaite, M. B. (1991) The Tait Flyping Conjecture, *Bull. Amer. Math. Soc.*, **25** 403–412; (1993) The Classification of Alternating Links, *Ann. Math.*, **138**, 113–171.
11. Rolfsen, D. (1976) *Knots and Links*, Publish or Perish, Berkeley.
12. Sundberg, C. and Thistlethwaite, M. (1998) The rate of Growth of the Number of Prime Alternating Links and Tangles, *Pac. J. Math.*, **182**, 329–358.
13. Tutte, W. T. (1963) A Census of Planar Maps, *Can. J. Math.*, **15**, 249–271.
14. Zvonkin, A. (1997) Matrix Integrals and Map Enumeration: An Accessible Introduction, *Math. Comp. Modelling*, **26**, 281–304.
15. Kazakov, V. A. and Migdal, A. A. (1988) Recent progress in the theory of non-critical strings, *Nucl. Phys.* B, **311**, 171–190.
16. Kazakov, V. A. and Zinn-Justin, P. (1999) Two-Matrix Model with $ABAB$ Interaction, *Nucl. Phys.* B, **546**, 647; `hep-th/9808043`.
17. Zinn-Justin, P. and Zuber, J.-B. (in print) Matrix Integrals and the Counting of Tangles and Links, proceedings of the 11th International Conference on Formal Power Series and Algebraic Combinatorics, Barcelona June 1999; *Discrete Mathematics*; `math-ph/9904019`.
18. Zinn-Justin, P. and Zuber, J.-B. (2000) On the Counting of Colored Tangles, *Journal of Knot Theory and its Ramifications*, **9**, 1127–1141; `math-ph/0002020`.
19. Zinn-Justin, P. (2001) Some Matrix Integrals related to Knots and Links, in *Proceedings of the 1999 semester of the MSRI "Random Matrices and their Applications"*, MSRI Publications Vol. 40 (2001); `math-ph/9910010`.
20. Zinn-Justin, P. (2000) The Six-Vertex Model on Random Lattices, *Europhys. Lett.*, **50**, 15–21; `cond-mat/9909250`.
21. Kostov, I. (2000) Exact solution of the Six-Vertex Model on a Random Lattice, *Nucl. Phys.* B, **575**, 513–534; `hep-th/9911023`.
22. Kauffman, L. H. (1994) *Knots and Physics*, World Scientific Pub. Co.
23. Kauffman, L. H. (1998) Virtual Knot Theory; `math-gt/9811028`.
24. Jacobsen, J. L. and Zinn-Justin, P. (2001) A Transfer Matrix approach to the Enumeration of Knots; `math-ph/0102015`.
25. Jacobsen, J. L. and Zinn-Justin, P. (2001) A Transfer Matrix approach to the Enumeration of Colored Links; `math-ph/0104009`.
26. Eynard, B. and Kristjansen, C. (1996) More on the exact solution of the $O(n)$ model on a random lattice and an investigation of the case $|n| > 2$, *Nucl. Phys.* B, **466**, 463–487; `hep-th/9512052`.

27. Kostov, I. K. (1989) *Mod. Phys. Lett.* A, **4**, 217; Gaudin, M. and Kostov, I. K. (1989) *Phys. Lett.* B, **220**, 200; Kostov, I. K. and Staudacher, M. (1992) *Nucl. Phys.* B, **384**, 459.
28. Knizhnik, V. G., Polyakov, A. M. and Zamolodchikov, A. B. (1988) Fractal structure of 2D quantum gravity, *Mod. Phys. Lett.* A, **3**, 819–826; David, F. (1988) Conformal field theories coupled to 2D gravity in the conformal gauge, *Mod. Phys. Lett.* A, **3**, 1651–1656; Distler, J. and Kawai, H. (1989) Conformal field theory and 2D quantum gravity, *Nucl. Phys.* B, **321**, 509.
29. Zinn-Justin, P. (2001) The General $O(n)$ Quartic Matrix Model and its Application to counting Tangles and Links; `math-ph/0106005`.

INVARIANCE PRINCIPLES
FOR NON-UNIFORM RANDOM MAPPINGS AND TREES

DAVID ALDOUS (aldous@stat.berkeley.edu)
and JIM PITMAN* (pitman@stat.berkeley.edu)
Department of Statistics
University of California
367 Evans Hall # 3860
Berkeley, CA 94720-3860

Abstract. In the context of uniform random mappings of an n-element set to itself, Aldous and Pitman (1994) established a functional invariance principle, showing that many $n \to \infty$ limit distributions can be described as distributions of suitable functions of reflecting Brownian bridge. To study non-uniform cases, in this paper we formulate a ***sampling invariance principle*** in terms of iterates of a fixed number of random elements. We show that the sampling invariance principle implies many, but not all, of the distributional limits implied by the functional invariance principle. We give direct verifications of the sampling invariance principle in two successive generalizations of the uniform case, to p-mappings (where elements are mapped to i.i.d. non-uniform elements) and P-mappings (where elements are mapped according to a Markov matrix). We compare with parallel results in the simpler setting of random trees.

Key words: Brownian bridge, Brownian excursion, random mapping, random tree, weak convergence.

1. Introduction

A function M from the n-element set $[n] := \{1, 2, \ldots, n\}$ to itself, in this context called a *mapping*, induces a digraph on vertex-set $[n]$ whose edges are $(i, M(i))$, $i \in [n]$. From a random function M we get a random digraph, and the subject of *random mappings* concerns exact and asymptotic properties of such random graphs, most commonly under the *uniform model* where M is uniform on all n^n mappings [10, 30, 34, 39], but also under various non-uniform models. Saying M is uniform is equivalent to saying that $M(1), M(2), \ldots, M(n)$ are independent uniform on $[n]$, so a natural non-uniform model can be defined by requiring $M(1), M(2), \ldots, M(n)$ to be independent with some general probability distribution p on $[n]$. Such ***p-mappings*** are the subject of both classical and current research, reviewed briefly in Section 1.1. A more general model is to take $M(1), \ldots, M(n)$ independent on $[n]$ with probabilities $P(M(i) = j) = p_{ij}$ for

* Research supported in part by N.S.F. Grants DMS-9970901 and DMS-0071448

V. A. Malyshev and A. M. Vershik (eds.),
Asymptotic Combinatorics with Application to Mathematical Physics, 113–147.

some Markov matrix $P = (p_{ij})$; call this a *P-mapping.* One might guess that some $n \to \infty$ asymptotic results for uniform random mappings would extend to p-mappings or P-mappings, under appropriate conditions on the sequence $p^{(n)}$ or $P^{(n)}$, or to other models of non-uniform random mappings. The purpose of this paper is to set out a technical framework for studying such questions and specifying limit distributions.

Classical work on the uniform model focussed on specific statistics of mappings, such as component sizes and cycle lengths, which in the uniform case scale as order n and order $n^{1/2}$, and on joint distributions of such statistics [41]. As an extension, in the uniform model Aldous and Pitman [7] gave a Brownian bridge limit theorem which encompasses simultaneously limit distributions for many different statistics which scale as order n and order $n^{1/2}$. That *functional invariance principle*, based on coding mappings as walks, is reviewed in Section 4. A drawback is that the statement of the functional invariance principle is complicated and seemingly rather arbitrary.

In Section 3 we introduce a more direct method for studying random mappings, which we call the *sampling invariance principle.* This method is based on studying $n \to \infty$ asymptotics of the iterates of a fixed number of elements. Our central result is that a variety of asymptotic results extend from the uniform case to any model of non-uniform random mappings which satisfies the sampling invariance principle. Theorem 15 formulates this generally, and then we set down asymptotic distributional results more explicitly. Parallel results for random trees are simpler; we start by reviewing these in Section 2. The mathematical content of Theorem 15 (and the parallel Proposition 7 for trees) is that the sampling invariance principle can be reinterpreted as weak convergence of random functions, but using a weaker topology than is used in the functional invariance principle.

The essence of the sampling invariance principle is that, for fixed k, the union of the orbits of k elements converges (as $n \to \infty$, after rescaling) to a random "mapping with edge-lengths" $G[k]$ whose distribution has a simple explicit form. Such random graphs with edge-lengths, which also arise in the context of trees (Section 2.1), are perhaps of independent interest.

Our results are "abstract" in that instead of starting from a specific model, we are starting from the assumption that the model satisfies the sampling invariance principle, and exploring the non-obvious implications of that principle. For illustration we give two non-uniform random mapping models, which can be regarded as successive generalizations of the uniform model, and which can be shown directly to satisfy the sampling invariance principle. Theorem 25, proved in Section 6.1 as a simple consequence of Poisson approximation for the non-uniform birthday problem, establishes the sampling invariance principle for p-mappings under a natural condition. Then in Section 6.2 we indicate some conditions on P-mappings which imply the sampling invariance principle. It seems likely that other models of non-uniform random mappings could be shown to satisfy the sampling invariance principle, and this is a natural topic for future research.

1.1. p-MAPPINGS AND p-TREES

Though our main results are not tied to a particular model of random mappings, our approach was motivated by consideration of the p-mappings model. Older papers on that model focused on exact combinatorial properties related to Burtin's lemma [17, 32, 47] (see [45] for recent systematic discussion of combinatorial properties) and on asymptotics in the special case where all but one of the p-values are equal [12, 40, 48]. Asymptotics for general p were first considered explicitly only recently, when O'Cinneide and Pokrovskii [42] proved convergence of the rescaled height profile to an unspecified limit (our Corollary 19 reproves this and specifies the limit). However, asymptotics for the closely related p-trees model of random trees (Section 2.3(b)) have been extensively studied [18, 11, 9, 8]. It has recently become clear that an efficient way to study p-mapping asymptotics is to exploit the Joyal bijection between marked trees and mappings, which enables one to deduce asymptotics for p-mappings from already known asymptotics for p-trees. This approach, to be developed elsewhere [6], turns out to give stronger information about p-mappings than does the approach in this paper. But that method seems tied to the particular p-mappings model.

A final remark on our methodology. The sampling invariance principle for random trees (involving spanning subtrees of random vertices: Section 2.1) appears as part of the circle of ideas around the Brownian continuum random tree [4], but is somewhat overshadowed by the stronger and more succinct functional invariance principle for trees (Section 2.2). But in the context of mappings, studying the orbits of a fixed number of vertices (which is the essence of the sampling invariance principle) is very natural and often easy; and it is the statement of the functional invariance principle which is harder to interpret. That is why it seems worthwhile exploring the consequences of the sampling invariance principle.

2. Invariance principles for random trees

Consider the assertion

- For certain models of non-uniform random trees T on $[n]$, the $n \to \infty$ asymptotic distributions of many statistics should be the same as for the uniform model.

An *invariance principle* is a way of formalizing such an assertion. In this Section we give a slightly new perspective on some known results. We emphasize two apparently different ways of thinking about asymptotics (looking at spanning subtrees in Section 2.1; coding trees as walks in Section 2.2) and then describe carefully their relationship in Section 2.4. In Section 2.3 we recall hypotheses under which these invariance principles are true or conjectured.

2.1. SPANNING SUBTREES AND THE SAMPLING INVARIANCE PRINCIPLE

Consider a rooted tree T on vertex-set $[n]$. Take distinct vertices $\kappa_1, \ldots, \kappa_k$ and consider the spanning subtree on $\{\text{root}, \kappa_1, \ldots, \kappa_k\}$. Relabel vertex κ_i as i and

unlabel other vertices, while still marking the root. Picture the resulting tree as on the left of figure 1. Take each edge to have length 1; then delete unlabeled degree-2 vertices to obtain a "tree with edge-lengths" where the edge-lengths are integers. Call this tree $\mathrm{SPAN}(\kappa_1, \ldots, \kappa_k; T)$. For real $c > 0$ write $c \otimes \mathrm{SPAN}(\kappa_1, \ldots, \kappa_k; T)$ for the tree obtained by multiplying all edge-lengths by c. For asymptotics we anticipate getting a "tree with edge-lengths" $\mathbf{t}$ as pictured on the right in figure 1. Such a tree $\mathbf{t}$ has the following properties.
(i) There is a degree-1 root, and k other leaves labeled by $[k]$.
(ii) It is a binary tree, with unlabeled degree-3 branchpoints.
(iii) Each edge has a strictly positive real length.

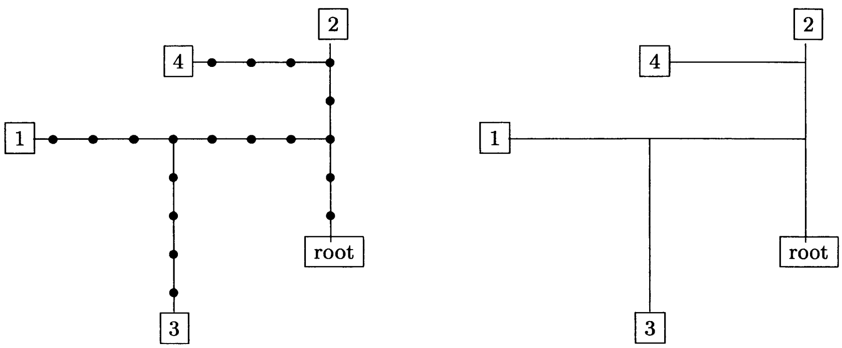


Figure 1. Instances of $\mathrm{SPAN}(\kappa_1, \ldots, \kappa_4; T)$ and of a $\mathbf{t} \in \mathbf{T}[4]$.

Write $\mathbf{T}[k]$ for the set of such graphs. A tree $\mathbf{t} \in \mathbf{T}[k]$ can be specified by its *shape* (pedantically, the shape is the equivalence class of isomorphic leaf-labeled rooted trees-without-edge-lengths) and by its $2k-1$ edge-lengths. Inductively, the number of shapes equals $(2k-3)!! := (2k-3)(2k-5)\cdots 3 \cdot 1$ because the k'th leaf can be attached at $2k-3$ different places. (Make the convention $(-1)!! = 1$.) Following [4] Section 4.3, define a distribution for a random graph $T[k]$ with edge lengths $\mathcal{L}_1, \ldots, \mathcal{L}_{2k-1}$ as follows. For each of the $(2k-3)!!$ possible shapes t,

$$P(\mathrm{shape}(T[k]) = t, \mathcal{L}_1 \in d\ell_1, \ \ldots \ , \mathcal{L}_{2k-1} \in d\ell_{2k-1}) \\ = \ f_{2k-1}(\ell_1, \ldots, \ell_{2k-1})\, d\ell_1 \ldots d\ell_{2k-1} \qquad (1)$$

where

$$f_{2k-1}(\ell_1, \ldots, \ell_{2k-1}) := \left(\sum_{j=1}^{2k-1} \ell_j \right) \exp\left(-\tfrac{1}{2} \left(\sum_{j=1}^{2k-1} \ell_j \right)^2 \right), \qquad (2)$$

and where we adopt some arbitrary convention for ordering the edges associated with each possible shape. The convention does not matter because $f_{2k-1}(\cdot)$ is symmetric. Definition (1)–(2) implies that $\mathrm{shape}(T[k])$ is uniform on the $(2k-3)!!$

possible shapes and that the edge-lengths are independent of the shape. Saying that (1) - (2) defines a *probability* density on $\mathbf{T}[k]$ is saying

$$(2k-3)!! \int \cdots \int f_{2k-1}(\ell_1, \ldots, \ell_{2k-1})\, d\ell_1 \ldots d\ell_{2k-1} = 1;$$

check by rewriting the integral as $\frac{1}{(2k-2)!} \int_0^\infty s^{2k-2} \exp(-s^2/2) ds$. (There is an alternate interpretation of the distribution of $T[k]$ using a *line-breaking construction*: [4] Lemma 21.)

Note there is a natural notion of convergence in $\mathbf{T}[k]$: $\mathbf{t}^{(n)} \to \mathbf{t}$ if shape$(\mathbf{t}^{(n)}) =$ shape$(\mathbf{t})$ ultimately and the edge-lengths converge. This convergence can be metrized. So convergence in distribution (written $\xrightarrow{d}$) of random trees with edge lengths means weak convergence with respect to the underlying metric.

By a *weighted tree* on vertices $[n]$ we mean a tree together with a probability distribution $q = (q_i)$ on the vertex-set $[n]$. Picture the q_i as deterministic "weights" on vertices, which for a random tree do not depend on the realization of the tree. We may consider the uniform distribution as a "default" choice of q, but allowing more generality may be useful. In the context of $n \to \infty$ asymptotics for such *random weighted trees* we always assume, without further mention, that the distributions $q = q^{(n)}$ satisfy

$$\max_i q_i \to 0 \text{ as } n \to \infty. \tag{3}$$

Definition 1. A model for random weighted trees (T, q) on $[n]$ satisfies the sampling invariance principle with scaling constants $c = c^{(n)}$ if, as $n \to \infty$,

$$c \otimes \mathrm{SPAN}(\kappa_1, \ldots, \kappa_k; T) \xrightarrow{d} T[k], \quad k \geq 1 \tag{4}$$

where $\{\kappa_1, \ldots, \kappa_k\}$ are independent of each other and of T with distribution $q^{(n)}$.

The word "sampling" is intended to convey the idea of "spanning tree on randomly sampled vertices". Note that for finite n the tree SPAN$(\kappa_1, \ldots, \kappa_k; T)$ might not be in $\mathbf{T}[k]$, for instance if two of the κ_i coincide. However, one can make sense of convergence in distribution of random objects $\mathcal{X}_n$ even if the objects are well-defined only on events A_n with $P(A_n) \to 1$ (for instance by appending a "fictitious state"), and we adopt this view throughout the paper. Note that hypothesis (3) is exactly what is needed to ensure $\lim_n P(\kappa_1, \ldots, \kappa_k \text{ distinct }) = 1$.

Examples of models where Definition 1 is satisfied will be given shortly (section 2.3), after we recall an alternative notion of *invariance principle*.

2.2. CODING TREES AS WALKS: THE FUNCTIONAL INVARIANCE PRINCIPLE

In a rooted *ordered* tree, the children of each vertex are ordered as first, second, third .... For a rooted ordered tree on n vertices there is a *depth-first* relabeling of the vertices as $\{0, 1, \ldots, n-1\}$ defined as follows. Label the root as 0. Having labeled some vertex as i, give label $i+1$ to
(i) the first child of i, if any; else

(ii) the first not-yet-labeled child of parent(i), if any; else
(iii) the first not-yet-labeled child of parent(parent(i)); and so on.
After thus relabeling vertices in depth-first order, define a walk

$$w(i) = \text{height}(i), \ 0 \le i \le n-1$$

where the *height* of vertex i is its distance to the root. Note that the walk determines the tree, because vertex i is the next child of vertex $\max\{i' < i : w(i') = w(i) - 1\}$.

Given a weight function q on the vertices, we can define a rescaled walk $(\widetilde{w}(t), 0 \le t \le 1)$ by

$$\widetilde{w}(t) = w(i) \text{ for } q_0 + q_1 + \cdots + q_{i-1} \le t < q_0 + q_1 + \cdots + q_i. \tag{5}$$

So $\widetilde{w}$ is an element of the usual space $D[0,1]$ of right-continuous functions with left limits [16]. The functional invariance principle (which we are about to define) relates to ordered trees, whereas the sampling invariance principle was defined for unordered trees. An unordered tree T can be made into an ordered tree by putting the children of each vertex into uniform random order. The resulting depth-first labeling of vertices of T by $\{0, 1, \ldots, n-1\}$ will be called the *randomized depth-first ordering* of T.

Write $B^{\text{exc}} = (B_t^{\text{exc}}, 0 \le t \le 1)$ for *standard Brownian excursion* [13, 15]. Consider the following property for a model of random weighted trees (T, q) on $[n]$ and constants $c = c^{(n)}$.

Property 2. The rescaled walk $\widetilde{W}(t)$ defined by (5) based on the randomized depth-first ordering satisfies

$$(c\widetilde{W}(t), \ 0 \le t \le 1) \ \xrightarrow{d} \ (2B_t^{\text{exc}}, \ 0 \le t \le 1) \text{ as } n \to \infty$$

with respect to a specified metric on $D[0,1]$.

Definition 3. A model for random weighted trees (T, q) on $[n]$ satisfies the functional invariance principle with scaling constants $c = c^{(n)}$ if Property 2 holds for the usual (Skorokhod J_1) metric on $D[0,1]$.

We shall shortly show (Proposition 7) that the functional invariance principle implies the sampling invariance principle. But before continuing the theoretical development to show this, let us discuss some specific models of random trees where these invariance principles are known or conjectured.

2.3. MODELS OF NON-UNIFORM RANDOM TREES

Model (a). The uniform random rooted tree on $[n]$ is a special case (where ξ has Poisson distribution) of the CBP(n) *model* (here CBP stands for *conditioned branching process*). This is the Galton–Watson branching process, with some offspring distribution ξ satisfying

$$E\xi = 1; \ 0 < \sigma^2 := \text{var } \xi < \infty ,$$

conditioned to have total size equal to n.

Theorem 4. *The CBP(n) model (with uniform weights q) satisfies the functional invariance principle with scaling constants $\sigma n^{-1/2}$.*

This was proved in [4] with a slightly different walk coding, though it is easy to deduce the stated form. See [38] for a more direct proof and further references; and see [21, 22] for cases with infinite variance. Because of the special combinatorial structure of the CBP(n) model one can obtain more refined results, for instance the local limit theorem corresponding to Corollary 10 [20].

Model (b). Given a probability distribution p on $[n]$, a *p-tree* [18, 45] is a random rooted tree T on vertex-set $[n]$ whose distribution is defined by

$$P(T = \mathbf{t}) = \prod_{i \in [n]} p_i^{C_i \mathbf{t}} \tag{6}$$

where $C_i\mathbf{t}$ denotes the number of children of vertex i in $\mathbf{t}$. For a probability distribution p on $[n]$ write

$$c_p := \sqrt{\sum_i p_i^2}.$$

For a sequence $p^{(n)}$ of probability distributions on $[n]$, introduce the condition

$$\frac{\max_i p_i}{c_p} \to 0 \text{ as } n \to \infty. \tag{7}$$

For reasons to be touched upon in Section 7(b), one can view condition (7) for p-trees as closely analogous to Lindeberg's condition in the central limit theorem.

Theorem 5. *Under assumption* (7), *a sequence of p-trees with weight functions $q = p$ satisfies the sampling invariance principle with scaling constants c_p.*

This is obtained by combining Corollary 9 of Camarri and Pitman [18], who specified the limit in terms of a line-breaking construction, with Lemma 21 of Aldous [4], which obtains the formula stated for $T[k]$ from the line-breaking construction. One can deduce from [18] that (7) is also *necessary* for the sampling invariance principle. But assumption (7) is not enough to imply the functional invariance principle for p-trees (the example for p-mappings in Section 6.1 can be adapted to p-trees). We do not know the precise necessary and sufficient condition.

In the analogous result for random mappings (Theorem 25) the weight functions can be arbitrary. This is surely also true in Theorem 5, though we have not attempted a proof.

Model (c). Attach i.i.d. costs to the edges of the complete graph on $[n]$, and then let T be the minimum-cost spanning tree, rooted at 1 say. Frieze [25] studied the total cost of T, and Aldous [3] studied some asymptotic distributions associated with the tree T itself, but the following conjecture remains open.

Conjecture 6. *T satisfies the sampling invariance principle with uniform weights and with scaling constants $n^{-1/2}$.*

Model (d). Given an irreducible Markov transition matrix $P = (p_{ij})$ on $[n]$, one can define a *P-tree* via

$$P(T = \mathbf{t}) \propto \prod_{(i,j)\in\mathbf{t}} p_{ij}$$

where we are regarding $\mathbf{t}$ as a set of edges directed toward the root. Such trees arise as part of a circle of ideas around the *Markov chain tree theorem* [37]. It seems plausible that one can find conditions on P that imply invariance principles, but this setting has apparently not been studied.

2.4. RELATING THE SAMPLING AND FUNCTIONAL INVARIANCE PRINCIPLES

The definitions of the sampling invariance principle and the functional invariance principle look quite different, but we now give a result revealing a close relationship. On the function space $D[0,1]$, the usual Skorokhod metric is stronger than the L_0 metric

$$||f_1 - f_2||_0 := \int_0^1 \min(1, |f_1(t) - f_2(t)|)\ dt.$$

It is easy to check there is yet another metric on $D[0,1]$, which we call the "$*$-metric", intermediate between the L_0 and the Skorokhod metrics, with the property

for $f_n, f \in D[0,1]$ with f continuous, $f_n \to^* f$ iff

$$\begin{aligned} ||f_n - f||_0 &\to 0 \text{ and} \\ \inf_{a\le t\le b} f_n(t) &\to \inf_{a\le t\le b} f(t) \text{ for all intervals } [a,b] \text{ of positive length.} \end{aligned} \tag{8}$$

Proposition 7. *A model for random weighted trees (T,q) on $[n]$ satisfies the sampling invariance principle with scaling constants $c = c^{(n)}$ if and only if Property 2 holds for the $*$-metric.*

Definition 3 and this alternate characterization of sampling invariance make it clear that the functional invariance principle does imply the sampling invariance principle. Proposition 7 is new, though the key underlying conceptual fact (Lemma 8) is known; let us explain this fact first.

Given $f : [0,1] \to [0,\infty)$ and distinct $\{u_1, u_2, \ldots, u_k\} \subset (0,1)$, one can define a tree-with-edge-lengths $\mathbf{t}(u_1, \ldots, u_k; f)$ by specifying
(i) leaf i is at height (distance from root) $f(u_i)$;
(ii) if $u_i < u_j$ then the branchpoint between the paths from the root to i and to j is at height $\min_{u_i\le t\le u_j} f(t)$.

We can make $\mathbf{t}(u_1, \ldots, u_k; f)$ into an ordered tree by giving the leaves i the order inherited from the natural order of the u_i in $(0,1)$.

Now let $U_1, \ldots, U_k$ be i.i.d. uniform $(0,1)$ random variables, independent of B^{exc}. Recall $T[k]$ from Section 2.1; we can make $T[k]$ into an ordered tree by putting, independently at each branchpoint, the two edges leading away from the root into random order.

Lemma 8 ([4] Corollary 22). $\mathbf{t}(U_1, \ldots, U_k; 2B^{\mathrm{exc}}) \stackrel{d}{=} T[k]$.

Proof of Proposition 7. Now for a tree T and a weight function q, make the tree ordered as in Section 2.2, and write

$$\kappa(u) := \min\{i : q_0 + q_1 + \ldots + q_i \geq u\}. \tag{9}$$

So $\kappa(\cdot)$ depends on the tree via the ordering of vertices. Take $(U_1, \ldots, U_k)$ independent of the random trees T. So $\kappa(U_1), \ldots, \kappa(U_k)$ are independent random vertices chosen from distribution q. Therefore $\mathrm{SPAN}(\kappa_1, \ldots, \kappa_k; T) \stackrel{d}{=} \mathrm{SPAN}(\kappa(U_1), \ldots, \kappa(U_k); T)$. But from their definitions, $\mathrm{SPAN}(\kappa(U_1), \ldots, \kappa(U_k); T)$ and $\mathbf{t}(U_1, \ldots, U_k; \widetilde{W})$ are almost the same; the difference is that the heights of branchpoints in the latter are exactly 1 less than their heights in the former, and because $c \to 0$ this difference vanishes asymptotically. Thus by Lemma 8 the sampling invariance principle is equivalent to: for each k,

$$\mathbf{t}(U_1, \ldots, U_k; c\widetilde{W}) \xrightarrow{d} \mathbf{t}(U_1, \ldots, U_k; 2B^{\mathrm{exc}}). \tag{10}$$

Write $0 < V_1 < V_2 < \ldots < V_k < 1$ for the order statistics of $(U_1, \ldots, U_k)$. Consider the assertion

$$\begin{aligned} &c\left(\widetilde{W}_n(V_1), \inf_{V_1 \leq t \leq V_2} \widetilde{W}_n(t), \widetilde{W}_n(V_2), \inf_{V_2 \leq t \leq V_3} \widetilde{W}_n(t), \ldots, \widetilde{W}_n(V_k)\right) \\ &\xrightarrow{d} (2B^{\mathrm{exc}}(V_1), \inf_{V_1 \leq t \leq V_2} 2B^{\mathrm{exc}}(t), 2B^{\mathrm{exc}}(V_2), \inf_{V_2 \leq t \leq V_3} 2B^{\mathrm{exc}}(t), \ldots, 2B^{\mathrm{exc}}(V_k)). \end{aligned} \tag{11}$$

Using the fact that (V_i) is independent of the permutation associating the (U_i) with the (V_i), and the random ordering of branches within trees, we see that (11) implies (10). Conversely, since in (10) the trees $\mathbf{t}(\cdot)$ are ordered, from a realization of $\mathbf{t}(\cdot)$ we can determine the permutation associating the (U_i) with the (V_i), and it follows that (10) implies (11). □

We have now reformulated the sampling invariance principle as (11). Proposition 7 is a consequence of this reformulation, together with the following reformulation of $*$-convergence. (Note that in our setting the first and last components of the vectors below are automatically tending to zero).

Lemma 9. *Let $X_n(t)$ and $X(t)$ be processes in $D[0,1]$, and suppose $X(\cdot)$ has continuous paths. Then $X_n \xrightarrow{d} X$ with respect to the $*$-metric if and only if for each k*

$$\begin{aligned} &\Big(\inf_{0 \leq t \leq V_1} X_n(t), X_n(V_1), \inf_{V_1 \leq t \leq V_2} X_n(t), X_n(V_2), \\ &\qquad \inf_{V_2 \leq t \leq V_3} X_n(t), \ldots, X_n(V_k), \inf_{V_k \leq t \leq 1} X_n(t)\Big) \\ &\xrightarrow{d} \Big(\inf_{0 \leq t \leq V_1} X(t), X(V_1), \inf_{V_1 \leq t \leq V_2} X(t), X(V_2), \\ &\qquad \inf_{V_2 \leq t \leq V_3} X(t), \ldots, X(V_k), \inf_{V_k \leq t \leq 1} X(t)\Big). \end{aligned} \tag{12}$$

Proof. The "only if" is immediate. For "if", the key fact (proved as Proposition 29 in the Appendix) is that convergence in distribution with respect to the L_0 metric is equivalent to

$$(X_n(V_1), X_n(V_2), \ldots, X_n(V_k)) \xrightarrow{d} (X(V_1), X(V_2), \ldots, X(V_k)), \quad k \geq 1. \tag{13}$$

So in particular, assumption (12) implies $X_n \xrightarrow{d} X$ in L_0. Now let us be more precise about notation and write $(V_{k,i}, 1 \leq i \leq k)$ for the order statistics of $(U_i, 1 \leq i \leq k)$. Fix $a < b$ and consider $i(k) \sim ak$ and $j(k) \sim bk$ as $k \to \infty$. For fixed k,

$$\inf_{V_{k,i(k)} \leq t \leq V_{k,j(k)}} X_n(t) \xrightarrow{d} \inf_{V_{k,i(k)} \leq t \leq V_{k,j(k)}} X(t)$$

and so by taking $k = k_n \to \infty$ sufficiently slowly

$$\inf_{V_{k_n,i(k_n)} \leq t \leq V_{k_n,j(k_n)}} X_n(t) \xrightarrow{d} \inf_{a \leq t \leq b} X(t),$$

using path-continuity of X. We could choose $i(k)$ such that $V_{k,i(k)} > a$ ultimately, or such that $V_{k,i(k)} < a$ ultimately, and analogously for $j(k)$; so

$$\inf_{a \leq t \leq b} X_n(t) \xrightarrow{d} \inf_{a \leq t \leq b} X(t). \tag{14}$$

Using the Skorokhod representation theorem ([24] Theorem 3.1.8) we may assume $X_n \xrightarrow{a.s.} X$ in L_0, and then convergence in (14) must be a.s. also. Since $a < b$ is an arbitrary interval, it follows that $X_n \xrightarrow{a.s.} X$ in the $*$-metric. □

Remark. The construction of a tree-with-edge-lengths $\mathbf{t}(u_1, \ldots, u_k; 2B^{\text{exc}})$ with an arbitrary finite number k of leaves labeled by $\{u_1, \ldots, u_k\} \subset [0,1]$ extends to the construction [4] of the Brownian *continuum random tree* (CRT) whose leaves are indexed by almost all $u \in [0,1]$, but we do not emphasize that formalism in this paper.

2.5. DISTRIBUTIONAL LIMITS IMPLIED BY INVARIANCE PRINCIPLES

Here we briefly recall some instances of what can be deduced from each type of invariance. Write height(i, T) for the height of vertex i in tree T. In the context of an invariance principle, define the *rescaled cumulative height profile* by

$$H_n(u) := \sum_{i=1}^{n} q_i^{(n)} 1_{(\text{height}(i,T) \leq u/c_n)}.$$

Regard H_n as a random element of the space $\mathcal{D}$ of distribution functions, where $\mathcal{D}$ is equipped with the topology of "convergence in distribution". Note that the map $D[0,1] \to \mathcal{D}$ defined by

$$f \to H_f: \quad \text{where } H_f(u) := \int_0^1 1_{(f(t) \leq u)}\, dt$$

is L_0-continuous at almost all realizations of B^{exc}. So we can use the continuous mapping theorem to get

Corollary 10. *Consider a model for random weighted trees (T, q) on $[n]$ which satisfies the sampling invariance principle with scaling constants c. Then*

$$H_n \xrightarrow{d} H$$

as random elements of $\mathcal{D}$, where

$$H(u) := \int_0^1 1_{(2B_t^{\text{exc}} \leq u)} \, dt; \quad 0 \leq u < \infty.$$

See [27, 31] for discussion of the explicit distribution of H.

The map $f \to \sup_{0 \leq t \leq 1} f(t)$ is continuous with respect to the Skorokhod topology but not with respect to the $*$-topology, so we cannot deduce the following from the *sampling* invariance principle.

Corollary 11. *Consider a model for random weighted trees (T, q) on $[n]$ which satisfies the functional invariance principle with scaling constants c. Then*

$$c \cdot \max_i \text{height}(i, T) \xrightarrow{d} \sup_{0 \leq t \leq 1} 2B_t^{\text{exc}}.$$

See [14] Section 4.1 for discussion of the explicit distribution of $\sup_t B_t^{\text{exc}}$.

To visualize the distinction between the sampling and functional invariance principles, consider trees $T^{(n)}$ satisfying the functional invariance principle with respect to uniform weights. Take $m = o(n)$ and make trees $\hat{T}^{(m+n)}$ by linking an arbitrary m-vertex tree $\tilde{T}^{(m)}$ to the root of $T^{(n)}$. Then $\hat{T}^{(m+n)}$ will still satisfy the *sampling* invariance principle, but in general not the functional invariance principle. One can easily make examples where the convergence of maximum heights assertion in Corollary 11 fails for $\hat{T}^{(m+n)}$. In the random walk coding, the point is that the walk excursion for $\tilde{T}^{(m)}$ is vanishingly short and so is not noticed by L_0 convergence. At the technical level, if one knows that the weak invariance principle holds, then to prove the strong invariance principle one needs only to prove tightness of the rescaled walks in the Skorokhod metric. It would be interesting to find useful sufficient conditions for tightness in our random tree context.

3. Formulating the sampling invariance principle for random mappings

In a sense, the rest of the paper is devoted to describing the "random mappings" analogs of the results for random trees in Section 2. The reader may wish first to look at Section 6 for specific models of random mappings.

For a mapping M and an element $i \in [n]$, consider the iterates $M^0(i) = i$, $M^1(i) = M(i)$, $M^{j+1}(i) = M(M^j(i))$, $j \geq 1$. There is a smallest number $L_1 + L_2$ such that

$$M^{L_1+L_2}(i) = M^{L_1}(i) \text{ for some } 0 \leq L_1 < L_1 + L_2$$

and the associated L_1 is unique. As shown in the left of figure 2, we picture the set of interates of i as a line of length L_1 attached to a cycle of length $L_2 - L_1$.

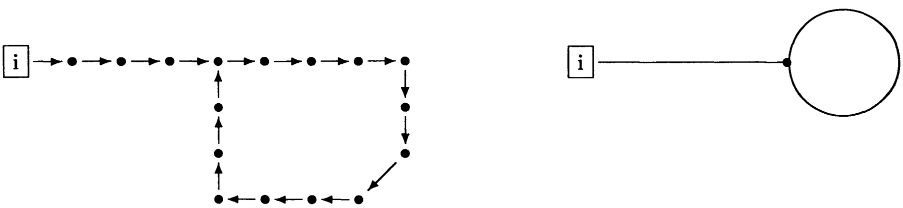


Figure 2.

For the *uniform* random mapping M, the elementary "birthday problem" argument shows

$$P(L_1 = x, L_2 = y) = (n)_{x+y}/n^{x+y+1}, \quad x = 0,1,2,\ldots;\ y = 1,2,\ldots$$

where $(n)_k := n(n-1)(n-2)\cdots(n-k+1)$. In the $n \to \infty$ limit,

$$n^{-1/2}(L_1, L_2) \xrightarrow{d} (\mathcal{L}_1, \mathcal{L}_2), \text{ where } f_{\mathcal{L}_1,\mathcal{L}_2}(\ell_1, \ell_2) = \exp(-(\ell_1 + \ell_2)^2/2). \tag{15}$$

We picture the limit as a "mapping with edge-lengths", as in right of figure 2: there are two edges, a line of length $\mathcal{L}_1$ and a loop of length $\mathcal{L}_2$.

Instead of starting with a single vertex, one can fix k distinct vertices $\kappa_1, \ldots, \kappa_k$ and consider the graph of all iterates $(M^j(\kappa_i), j \geq 0, 1 \leq i \leq k)$. Relabel vertex κ_i as i and unlabel other vertices. Picture the resulting graph as on the top of figure 3, which shows an example where no κ_j falls in the orbit of any κ_i, $i < j$. Take each edge to have length 1; then delete unlabeled degree-2 vertices to obtain a "mapping with edge-lengths" where the edge-lengths are integers. Call this graph ORBITS$(\kappa_1, \ldots, \kappa_k; M)$. For asymptotics, after suitably scaling edge-lengths we anticipate getting a "mapping with edge-lengths" $\mathbf{g}$ as pictured on the bottom in figure 3.

Such a graph $\mathbf{g}$ has the following properties.

(i) Each component consists of some number $\gamma \geq 1$ of trees attached to a directed cycle consisting of γ edges.

(ii) Each tree is an unordered binary tree, with labeled leaves. The degree-3 branchpoints within trees or where trees meet cycles are unlabeled.

(iii) The set of leaf-labels is $[k]$.

(iv) Each edge has a strictly positive real length.

Write $\mathbf{G}[k]$ for the set of such graphs, which one could call *mappings with edge lengths.* A graph $\mathbf{g} \in \mathbf{G}[k]$ can be specified by its *shape* (pedantically, the shape is the equivalence class of isomorphic leaf-labeled graphs-without-edge-lengths) and by its edge-lengths. Inductively, the number of edges equals $2k$ (adding a new leaf creates two extra edges) and the number of shapes equals $(2k-1)!! := (2k-1)(2k-3)(2k-5)\cdots 3 \cdot 1$ because the k'th leaf can be attached at $2k-1$ different places (the $2k-2$ existing edges, or a new component). This closely parallels the discussion of tree-with-edge-lengths in Section 2.1. Analogous to the

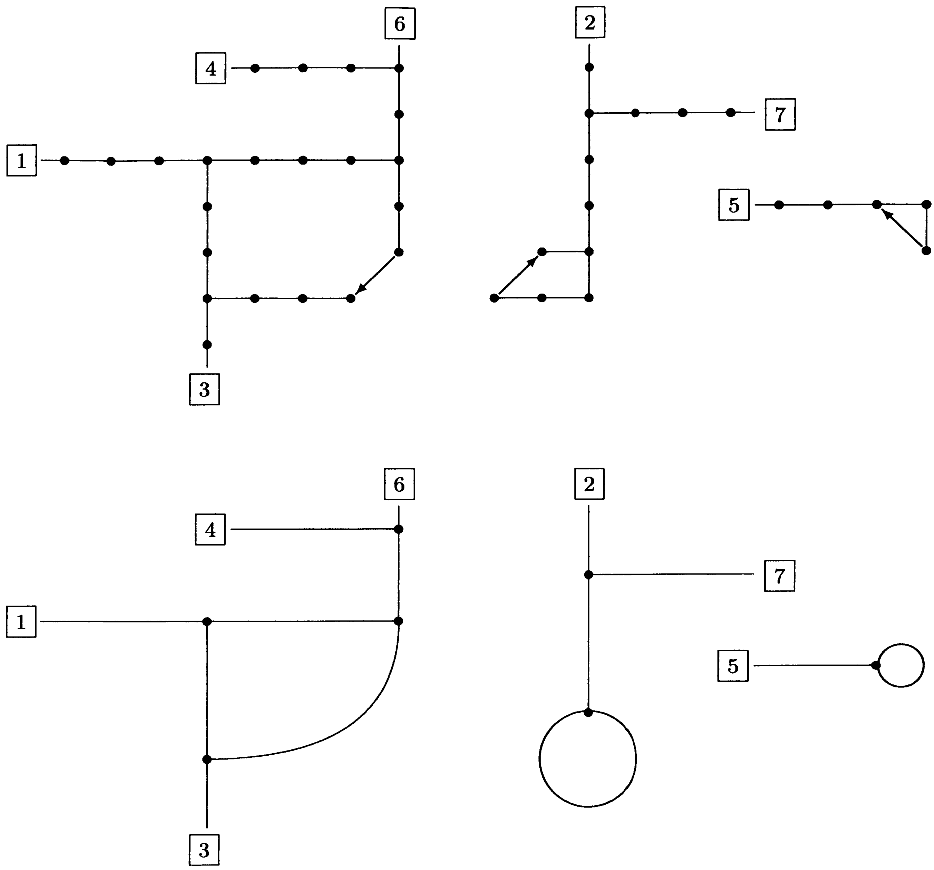


Figure 3. Instances of ORBITS($\kappa_1, \ldots, \kappa_7; M$) and of a $\mathbf{g} \in \mathbf{G}[7]$.

definition (1) of $T[k]$, we now define a distribution for a random mapping with edge lengths, $G[k]$, as follows.

For each possible shape g,

$$P(\text{shape}(G[k]) = g, \mathcal{L}_1 \in d\ell_1, \ldots, \mathcal{L}_{2k} \in d\ell_{2k}) = f_{2k}^*(\ell_1, \ldots, \ell_{2k})\, d\ell_1 \ldots d\ell_{2k} \quad (16)$$

where

$$f_{2k}^*(\ell_1, \ldots, \ell_{2k}) := \exp\left(-\tfrac{1}{2}\left(\sum_{j=1}^{2k} \ell_j\right)^2\right). \quad (17)$$

Compare with (2) and note the missing prefactor in (17); a calculation at (19) later will illuminate the connection between f_{2k}^* and f_{2k-1}. As before, definition (16) implies that shape($G[k]$) is uniform on the $(2k-1)!!$ possible shapes and that edge-lengths are independent of shape.

We can introduce *weighted* mappings in the same way as weighted trees: there is a probability distribution (weight function) $q = q^{(n)}$ on $[n]$. We can talk about convergence in $\mathbf{G}[k]$ as we did in $\mathbf{T}[k]$: $\mathbf{g}^{(n)} \to \mathbf{g}$ if shape($\mathbf{g}^{(n)}$) = shape($\mathbf{g}$) ultimately and the edge-lengths converge. We can now copy the format of Definition 1.

Definition 12. A model of random weighted mappings (M, q) on $[n]$ satisfies the sampling invariance principle with scaling constants $c = c^{(n)}$ if, as $n \to \infty$,

$$c \otimes \mathrm{ORBITS}(\kappa_1, \ldots, \kappa_k; M) \xrightarrow{d} G[k], \quad k \geq 1 \tag{18}$$

where $\kappa_1, \ldots, \kappa_k$ are independent of each other and of M with distribution $q^{(n)}$.

As noted below Definition 1, our standing hypothesis (3) on the weights q implies $\lim_n P(\kappa_1, \ldots, \kappa_k \text{ distinct }) = 1$ and we only require the left side to take values in $\mathbf{G}[k]$ on events of probability $\to 1$ as $n \to \infty$.

3.1. A CONSTRUCTION OF $G[k]$

Motivation for studying the random mapping with edge-lengths $G[k]$ with distribution (17) comes from its appearance as the limit (18) for the uniform random mapping (this is elementary, or a special case of Theorem 25 below). Now we already mentioned that the analogous random tree with edge-lengths $T[k]$ has a direct line-breaking construction [4, 9]. We will show below that there is a simple way to construct $G[k]$ from $T[k+1]$. This construction, illustrated in figure 4, is a "graphs with edge lengths" analog of Joyal's bijection exploited in [6], though we will not elaborate on the analogy here.

Start with a tree with edge lengths $\mathbf{t} = \mathbf{t}^1$ with leaves labeled $0, 1, \ldots, k$. In the figure, leaf i is denoted $\boxed{\mathrm{i}}$. Call the path from the root to leaf 0 the *spine*. Let j_1 be the junction where the path from leaf 1 to the root meets the spine. At j_1 cut away the edge leading toward leaf 0. Make the line from j_1 to root into a cycle by identifying the root with j_1, to form a component $\mathbf{g}^1$. In the remaining part of the original tree, make a new root at the endpoint previously at j_1 to define a new tree with edge lengths, say $\mathbf{t}^2$. Repeat recursively, letting j_2 be the junction where the path in $\mathbf{t}^2$ from the lowest-numbered leaf (except 0) meets the spine, to construct another component $\mathbf{g}^2$ with a cycle defined by identifying its root with j_2. Continue until the remaining tree $\mathbf{t}^{j+1}$ consists only of leaf 0 and a single edge; discarding that tree, the components $\mathbf{g}^1, \ldots, \mathbf{g}^j$ form a mapping with edge lengths. Cycles are directed according to the direction along the original spine from the root to leaf 0.

Abusing notation slightly, write $\mathbf{T}[0, k]$ for the set of trees with edge lengths with leaf-set $\{0, 1, \ldots, k\}$.

Lemma 13. *The map $J : \mathbf{T}[0, k] \to \mathbf{G}[k]$ described above carries the probability density f_{2k+1} defined at* (2) *to the density f^*_{2k} defined at* (17).

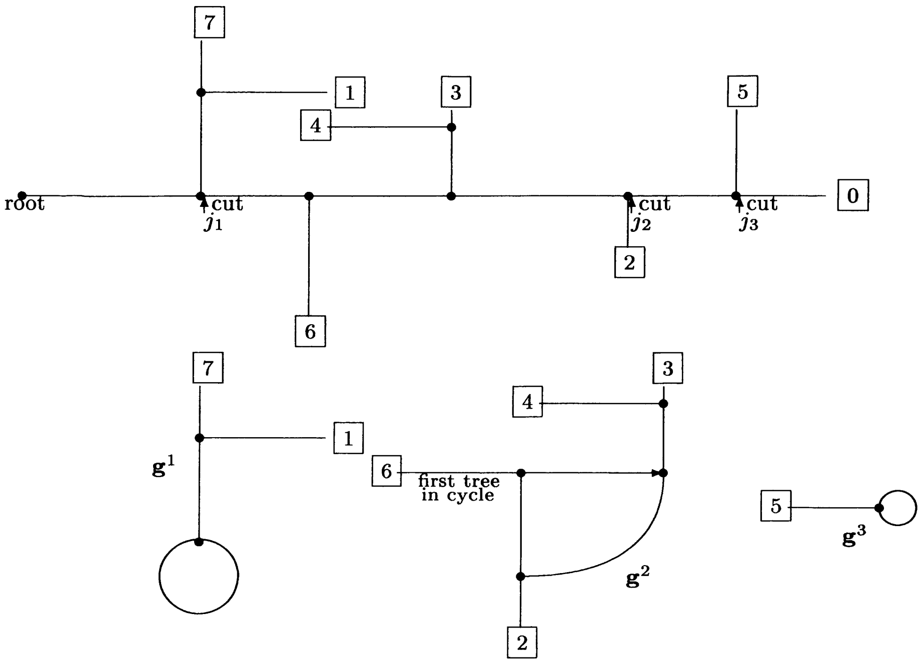


Figure 4. The map $J : \mathbf{T}[0,k] \to \mathbf{G}[k]$ which takes the distribution of $T[k+1]$ to the distribution of $G[k]$.

Proof. It is straightforward to check the map is a bijection between the two sets of *shapes*, each having cardinality $(2k-1)!!$. Checking the assertion about densities reduces to checking the consequence of integrating out the contribution from the length ℓ_0 of the discarded edge incident to 0, that is to checking

$$f^*_{2k}(\ell_1, \ldots, \ell_{2k}) = \int_0^\infty (\sum_{j=0}^{2k} \ell_j) \exp\left(-\tfrac{1}{2}\left(\sum_{j=0}^{2k} \ell_j\right)^2\right) d\ell_0. \tag{19}$$

But writing $s = \sum_{j=1}^{2k} \ell_j$, this is just the integration by parts formula

$$\int_0^\infty (\ell_0 + s) \exp(-\tfrac{1}{2}(\ell_0 + s)^2)\, d\ell_0 = \exp(-\tfrac{1}{2}s^2).$$

□

4. Brownian bridge and random mapping-walks

Here we recall from [7] how to code a mapping as a walk, and a version of the functional invariance principle for the uniform model of random mappings.

4.1. CODING MAPPINGS AS WALKS PLUS MARKS

Recall from Section 2.2 the coding of a rooted ordered tree as a walk. The corresponding coding for mappings is more intricate. Here's the key conceptual idea from [7], with a different coding.

> To a mapping M on $[n]$ one can associate a ***walk*** and ***marks***. The walk and marks determine the mapping, up to vertex-labels. Given a mapping, the definition of the walk and marks involves three levels of choices of orderings.
> (a) The trees in the mapping's digraph are naturally "unordered trees"; we need to make them ordered trees.
> (b) The components of the digraph are unordered; we need to impose an order.
> (c) The trees attached to a cycle in a component are (only) cyclically ordered; we need to specify a "first" tree.
> The mapping and these choices determine the walk. Some choices will make the probabilistic structure more tractable.

Here are the details, deferring order choices until later. A *walk* is a sequence $0 = w(0), w(1), w(2), \ldots, w(n-1)$ satisfying

$$w(i) \in \{0, 1, 2, 3, \ldots\}, \quad w(i+1) \leq w(i) + 1.$$

So maybe $w(i+1) < w(i)-1$. The *marks* are integers $0 = d(0) < d(1) < \ldots \leq n-1$, and for each marked integer $d(j)$ we require $w(d(j)) = 0$.

Using a mapping and order choices to define a walk and marks. An order on components and an order of trees within each component specify an order on all the trees, so breaking cyclic edges makes a forest (consisting of trees, whose roots are the original cyclic vertices) whose trees are in specified order. To each tree is associated a walk, as in Section 2.2. Concatenating the walks $\tilde{w}$ for each tree, in the specified order of trees, gives the walk w for the mapping M, and defines a relabeling of the vertices as $\{0, 1, \ldots, n-1\}$. Then mark each i for which, after this relabeling, vertex i is in a different component of the mapping than all vertices $i' < i$. So the marked vertices are the first vertex of each component.

Using the walk and marks to define a mapping. We will label vertices as $\{0, 1, \ldots, n-1\}$. The components of the digraph of M will be $\{0, 1, \ldots, d(1)-1\}$, $\{d(1), d(1)+1, \ldots, d(2)-1\}$, ... The value $w(i)$ will be the height of vertex i above the cycles of M. So the cyclic vertices i will be exactly the vertices with $w(i) = 0$. If $i < j$ are the positions of successive visits of $w(\cdot)$ to 0, then $(\tilde{w}(0), \tilde{w}(1), \ldots, \tilde{w}(j-i-1)) = (w(i), w(i+1), \ldots, w(j-1))$ is a walk which determines a tree, and the order of the excursions within each component determines the cyclic order in which trees are attached to the cycle.

Together with the walk we define

$$l(j) := \sum_{i=0}^{i} 1_{(w(i)=0)}. \tag{20}$$

So $l(j)$ is the number of vertices $i \leq j$ which are cyclic.

We now specify our order choices, for coding a mapping as a walk. Other choices are possible, as discussed in [1]. The choices will use external randomization.
(a) Within each tree, children of each vertex are put in uniform random order.
(b) Order components in q-biased random order. That is, choose random $v_1 \in [n]$ according to $q^{(n)}$, and let C_1 be the component containing vertex v_1. Then choose random $v_2 \in [n] \setminus C_1$ with probabilities proportional to $q^{(n)}$, and let C_2 be the component containing vertex v_2; and so on.
(c) Within component C_j, put trees in cyclic order such that the tree containing vertex v_j is last.

Summary. Given a deterministic or random mapping on $[n]$, the construction above leads to a random walk $(w(i), 0 \leq i \leq n-1)$ and random marks $0 = d(0) < d(1) < \ldots \leq n-1$, and a relabeling of vertices by $\{0, 1, \ldots, n-1\}$. From the walk and marks we can reconstruct the mapping digraph, up to permutation of vertices. Figure 5 later illustrates such a walk and marks (more precisely, their rescaling described below). Call $(w(i), 0 \leq i \leq n-1)$ the *mapping-walk.*

Finally, for a component C write $q(C) := \sum_{i \in C} q_i$ for the weight of C, and write cycle(C) for the cycle length in component C. We will give results for the joint distributions

$$((q(C_j), c \cdot \text{cycle}(C_j)), j \geq 1) \tag{21}$$

where as above the components (C_j) are in q-biased random order.

4.2. RESCALING MAPPING-WALKS

To discuss weak convergence as $n \to \infty$ of mapping-walks $(w(i), 0 \leq i \leq n) = (w^{(n)}(i), 0 \leq i \leq n)$ associated with random mappings $M = M^{(n)}$, we need to introduce rescalings. Given weights $q(i) = q^{(n)}(i)$, set

$$\tau(i) := q(0) + q(1) + \cdots + q(i-1)$$

and define the rescaled walk $(\bar{w}(t), 0 \leq t \leq 1)$ by

$$\bar{w}(t) := w(i) \text{ for } \tau(i) \leq t < \tau(i+1).$$

Define a rescaled "cyclic vertex counting" process $(\bar{l}(t), 0 \leq t \leq 1)$ by

$$\bar{l}(t) := l(i) \text{ for } \tau(i) \leq t < \tau(i+1).$$

And rescale the marks $0 = d(0) < d(1) < \ldots$ by defining

$$\bar{d}(j) := \tau(d(j)), \ j = 0, 1, 2, \ldots$$

so that $0 = \bar{d}(0) < \bar{d}(1) < \ldots < 1$.

4.3. THE BROWNIAN BRIDGE LIMIT

Write $B^{|\mathrm{br}|} = (B_t^{|\mathrm{br}|}, 0 \leq t \leq 1)$ for *reflecting Brownian bridge*, obtained from standard Brownian motion (B_t) via

$$B_t^{|\mathrm{br}|} := |B_t - tB_1|.$$

It turns out that the role of Brownian excursion for random trees is played by reflecting Brownian bridge for random mappings. To code the component structure of the mappings, we require some external randomization provided by an infinite i.i.d. uniform $(0,1)$ sequence $(\widetilde{U}_i)$. Define $D_0 = 0$,

$$D_1 := \inf\{t \geq \widetilde{U}_1 : B_t^{|\mathrm{br}|} = 0\}$$

and then for $r = 1, 2, \ldots$ define

$$D_{r+1} := \inf\{t \geq D_r + \widetilde{U}_{r+1}(1 - D_r) : B_t^{|\mathrm{br}|} = 0\}. \tag{22}$$

Write $(L(t), 0 \leq t \leq 1)$ for *local time at* 0 for $B^{|\mathrm{br}|}$, normalized so that $P(L(1) > s) = \exp(-s^2/2)$, $s > 0$ [36].

Aldous and Pitman proved a variety of results for the uniform model of random mappings, including the following.

Theorem 14 ([7]). *For the uniform random mapping M on $[n]$, with uniform weights, as $n \to \infty$*

$$n^{-1/2}\bar{w}^{(n)} \xrightarrow{d} 2B^{|\mathrm{br}|} \tag{23}$$

$$n^{-1/2}\bar{l}^{(n)} \xrightarrow{d} L \tag{24}$$

in the sense of weak convergence of processes in the usual Skorokhod metric on $D[0,1]$. Moreover

$$\left(\left(q(C_j^{(n)}), n^{-1/2} \cdot \mathrm{cycle}(C_j^{(n)})\right), j \geq 1\right)$$
$$\xrightarrow{d} \left((D_j - D_{j-1}, L(D_j) - L(D_{j-1})), j \geq 1\right) \tag{25}$$

The proof of Theorem 14 in [7] used a different, but asymptotically equivalent, way of coding trees as walks. Also, in section 4.1 we used a specific choice of ordering of components to define the mapping-walk. It turns out there is an alternate choice based on size-biasing of *cycles* which also leads to $B^{|\mathrm{br}|}$ asymptotics. These random mapping considerations lead to two different recursive decompositions of $B^{|\mathrm{br}|}$, whose structure is explored in [1], including explicit descriptions of the distribution of the right side of (25).

Note that the total cycle length is $\bar{l}^{(n)}(1) = \sum_j \mathrm{cycle}(C_j^{(n)})$. The asymptotic result

$$n^{-1/2} \cdot \sum_j \mathrm{cycle}(C_j^{(n)}) \xrightarrow{d} L(1) \tag{26}$$

follows from (24) but cannot be deduced directly from (25).

5. Functional implications of the sampling invariance principle for random mappings

We finally arrive at the central point of the paper. Though the assertion of convergence to $B^{|\mathrm{br}|}$ in Theorem 14 is elegant, the choice of extra properties (24,25) seem somewhat arbitrary, as do the details of the mapping-walk definition. In contrast, the sampling invariance principle is a "natural" assertion without any arbitrary choices. One anticipates, by analogy with Proposition 7 for random trees, that the sampling invariance principle itself should imply results in the general format of Theorem 14. Theorem 15 and Proposition 20 provide such results.

Theorem 15. *Suppose a model of random weighted mappings (M,q) on $[n]$ satisfies the sampling invariance principle with scaling constants $c = c^{(n)}$. Then as $n \to \infty$*

$$((q(C_j), c \cdot \mathrm{cycle}(C_j)), j \geq 1) \xrightarrow{d} ((D_j - D_{j-1}, L(D_j) - L(D_{j-1})), j \geq 1) \quad (27)$$

and

$$c\bar{w} \xrightarrow{d} 2B^{|\mathrm{br}|} \quad (28)$$

in the sense of weak convergence of processes in the $*$-metric defined in Section 2.4.

Remarks. Here (27) just repeats (25). The conclusion of Theorem 15 is weaker than Theorem 14 in two ways. First, in (28) we have a weaker topology on function space than in (23) – in the trees setting, this was the *only* difference. Second, we do not have convergence (26) of total cycle length (nor the refinement (24)). See Section 5.1 for further discussion of asymptotic distributions.

The proof of Theorem 15, which occupies the rest of this Section, follows the general outline of the proof of Proposition 7, with extra complications dealing with component ordering conventions. We will need an analog of Lemma 8, stated as Lemma 16 below, which links sampling invariance to Brownian bridge. We then proceed to the analog of (11), stated as Lemma 17. First we need to specify an operation $\mathbf{g}(\cdot)$ (analogous to $\mathbf{t}(\cdot)$ in Section 2.4) which describes how graphs-with-edge-lengths are obtained from functions and marks.

Construction (illustrated in figure 5). Given:
- a function $f : [0,1] \to [0,\infty)$;
- a non-decreasing function $l : [0,1] \to [0,\infty)$ with $l(0) = 0$ and whose points of increase are contained in $\{t : f(t) = 0\}$;
- marks $0 = d(0) < d(1) < \ldots \leq 1$, such that each $f(d(i)) = 0$;
- distinct $u_1, \ldots, u_k \subseteq (0,1)$

we construct a graph-with-edge-lengths $\mathbf{g}(u_1, \ldots, u_k; f, l, d)$ as follows.

(i) For each excursion of f from 0 which contains at least one of the u_i, construct the associated tree-with-edge-lengths $\mathbf{t}$ as defined above Lemma 8 (with the excursion interval I in place of $[0,1]$ and $\{i : u_i \in I\}$ in place of $[k]$ as the set

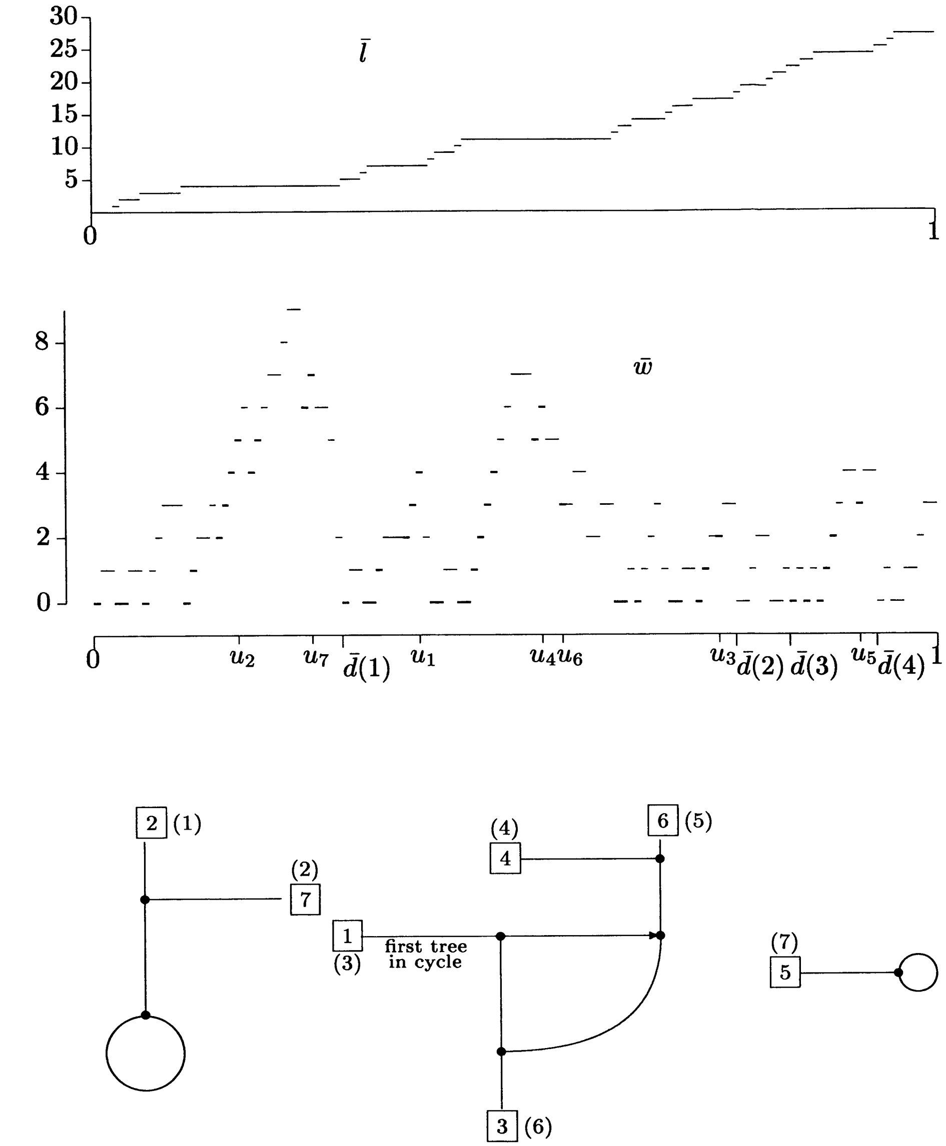


Figure 5. The rescaled walk $\bar{w}$ coding a certain mapping m on $[125]$, with equal weights, and the rescaled marks $\bar{d}(r)$. The points u_i identify seven vertices which can be labeled as $\{\kappa_1, \dots, \kappa_7\}$. The set $\mathrm{ORBITS}(\kappa_1, \dots, \kappa_7; m)$ is the digraph shown at the top of Figure 3. The walk steps corresponding to vertices in this sub-digraph are shown in thick lines. The top graph is the rescaled process $\bar{l}$ which counts cyclic vertices. Shown below is the graph-with-edge-lengths $\mathbf{g}(u_1, \dots, u_7; \bar{w}, \bar{l}, \bar{d})$, with leaf labels $G[k]$ and the corresponding leaf-ordered graph $\mathbf{g}^{\mathrm{ord}}(u_1, \dots, u_7; \bar{w}, \bar{l}, \bar{d})$, with leaf labels (i).

of leaves). Write $l(\mathbf{t})$ for $l(\cdot)$ evaluated at the starting point left$(\mathbf{t})$ of the excursion coding $\mathbf{t}$.

(ii) For trees $\mathbf{t}^j, \dots, \mathbf{t}^J$ such that left$(\mathbf{t}^j) < \dots <$ left$(\mathbf{t}^J)$ and such that these are all the trees $\mathbf{t}$ with left$(\mathbf{t}) \in [d(r-1), d(r))$ for some r, create a cyclic path between their roots, where the cyclic edge from root$(\mathbf{t}^i)$ to root$(\mathbf{t}^{i+1})$ has length $l(\mathbf{t}^{i+1}) - l(\mathbf{t}^i)$ for $j \leq i < J$, and the cyclic edge from root$(\mathbf{t}^J)$ to root$(\mathbf{t}^j)$ has length $l(d(r)) - l(\mathbf{t}^J) + l(\mathbf{t}^j) - l(d(r-1))$.

Observe that external randomization appears twice in the statement of Theorem 15: in the hypothesis of sampling invariance, and in the conclusion where the mapping-walk is defined using q-biased random order of components. For the proof (but not the statement) it will be important to take these two randomization operations to be independent. So we now introduce i.i.d. uniform $(0,1)$ random variables (U_i), independent of the mappings and mapping-walks in section 4.1, and independent of $B^{|\mathrm{br}|}$ and $(\tilde{U}_i, D_i)$ in section 4.3.

Lemma 16. $\mathbf{g}(U_1, \dots, U_k; 2B^{|\mathrm{br}|}, L, D) \stackrel{d}{=} G[k]$, *where $L(t)$ is local time for $B^{|\mathrm{br}|}$ at 0 and $D = (D_0, D_1, \dots)$ are the marks defined at* (22).

Historically, Lemma 8 was first proved [4] as a consequence of the functional invariance principle for uniform random trees, and later reproved [26, 43] directly via excursion theory for Brownian excursion. Doubtless Lemma 16 could also be proved via excursion theory for Brownian bridge, but we shall deduce it during the course of the next proof from Theorem 14 for uniform random mappings. See [43] for further results in the spirit of Lemmas 8 and 16 for Brownian bridge and meander.

Lemma 17. *Under the hypotheses of Theorem 15,*

$$c\left(\bar{w}(V_1), \inf_{V_1 \leq t \leq V_2} \bar{w}(t), \bar{w}(V_2), \inf_{V_2 \leq t \leq V_3} \bar{w}(t), \dots, \bar{w}(V_k)\right)$$

$$\stackrel{d}{\to} \left(2B^{|\mathrm{br}|}(V_1), \inf_{V_1 \leq t \leq V_2} 2B^{|\mathrm{br}|}(t), 2B^{|\mathrm{br}|}(V_2), \inf_{V_2 \leq t \leq V_3} 2B^{|\mathrm{br}|}(t), \dots, 2B^{|\mathrm{br}|}(V_k)\right)$$

where $(V_1, \dots, V_k)$ are the order statistics of $(U_1, \dots, U_k)$.

Proof of Lemmas 16 and 17. First fix n. For $\kappa(u)$ defined at (9), $\kappa(U_1), \dots, \kappa(U_k)$ are independent random vertices chosen from distribution q, as required for the sampling invariance principle. Moreover, from their definitions ORBITS$(\kappa(U_1), \dots, \kappa(U_k); M)$ is the same graph-with-edge-lengths (except for heights of branch-points differing by 1) as $\mathbf{g}(U_1, \dots, U_k; \bar{w}, \bar{l}, \bar{d})$, where $\bar{w}$ is the rescaled mapping-walk defined in section 4.2, and $\bar{l}$ and $\bar{d}$ are also as defined there. Letting $n \to \infty$, the sampling invariance principle implies

$$c \otimes \mathbf{g}(U_1, \dots, U_k; \bar{w}, \bar{l}, \bar{d}) \stackrel{d}{\to} G[k] \text{ as } n \to \infty. \tag{29}$$

Now in the special case of uniform random mappings, it is not hard to see that the argument in [7] to prove Theorem 14 implies

$$c \otimes \mathbf{g}(U_1, \dots, U_k; \bar{w}, \bar{l}, \bar{d}) \stackrel{d}{\to} \mathbf{g}(U_1, \dots, U_k; 2B^{|\mathrm{br}|}, L, D). \tag{30}$$

Thus the right sides of (29) and (30) must be equal in distribution, establishing Lemma 16. Returning to the general case of mappings satisfying the sampling invariance principle, we have now established (30). Now the order statistics $(V_1, \ldots, V_k)$ are related to $(U_1, \ldots, U_k)$ via some permutation $V_i = U_{\pi(i)}$. Construct $\mathbf{g}^{\mathrm{ord}}(U_1, \ldots, U_k; 2B^{|\mathrm{br}|}, L, D)$ and $\mathbf{g}^{\mathrm{ord}}(U_1, \ldots, U_k; \bar{w}, \bar{l}, \bar{d})$ from $\mathbf{g}(U_1, \ldots, U_k; 2B^{|\mathrm{br}|}, L, D)$ and $\mathbf{g}(U_1, \ldots, U_k; \bar{w}, \bar{l}, \bar{d})$ by relabeling leaf $\pi(i)$ as leaf i, to get a "leaf-ordered" graph (illustrated in figure 5 with leaf i labeled (i)). Clearly $\mathbf{g}^{\mathrm{ord}}(U_1, \ldots, U_k; 2B^{|\mathrm{br}|}, L, D)$ determines the $(2k-1)$-vector appearing on the right side in Lemma 17, and $\mathbf{g}^{\mathrm{ord}}(U_1, \ldots, U_k; \bar{w}, \bar{l}, \bar{d})$ determines the $(2k-1)$-vector appearing on the left side. So to prove Lemma 17 it is enough to prove the "ordered" version of (30):

Lemma 18. *As $n \to \infty$*

$$c \otimes \mathbf{g}^{\mathrm{ord}}(U_1, \ldots, U_k; \bar{w}, \bar{l}, \bar{d}) \xrightarrow{d} \mathbf{g}^{\mathrm{ord}}(U_1, \ldots, U_k; 2B^{|\mathrm{br}|}, L, D).$$

Proof. First fix n. The external randomization used to order components of a mapping M as part of the walk coding can be implemented as follows. Take i.i.d. $(\tilde{\kappa}_i)$ with distribution q. Take C_1 to be the component containing $\tilde{\kappa}_1$ and then take C_r to be the component containing $\tilde{\kappa}_{I_r}$ where

$$I_r = \min\{i : \tilde{\kappa}_i \notin \cup_{1 \le j < r} C_j\}.$$

Fixing K and k, and letting $n \to \infty$, the sampling invariance principle with $K + k$ sampled vertices implies

$$c \otimes \mathrm{ORBITS}(\tilde{\kappa}_1, \ldots, \tilde{\kappa}_K, \kappa(U_1), \ldots, \kappa(U_k); M) \xrightarrow{d} G[K + k].$$

But from a realization of $\mathrm{ORBITS}(\tilde{\kappa}_1, \ldots, \tilde{\kappa}_K, \kappa(U_1), \ldots, \kappa(U_k); M)$ we can not only derive (exactly) the realization of $\mathrm{ORBITS}(\kappa(U_1), \ldots, \kappa(U_k); M)$, but also we get information about the permutation π taking U_i to $V_i = U_{\pi(i)}$. For instance, if a component contains $\tilde{\kappa}_1$ then it is the first component. In fact, the order of tree-components is determined exactly *unless*

$$\exists 1 \le j \le k : \ \kappa(U_j) \text{ in different component from each } \tilde{\kappa}_i, \ 1 \le i \le K.$$

Moreover the distribution of within-tree leaves is determined by the random depth-first ordering imposed by that aspect of external randomization. So to prove that the limit distribution in Lemma 18 does not depend on the model of random mappings, it is enough to prove

$$\lim_K \lim_n P\left(\exists 1 \le j \le k : \begin{array}{c} \kappa_j \text{ in different component} \\ \text{from each } \tilde{\kappa}_i, \ 1 \le i \le K \end{array} \right) = 0. \tag{31}$$

The probability in question is bounded by

$$kP(\tilde{\kappa}_1 \text{ in different component from each } \tilde{\kappa}_i, \ 2 \le i \le K + 1). \tag{32}$$

By the sampling invariance principle, the $n \to \infty$ limit of the probability in (32) does not depend on the model of random mapping. Using Theorem 14, the limiting distribution of component weights is $(D_r - D_{r-1}, r \geq 1)$. So the quantity in (31) is bounded by

$$\lim_K \, kE(1 - D_1)^K$$

which clearly equals 0. So we have shown that the limit distribution in Lemma 18 does not depend on the model of random mappings. Use Theorem 14 again to identify the limit distribution as

$$\mathbf{g}^{\mathrm{ord}}(U_1, \ldots, U_k; 2B^{|\mathrm{br}|}, L, D). \qquad \square$$

Proof of Theorem 15. Having established Lemma 17, the same argument as used in the proof of Proposition 7 shows $c^{(n)}\bar{W}_n \xrightarrow{d} 2B^{|\mathrm{br}|}$. Regarding (27), we will prove

$$(q(C_1), c \cdot \mathrm{cycle}(C_1)) \xrightarrow{d} (D_1, L(D_1)) \tag{33}$$

and the full version is similar. Recalling $(V_i, 1 \leq i \leq k)$ are the order statistics of $(U_i, 1 \leq i \leq k)$, write

$$I_k := \max\{i : \kappa(V_i) \text{ in same component as } \kappa(V_1)\}.$$

Then $V_{I_k} \leq q(C_1) \leq V_{I_k+1}$. Now (30) implies that as $n \to \infty$ for fixed k,

$$(V_{I_k}, V_{I_k+1}) \xrightarrow{d} (V_{J_k}, V_{J_k+1}) \tag{34}$$

where $J_k := \max\{\ i\ :\ V_i \leq D_1\ \}$. But as $k \to \infty$ we clearly have $(V_{J_k}, V_{J_k+1}) \xrightarrow{d} (D_1, D_1)$ and this establishes

$$q(C_1) \xrightarrow{d} D_1. \tag{35}$$

Now the sampling invariance principle (for the orbit of a single vertex), and our "q-biased component order" convention, immediately imply that $c \cdot \mathrm{cycle}(C_1)$ converges in distribution to a limit which does not depend on the model; so by Theorem 14 the limit distribution is that of $L(D_1)$. Repeating the argument above, convergence (34, 35) holds jointly with convergence of $c \cdot \mathrm{cycle}(C_1)$ and the joint limit does not depend on the model; so again the joint limit distribution is that arising in the uniform model, $(D_1, L(D_1))$, and (33) is established. $\square$

5.1. DISTRIBUTIONAL LIMITS IMPLIED BY THE SAMPLING INVARIANCE PRINCIPLE

One can immediately use Theorem 15 to see that the sampling invariance principle implies convergence of rescaled cumulative height profiles, as for trees in Corollary 10.

Corollary 19. *Consider a model of random weighted mappings (M,q) on $[n]$ which satisfies the sampling invariance principle with scaling constants c. Define the* rescaled cumulative height profile *by*

$$H_n(u) := \sum_{i=1}^{n} q_i 1_{(\mathrm{height}(i,M)\le u/c)}.$$

Then

$$H_n \xrightarrow{d} H$$

as random elements of $\mathcal{D}$, where

$$H(u) := \int_0^1 1_{(2B_t^{|\mathrm{br}|}\le u)}\, dt; \quad 0 \le u < \infty.$$

See [28, 44] for discussion of the explicit distribution of H.

It is true that the sampling invariance principle implies convergence of rescaled tree-sizes, but this cannot be deduced from Theorem 15. Instead we can give a direct proof.

Proposition 20. *Consider a model of random weighted mappings (M,q) on $[n]$ which satisfies the sampling invariance principle with scaling constants c. Let*

$$\Delta_1^n \ge \Delta_2^n \ge \dots; \quad \sum_i \Delta_i^n = n$$

be the q-measures of the tree-components of the digraph of M^n. Then as $n \to \infty$

$$(\Delta_i^n, i \ge 1) \xrightarrow{d} (\Delta_i^\infty, i \ge 1) \tag{36}$$

where the right side denotes the lengths of excursions from 0 *for $B^{|\mathrm{br}|}$, in decreasing order.*

See [46] for discussion of the limit distribution.

Proof. Let $(U_j^n, j \ge 1)$ be independent vertices of M^n with distribution q. For each n and k there is a random partition $\mathcal{P}_k^n$ of $[k]$ defined by

> j_1 and j_2 are in the some component of the partition if $U_{j_1}^n$ and $U_{j_2}^n$ are in the same tree-component of the digraph of M^n.

By results on exchangeable random partitions going back to Kingman [33], convergence (36) is equivalent to the assertion that, for each fixed k,

$$\mathcal{P}_k^n \xrightarrow{d} \mathcal{P}_k^\infty \text{ as } n \to \infty \tag{37}$$

where $\mathcal{P}_k^\infty$ is the partition defined analogously in terms of i.i.d. uniform$(0,1)$ random variables and the excursions of $B^{|\mathrm{br}|}$; and where "convergence in distribution" has its elementary interpretation because the number of possible partitions is finite. Now implicit in the arguments of [7] is that (36) and hence (37) holds for the uniform model of random mappings with uniform weights. It is therefore sufficient to show that the $n \to \infty$ limit of $\mathcal{P}_k^n$ is the same in all models satisfying the sampling invariance principle. But this is clear from the definition: $c \otimes \text{ORBITS}(\kappa_1, \ldots, \kappa_k; M) \xrightarrow{d} G[k]$ implies the shapes converge, and the shape determines the partition. □

5.2. THE FUNCTIONAL INVARIANCE PRINCIPLE FOR RANDOM MAPPINGS

As we have already mentioned, it is not easy to decide exactly how to define the functional invariance principle for random mappings. Here is our tentative definition.

Definition 21. A model of random weighted mappings (M, q) on $[n]$ satisfies the functional invariance principle with scaling constants $c = c^{(n)}$ if, as $n \to \infty$, the assertions of Theorem 14 hold (with $n^{-1/2}$ replaced by $c^{(n)}$) and the assertion of Proposition 20 holds.

With this definition it is true – though we omit details – that the functional invariance principle implies the sampling invariance principle.

As with trees in Corollary 11, we can deduce convergence of maximal heights from the functional, but not sampling, invariance principle.

Corollary 22. *Consider a model of random weighted mappings (M, q) on $[n]$ which satisfies the functional invariance principle with scaling constants c. Then*

$$c \cdot \max_i \text{height}(i, M) \xrightarrow{d} \sup_{0 \le t \le 1} 2B_t^{|\text{br}|}.$$

The limit distribution of $\sup_t B_t^{|\text{br}|}$ here is of course the classical Kolmogorov-Smirnov statistic ([23] equation 7.8.11).

To visualize the distinction between the sampling and functional invariance principles, consider mappings $M^{(n)}$ satisfying the functional invariance principle with respect to uniform weights. Take $m = o(n)$ and make a mapping $\hat{M}^{(m+n)}$ by including an arbitrary mapping $\tilde{M}^{(m)}$ on $\{n+1, \ldots, n+m\}$. Then $\hat{M}^{(m+n)}$ will still satisfy the *sampling* invariance principle, but in general not the functional invariance principle. One can easily make examples of $\tilde{M}^{(m)}$ so that the convergence of maximum heights assertion in Corollary 22 fails, or the convergence of total cycle length (26) fails. Indeed this highlights a subtlety of the joint convergence of component weights and cycle lengths in (27); saying that the cycle lengths of the largest components converge is not saying that the longest cycle lengths converge.

Example 26 later provides a more concrete example.

6. Models for non-uniform random mappings

As already mentioned, the definition of the functional invariance principle for random mappings was motivated by Theorem 14, the case of uniform random mappings. There is no obvious mappings analog of CBP(n) trees, but p-trees and P-trees do have natural mapping analogs. In the next Sections we show that direct analysis of these models can lead to proofs of the sampling invariance principle.

6.1. THE SAMPLING INVARIANCE PRINCIPLE FOR p-MAPPINGS

Our proof will use Poisson approximation to evaluate asymptotics in the exact distribution of ORBITS($[k]; M$), where $[k] = \{1, 2, \ldots, k\}$. We start by isolating the Poisson approximation result we need.

For a probability distribution p on $[n]$ write

$$c_p := \sqrt{\sum_i p_i^2}.$$

It is well known (see e.g. [18] and papers cited there) that the elementary analysis of the uniform "birthday problem" extends to the non-uniform case under condition (7), as follows.

Lemma 23. *Let $(p^{(n)})$ satisfy (7). Suppose $m = m^{(n)}$ satisfies $mc_p \to s$ for some $0 < s < \infty$. For each n let $(\xi_i, 1 \le i \le m)$ be i.i.d.(p). Then*

$$P(\xi_i, 1 \le i \le m \textit{ are all distinct }) \to \exp(-s^2/2).$$

The precise result we need is the following variant.

Corollary 24. *In the setting of Lemma 23, fix $k \ge 1$. For each n let $i_1, j_1, i_2, j_2, \ldots, i_k, j_k$ be distinct elements of $[m]$. Then*

$$P\begin{pmatrix} \xi_i \notin [k], i \in [m];\ \xi_i \textit{ distinct for all} \\ i \in [m] \setminus \{j_1, \ldots, j_k\}; \xi_{j_u} = \xi_{i_u}, 1 \le u \le k \end{pmatrix} \sim c_p^{2k} \exp(-s^2/2).$$

Proof. First observe

$$P(\xi_i \notin [k], i \in [m]) \to 1 \tag{38}$$

because the complementary probability is at most $km \max_i p_i = o(mc_p) = o(1)$. Conditioning on the event in (38) is equivalent to replacing $(p^{(n)})$ by conditioned probabilities which satisfy the same hypotheses, so we can reduce the problem to proving

$$P\begin{pmatrix} \xi_i \text{ distinct for all } i \in [m] \setminus \{j_1, \ldots, j_k\}; \\ \xi_{j_u} = \xi_{i_u}, 1 \le u \le k \end{pmatrix} \sim c_p^{2k} \exp(-s^2/2). \tag{39}$$

By a similar conditioning argument and Lemma 23,

$$P(\xi_i, i \in [m] \setminus \{j_1, \ldots, j_k\} \text{ are distinct } | \xi_{i_1}, \ldots, \xi_{i_k}) \sim \exp(-s^2/2)$$

uniformly on $\{\xi_{i_1}, \ldots, \xi_{i_u}$ distinct $\}$. From the definition of c_p and (7),

$$P(\xi_{j_u} = \xi_{i_u}, 1 \le u \le k;\ \xi_{i_u} \text{ distinct}, 1 \le u \le k) \sim c_p^{2k}.$$

Combine these two relations to obtain (39). $\square$

Theorem 25 (Weak invariance principle for p-mappings). *For a sequence of p-mappings M, if the sequence $(p^{(n)})$ satisfies (7), then as $n \to \infty$*

$$c_p \otimes \mathrm{ORBITS}([k]; M) \xrightarrow{d} G[k], \quad k \geq 1.$$

In particular, the sampling invariance principle holds for any weight functions q, with scaling constants c_p.

The proof of the first assertion will be given soon. The second assertion follows from the first because, by relabeling vertices, the first holds for any deterministic choice of k vertices, and hence for the q-random choices in the sampling invariance principle.

Note in particular that Theorem 25 implies that, under (7), the rescaled cumulative height profile converges (Corollary 19). O'Cinneide and Pokrovskii [42] Theorem 2.1 used loosely similar methods to prove convergence (under the same hypotheses) of the rescaled cumulative height process to an unspecified limit. The next example shows that (7) is not enough to imply the functional invariance principle, though we speculate that a very slight strengthening would be enough.

Example 26. Condition (7) is not sufficient to imply the functional invariance principle for p-mappings.

Take

$$p_i^{(n,m)} = \tfrac{1}{2n}, 1 \leq i \leq n; \quad p_i^{(n,m)} = \tfrac{1}{2m}, n+1 \leq i \leq n+m.$$

Write $M = M_{n,m}$ for the associated p-mapping on $[n+m]$. Take $m = m(n)$ such that $m/n \to \infty$, and then (7) holds with $c(p^{(n,m)}) \sim \frac{1}{2}n^{-1/2}$. Now fix n. We shall show that, as $m \to \infty$,

$$\max_i \mathrm{height}(i, M) \to \infty \text{ in probability.} \tag{40}$$

By taking $m = m(n) \to \infty$ sufficiently fast, (40) implies that Corollary 22 fails for $M_{n,m}$, and hence the functional invariance principle fails.

Fix n. Choose $L = L(m) \to \infty$ so that

$$\tfrac{m}{2L} 4^{-L} \to \infty.$$

Take $i_1 = m + 1$ and define $A_1 = \{i_1, M(i_1), M^{(2)}(i_1), \ldots, M^{(L-1)}(i_1)\}$. Inductively, for $2 \leq r \leq m/(2L)$ let i_r be the minimum element of $\{n+1, \ldots, n+m\} \setminus \cup_{1 \leq s < r} A_s$ and let $A_r = \{i_r, M(i_r), M^{(2)}(i_r), \ldots, M^{(L-1)}(i_r)\}$. Then the conditional probability of the event

A_r consists of L elements, distinct from each other and from $[n] \cup \cup_{1 \leq s < r} A_s$

is at least $(1/4)^L$. So the number $N(m)$ of such events (for some $r \leq m/(2L)$) tends in probability to infinity as $m \to \infty$, by specification of $L(m)$. Now (40) follows easily.

Proof of Theorem 25. Fix $k \geq 1$. Fix $\mathbf{g} \in \mathbf{G}[k]$ with total edge-length $\sum \ell_i = s$. We shall prove Theorem 25 by proving the corresponding, formally stronger, *local* limit theorem

$$\text{if } c_p \otimes \mathbf{g}^{(n)} \to \mathbf{g} \text{ then } P(\mathrm{ORBITS}([k];M) = \mathbf{g}^{(n)}) \sim c_p^{2k} \exp(-s^2/2). \tag{41}$$

Here $\mathbf{g}^{(n)}$ denotes a possible value of ORBITS$([k];M)$ which is in $\mathbf{G}[k]$. Note that, although for finite n the graph ORBITS$([k];M)$ may not be in $\mathbf{G}[k]$ (for example, because it may have degree-4 vertices or non-leaf labeled vertices), combining (41) with the fact that $G[k]$ is a ***probability*** distribution will imply $P(\mathrm{ORBITS}([k];M) \in \mathbf{G}[k]) \to 1$ and hence we need only consider the case $\mathbf{g}^{(n)} \in \mathbf{G}[k]$.

Consider the chance that, for a random p-mapping, ORBITS$([7];M)$ is exactly the graph $\mathbf{g}$ at the top of figure 3. One can construct ORBITS$([7];M)$ from an i.i.d.(p) sequence $(\xi_j, 1 \leq j \leq n)$ as follows. Declare the iterates of 1, that is $M(1), M^2(1), M^3(1), \ldots$, to be the values $\xi_1, \xi_2, \ldots, \xi_{r_1}$ for $r_1 := \min\{j : \xi_j \in \{1, \xi_1, \ldots, \xi_{j-1}\}\}$. Then, if $2 \notin \{1, \xi_1, \ldots, \xi_{r_1 - 1}\}$, declare the iterates of 2 to be the subsequent ξ-values $\xi_{r_1+1}, \xi_{r_1+2}, \ldots, \xi_{r_2}$ until $r_2 := \min\{j > r_1 : \xi_j \in \{1, 2, \xi_1, \ldots, \xi_{j-1}\}\}$; and so on. So the probability under consideration is exactly the chance that the i.i.d.(p) sequence (ξ_i) starts with a "pattern" of the form

$$\overset{[1]}{\cdot} \cdot \cdot \, a \cdot \cdot \cdot \, b \cdot \cdot \cdot \cdot \cdot \, c \cdot \cdot \, a \overset{[2]}{\cdot} d \cdot \cdot \, e \cdot \cdot \cdot \cdot \cdot \, e \overset{[3]}{\cdot} c \overset{[4]}{\cdot} \cdot \cdot \, f \cdot b \overset{[5]}{\cdot} \cdot \, g \cdot \cdot \, g \overset{[6]}{f} \overset{[7]}{\cdot} \cdot \cdot \, d$$

whose meaning we now explain. The orbit of vertex 1 in figure 3 consists of a path of length 4 attached to a cycle of length 13. To create such an orbit we need $\xi_4 = \xi_{17}$ and we need the other values $\xi_1, \ldots, \xi_3, \xi_5, \ldots, \xi_{16}$ to be distinct and different from the former common value. This makes $r_1 = 17$. Similarly, the orbit of vertex 2 consists of a path of length 5 attached to a cycle of length 5; To create such an orbit we need $\xi_{27} = \xi_{22}$ and we need the other values $\xi_{18}, \ldots, \xi_{21}, \xi_{23}, \ldots, \xi_{26}$ to be distinct from each other and from previous ξ-values. This makes $r_2 = 27$. Vertex 3 is attached by a path of length 2 to a particular point of the cycle from vertex 1; this requires $\xi_{29} = \xi_k$ for a particular k (which turns out to be $k = 14$). And so on. The upshot is that the graph in figure 3 corresponds exactly to the case where $\xi_1, \ldots, \xi_{46}$ fit the "pattern" shown:

(i) the successive ξ-valuse denoted as $\cdots\cdot$ are distinct except where indicated otherwise; and they are distinct from [7];

(ii) $\xi_{i_u} = \xi_{j_u}$ for each of the 7 pairs (i_u, j_u) whose positions are denoted by symbols a, b, c, d, e, f, g: for instance, symbol a at positions $(4, 17)$ indicates that $\xi_{17} = \xi_4$, and symbol e at positions $(22, 27)$ indicates that $\xi_{22} = \xi_{27}$. In the pattern, superscript $[i]$ denotes position r_{i+1}, that is (because $M(i) = \xi_{r_{i+1}}$) the position of the ξ-value giving $M(i)$.

Applying this analysis to a general possible $\mathbf{g}^{(n)} \in \mathbf{G}[k]$, we see that $P(\mathrm{ORBITS}([k];M) = \mathbf{g}^{(n)})$ is precisely equal to a probability of the form appearing in Corollary 24, and so that Corollary implies (41). $\square$

6.2. RANDOM WALK P-MAPPINGS

Recall that a P-mapping is a random mapping M on $[n]$ such that $M(1),\ldots,M(n)$ are independent with $P(M(i)=j)=p_{ij}$ for a Markov transition matrix P. One can imagine qualitatively different hypotheses on P which would lead to the sampling invariance principle—note that in this context the stationary distribution of P would be a natural choice of weight function.

Here we shall indicate one possible type of hypothesis. Suppose (for some subsequence of $n\to\infty$) we have size-n groups G_n with group operation denoted by $*$. Suppose each $P=P^{(n)}$ is of the form $p_{ij}=\mu(i^{-1}*j)$ for some probability distribution $\mu=\mu^{(n)}$ on G_n. That is, P is the transition matrix of a random walk $X_m=\xi_1*\xi_2*\ldots*\xi_m$ whose steps ξ have distribution μ. Suppose there exist constants $\rho=\rho^{(n)}, t=t^{(n)}$ and γ not depending on n such that the following hold as $n\to\infty$.

(i) $t\to\infty,\ t/\rho\to 0,\ n^{-1/2}\rho\to 0$.

(ii) $n^{1/2}\sum_{i=1}^{\rho}P(X_i=\text{identity})\to 0$.

(iii) $E(N(\rho)|N(\rho)\geq 1)\to\gamma$, where $N(\rho):=\sum_{i=0}^{\rho}\sum_{j=0}^{\rho}1_{(X_i=Y_j)}$ for independent random walks (X_i) and (Y_i) with Y_0 independent uniform on G_n.

(iv) $\max_{g\in G_n}n^2|P(X_t=g)-n^{-1}|\to 0$.

Proposition 27 ([2], Prop. 33)**.** *Under the hypotheses above, as* $n\to\infty$

$$(\gamma n)^{-1/2}\otimes\text{ORBITS}(\text{identity};M)\ \stackrel{d}{\to}\ G[1].$$

Though we shall not give details, the analysis in [2] can be extended to show

Proposition 28. *Under the hypotheses above,* $(M,\text{uniform})$ *satisfies the sampling invariance principle with constants* $(\gamma n)^{-1/2}$*.*

As an illustrative example ([2] Example 34) take G_n to be the cyclic group on $\{0,1,\ldots,n-1\}$ and take $\mu(i)=\frac{1}{2n}$ for $i\neq 1$ and $\mu(1)=\frac{1}{2}+\frac{1}{2n}$. Here one can show the hypotheses hold with $\gamma=4/3$. Note $P(M(1)=2)=1/2+1/(2n)$; this "immediate dependence" implies we do not have

$$(\gamma n)^{-1/2}\otimes\text{ORBITS}(1,2;M)\ \stackrel{d}{\to}\ G[2]$$

in contrast to Theorem 25. This explains why we use randomly-chosen vertices κ_i in Definition 12.

7. Remarks

(a) The coding of trees as walks in this paper is via "depth first search" or the "exploration process", used also in e.g. [35]. See [38] for further references and the asymptotic equivalence of variant definitions. Note that a different family of *breadth-first* walks are used for other purposes, e.g. [5].

(b) The literature on asymptotics for p-trees [18, 11, 9, 8] develops a complete theory of all possible limits of p-trees without assumption (7), in which setting the limit tree-with-edge-lengths $T[k]$ will have some different distribution. This is the sense in which condition 7 is analogous to Lindeberg's condition in the central limit theorem. The method of deriving random p-mapping asymptotics from p-trees asymptotics via Joyal's bijection, mentioned in Section 1.1, should lead to a parallel complete description of all possible limits of p-mappings. We plan to investigate this elsewhere [6]. At a technical level, note that Theorem 25 holds for arbitrary weights $q^{(n)}$ whereas the method of [6] seems tied to the choice $q^{(n)} = p^{(n)}$.

(c) Conceptually, one can think of using $B^{|\mathrm{br}|}$ to construct a *continuum random mapping* analogously to the *continuum random trees* constructed from Brownian excursion. This idea also may be developed elsewhere.

(d) It is intriguing, and easy to check, that the distribution of $G[k]$ is (up to scaling constants) the *maximum entropy* distribution on its state space $\mathbf{G}[k]$ subject to the constraint $E($ sum of edge-lengths$)^2 =$ constant. The corresponding assertion is not true for $T[k]$ because of the prefactor in (2).

Acknowledgement. We thank Grégory Miermont for a careful reading of a draft.

References

1. Aldous, D. J. and Pitman, J. (2002) Two recursive decompositions of Brownian bridge related to the asymptotics of random mappings, Technical Report 595, Dept. Statistics, U.C. Berkeley, 2002.
2. Aldous, D. J. (1985) Self-intersections of random walks on discrete groups, *Math. Proc. Cambridge Philos. Soc.*, **98**, 155–177.
3. Aldous, D. J. (1990) A random tree model associated with random graphs, *Random Structures Algorithms*, **1** 383–402.
4. Aldous, D. J. (1993) The continuum random tree III, *Ann. Probab.*, **21**, 248–289.
5. Aldous, D. J. (1997) Brownian excursions, critical random graphs and the multiplicative coalescent, *Ann. Probab.*, **25**, 812–854.
6. Aldous, D. J., Miermont, G. and Pitman, J. (in preparation) Limit distributions for non-uniform random mappings, involving Brownian bridge and inhomogeneous continuum random trees.
7. Aldous, D. J. and Pitman, J. (1994) Brownian bridge asymptotics for random mappings, *Random Structures Algorithms*, **5**, 487–512.
8. Aldous, D. J. and Pitman, J. (1999) A family of random trees with random edge lengths, *Random Structures Algorithms*, **15**, 176–195.
9. Aldous, D. J. and Pitman, J. (2000) Inhomogeneous continuum random trees and the entrance boundary of the additive coalescent, *Probab. Th. Rel. Fields*, **118**, 455–482.
10. Baron, G., Drmota, M. and Mutafchiev, L. (1996) Predecessors in random mappings, *Combin. Probab. Comput.*, **5**, no. 4 317–335.
11. Bennies, J. and Pitman, J. (2001) Asymptotics of the Hurwitz binomial distribution related to mixed Poisson Galton-Watson trees, *Combinatorics, Probability and Computing*, **10**, 203–211.
12. Berg, S. and Mutafchiev, L. (1990) Random mappings with an attracting center: Lagrangian distributions and a regression function, *J. Appl. Probab.*, **27**, 622–636.
13. Bertoin, J. and Pitman, J. (1994) Path transformations connecting Brownian bridge, excursion and meander, *Bull. Sci. Math. (2)*, **118** 147–166.
14. Biane, P., Pitman, J. and Yor, M. (2001) Probability laws related to the Jacobi theta and Riemann zeta functions, and Brownian excursions, *Bull. Amer. Math. Soc.*, **38**, 435–465.
15. Biane, P. and Yor, M. (1987) Valeurs principales associees aux temps locaux Browniens, *Bull. Sci. Math. (2)*, **111** 23–101.

16. Billingsley, P. (1999) *Convergence of Probability Measures*, Wiley, second edition.
17. Burtin, Y. D. (1980) On a simple formula for random mappings and its applications, *J. Appl. Probab.*, **17**, 403–414.
18. Camarri, M. and Pitman, J. (2000) Limit distributions and random trees derived from the birthday problem with unequal probabilities, *Electron. J. Probab.*, **5**, paper 2, 1–18.
19. Cremers, H. and Kadelka, D. (1986) On weak convergence of integral functions of stochastic processes with applications to processes taking paths in $L(p, E)$, *Stochastic Process. Appl.*, **21**, 305–317.
20. Drmota, M. and Gittenberger, B. (1997) On the profile of random trees, *Random Structures and Algorithms*, **10**, 421–451.
21. Duquesne, T. (2000) A limit theorem for the contour process of conditioned Galton-Watson trees, Unpublished.
22. Duquesne, T. and Le Gall, J.-F. (2001) Random trees, Lévy processes and spatial branching processes, Technical Report DMA-01-23, Ecole Normale Superieure.
23. Durrett, R. *Probability: theory and examples*, Duxbury Press, Belmont, CA, second edition, 1996.
24. Ethier, S. N. and Kurtz, T. G. (1986) *Markov Processes: Characterization and Convergence*, Wiley, New York.
25. Frieze, A. M. (1985) On the value of a random minimum spanning tree problem, *Discrete Appl. Math.*, **10**, 47–56.
26. Le Gall, J.-F. (1993) The uniform random tree in a Brownian excursion, *Probab. Th. Rel. Fields*, **96**, 369–384.
27. Gittenberger, B. and Louchard, G. (1999) The Brownian excursion multidimensional local time density, *J. Appl. Probab.*, **36**, 350–373.
28. Gittenberger, B. and Louchard, G. (2000) On the local time density of the reflecting Brownian bridge, *J. Appl. Math. Stochastic Anal.*, **13**, 125–136.
29. Grinblat, L. S. (1976) A limit theorem for measurable random processes and its applications, *Proc. Amer. Math. Soc.*, **61**, 371–376.
30. Harris, B. (1993) A survey of the early history of the theory of random mappings, in *Probabilistic methods in discrete mathematics (Petrozavodsk, 1992)*, VSP, Utrecht, 1–22.
31. Jansons, K. M. (1997) The distribution of the time spent by a standard excursion above a given level, with applications to ring polymers near a discontinuity potential, *Elect. Comm. in Probab.*, **2**, 53–58.
32. Jaworski, J. (1984) On a random mapping (T, P_j), *J. Appl. Probab.*, **21**, 186–191.
33. Kingman, J. F. C. (1978) The representation of partition structures, *J. London Math. Soc.*, **18**, 374–380.
34. Kolchin, V. F. (1986) *Random Mappings*, Optimization Software, New York, (Translation of Russian original).
35. Le Gall, J. F. and Le Jan, Y. (1998) Branching processes in Lévy processes: the exploration process, *Ann. Probab.*, **26**, 213–252.
36. Lévy, P. (1939) Sur certains processus stochastiques homogènes. *Compositio Math.*, **7**, 283–339.
37. Lyons, R. and Peres, Y. (in press) *Probability on Trees and Networks*, Cambridge University Press.
38. Marckert, J.-F. and Mokkadem, A. (2001) The depth first processes of Galton–Watson trees converge to the same Brownian excursion, Univ. de Versailles.
39. Mutafchiev, L. R. (1993) Random mappings and trees: relations between asymptotics of moments of certain characteristics, in *Probabilistic methods in discrete mathematics (Petrozavodsk, 1992)*, VSP, Utrecht, 353–365.
40. Mutafchiev, L. R. (1987) On random mappings with a single attracting centre, *J. Appl. Probab.*, **24**, 258–264.
41. Mutafciev, L. (1983) Probability distributions and asymptotics for some characteristics of random mappings, in W. Grossman et al. (Eds.), *Proc. 4th Pannonian Symp. on Math. Statist.*, 227–238.
42. O'Cinneide, C. A. and Pokrovskii, A. V. (2000) Nonuniform random transformations, *Ann. Appl. Probab.*, **10**, no. 4, 1151–1181.

43. Pitman, J. (1999) Brownian motion, bridge, excursion and meander characterized by sampling at independent uniform times, *Electron. J. Probab.*, **4**, paper 11, 1–33.
44. Pitman, J. (1999) The SDE solved by local times of a Brownian excursion or bridge derived from the height profile of a random tree or forest, *Ann. Probab.*, **27**, 261–283.
45. Pitman, J. (2001) Random mappings, forests and subsets associated with Abel–Cayley–Hurwitz multinomial expansions, *Séminaire Lotharingien de Combinatoire*, issue 46:45; `http://www.mat.univie.ac.at/~slc/`.
46. Pitman, J. and Yor, M. (1997) The two-parameter Poisson–Dirichlet distribution derived from a stable subordinator, *Ann. Probab.*, **25**, 855–900.
47. Ross, S. M. (1981) A random graph, *J. Appl. Probab.*, **18**, 309–315.
48. Stepanov, V. E. (1971) Random mappings with a single attracting center, *Theory Probab. Appl.*, **16**, 155–162.

Appendix

8. Appendix: Weak convergence on L_0

Recall L_0 is the space of measurable functions $f : [0,1] \to R$ with metric

$$||f_1 - f_2||_0 := \int_0^1 \min(1, |f_1(t) - f_2(t)|)\, dt.$$

Weak convergence of stochastic processes with respect to the L_0 metric has been discussed in several places [19, 29]; in particular it is known that convergence of finite dimensional distributions implies convergence in L_0. We need the following, apparently new, result which characterizes L_0 weak convergence as convergence of *random* finite dimensional distributions.

Proposition 29. *Let X and X_n, $n = 1, 2, \ldots$, be random elements of L_0. For each k let $0 < V_{k,1} < V_{k,2} < \ldots < V_{k,k} < 1$ be the order statistics of k independent uniform$(0,1)$ random variables. Then $X_n \xrightarrow{d} X$ in L_0 if and only if for each fixed k,*

$$(X_n(V_{k,1}), X_n(V_{k,2}), \ldots, X_n(V_{k,k})) \xrightarrow{d} (X(V_{k,1}), X(V_{k,2}), \ldots, X(V_{k,k}))$$

as $n \to \infty$.

The "only if" part is immediate. The proof of the "if" part occupies the rest of this Appendix. Fix B and set

$$F_B := \{f \in L_0 : \ \mathrm{esssup}_t |f(t)| \leq B\}.$$

On F_B the L_0 metric is equivalent to the L_1 metric

$$||f_1 - f_2||_1 := \int_0^1 |f_1(t) - f_2(t)|\, dt.$$

For $m \geq 1$ define $A_m : F_B \to F_B$ by averaging over dyadic rational intervals:

$$(A_m f)(t) = 2^m \int_{i2^{-m}}^{(i+1)2^{-m}} f(u)\, du, \quad i2^{-m} \leq t < (i+1)2^{-m}.$$

The following compactness criterion is straightforward.

Lemma 30. *A sequence $\{f_n, n \geq 1\} \subset F_B$ is relatively compact if and only if*

$$\lim_m \limsup_n ||f_n - A_m f_n||_1 = 0.$$

For $f \in F_B$ and $k \geq 1$ define a random element $\tilde{f}^k$ of F_B by

$$\tilde{f}^k(t) = f(V_{k,i}), \quad \tfrac{i-1}{k} \leq t < \tfrac{i}{k}.$$

Because a measurable function can be approximated by functions constant on intervals, it is straightforward to show

Lemma 31. $\lim_k E||\tilde{f}^k - f||_1 = 0, \quad f \in F_B.$

The technical heart of the argument is the following lemma, whose proof we defer.

Lemma 32. *Fix B and m. Then*

$$||f - A_m f||_1 \leq E||\tilde{f}^k - A_m \tilde{f}^k||_1 + \delta(k, B, m) \quad \forall f \in F_B, \ \forall k \geq 1$$

where the constants $\delta(k, B, m)$ satisfy $\lim_k \delta(k, B, m) = 0$.

To prove the Proposition, first truncate at $\pm B$ to reduce to the case where X_n and X take values in F_B; the general case then follows by letting $B \uparrow \infty$. By hypothesis,

$$\widetilde{X}_n^k \xrightarrow{d} \widetilde{X}^k \text{ as } n \to \infty, \ k \text{ fixed.}$$

Since A_m is a continuous function, we deduce

$$\lim_n E||\widetilde{X}_n^k - A_m \widetilde{X}_n^k||_1 = E||\widetilde{X}^k - A_m \widetilde{X}^k||_1. \tag{42}$$

Now

$$\begin{aligned} \limsup_n E||X_n - A_m X_n||_1 &\leq E||\widetilde{X}^k - A_m \widetilde{X}^k||_1 + \delta(k, B, m) \\ &\qquad \text{by Lemma 32 and (42)} \\ &\leq E||X - A_m X||_1 \\ &\qquad \text{letting } k \to \infty, \text{ using Lemma 31.} \end{aligned}$$

So

$$\lim_m \limsup_n E||X_n - A_m X_n||_1 \leq \lim_m E||X - A_m X||_1.$$

But the right side equals zero by applying Lemma 30 to X alone; and so Lemma 30 implies that the sequence $\{X_n, n \geq 1\}$ is tight in L_1. To prove convergence it therefore suffices to verify the *identifiability* result

$$\text{if } (X(V_{k,1}), X(V_{k,2}), \ldots, X(V_{k,k})) \stackrel{d}{=} (Y(V_{k,1}), Y(V_{k,2}), \ldots, Y(V_{k,k})), k \geq 1$$
$$\text{then } X \stackrel{d}{=} Y$$

and this is straightforward.

Proof of Lemma 32. The desired inequality can be split into two parts:

$$||f - A_m f||_1 \leq E||\tilde{f}^k - A_m f||_1 + \delta_1(k, B, m) \tag{43}$$

$$E||A_m f - A_m \tilde{f}^k||_1 \leq \delta_2(k, B, m) \tag{44}$$

where we want $\lim_k \delta(k, B, m) = 0$ in each case. We will prove these for $m = 1$, the general case being similar. Note $A_1 f$ is constant on the intervals $[0, 1/2)$ and $[1/2, 1]$. To study (43), the contribution to $E||\tilde{f}^k - A_1 f||_1$ from the interval $[0, 1/2)$ equals

$$E \int_0^{1/2} |\tilde{f}^k(t) - A_1 f(1/4)| \, dt.$$

By a Fubini argument, this equals

$$\int_0^1 |f(s) - A_1 f(1/4)| \, \rho_k(s) ds, \text{ for } \rho_k(s) ds = k^{-1} \sum_{i=1}^{k/2} P(V_{k,i} \in [s, s + ds]). \tag{45}$$

And we can write the contribution to $||f - A_1 f||_1$ from the interval $[0, 1/2)$ as

$$\int_0^1 |f(s) - A_1 f(1/4)| 1_{(s < 1/2)} \, ds. \tag{46}$$

Using the definition of the $V_{k,i}$ as uniform order statistics one can show $\rho_k(s) = P(V_{k-1,k/2} > s)$ and we lose nothing in assuming k is even. Now observe

$$\text{quantity (46)} - \text{quantity (45)} \leq \int_0^{1/2} |f(s) - A_1 f(1/4)| \; P(V_{k-1,k/2} < s) \, ds.$$

The integrand is bounded by $2B$, and combining with the symmetric contribution from $[1/2, 1]$ we see

$$||f - A_1 f||_1 - E||\tilde{f}^k - A_1 f||_1 \leq 2B \; E|V_{k-1,k/2} - 1/2|.$$

Since clearly $V_{k-1,k/2} \xrightarrow{d} 1/2$ as $k \to \infty$, we have a bound of the required form (43).

To argue (44), the value of $A_1 \tilde{f}^k$ on $[1/2, 1]$ equals

$$\tfrac{2}{k} \sum_{i=1}^{k/2} f(V_{k,i}) := \alpha_k, \text{ say.}$$

So the contribution to (44) from $[1/2, 1]$ can be bounded by

$$\left| \tfrac{1}{2} E\alpha_k - \int_0^{1/2} f(t) dt \right| + \tfrac{1}{2} E|\alpha_k - E\alpha_k|$$

$$\leq \left| \tfrac{1}{2} E\alpha_k - \int_0^{1/2} f(t) dt \right| + \tfrac{1}{2} \sqrt{\operatorname{var} \alpha_k}. \tag{47}$$

But

$$E\alpha_k = 2\int_0^1 f(s)\rho_k(s)\, ds \tag{48}$$

for $\rho_k(\cdot)$ as before. So the first term in (47) equals

$$\left|\int_0^1 f(s)(\rho_k(s) - 1_{(s\le 1/2)})\, ds\right| \le B\int_0^1 |\rho_k(s) - 1_{(s\le 1/2)}|\, ds = BE|V_{k-1,k/2} - \tfrac{1}{2}|$$

and this bound is of the required form. For the second term of (47), a brief calculation (whose details we omit) gives

$$E\alpha_k^2 = \tfrac{2}{k}E\alpha_k + \tfrac{4(k-1)}{k}\int_0^1\int_0^1 f(s_1)f(s_2)P(V_{k-2,k/2-1} > \max(s_1,s_2))\, ds_1 ds_2.$$

So using (48),

$$\begin{aligned}
\operatorname{var}\alpha_k &= E\alpha_k^2 - (E\alpha_k)^2\\
&= \tfrac{2}{k}E\alpha_k - \tfrac{4}{k}\int_0^1\int_0^1 f(s_1)f(s_2)P(V_{k-2,k/2-1} > \max(s_1,s_2))\, ds_1 ds_2\\
&\quad + 4\int_0^1\int_0^1 f(s_1)f(s_2)\big(P(V_{k-2,k/2-1} > \max(s_1,s_2)) - \rho_k(s_1)\rho_k(s_2)\big)\, ds_1 ds_2.
\end{aligned}$$

Because $|f| \le B$ we then see

$$\begin{aligned}
\operatorname{var}\alpha_k &\le \tfrac{2B}{k} + \tfrac{4B^2}{k}\\
&\quad + 4B^2\int_0^1\int_0^1 \big|P(V_{k-2,k/2-1} > \max(s_1,s_2)) - \rho_k(s_1)\rho_k(s_2)\big|\, ds_1 ds_2.
\end{aligned}$$

Since $V_{k-2,k/2-1} \xrightarrow{d} 1/2$ as $k\to\infty$, this bound $\to 0$ as $k\to\infty$, as required. $\square$

Part TWO

Integrable Models (of Statistical Physics and Quantum Field Theory)

RENORMALIZATION GROUP SOLUTION OF FERMIONIC DYSON MODEL

M. D. MISSAROV* (moukadas.missarov@ksu.ru)
Kazan State University
Department of Computational Mathematics and Cybernetics
Kazan
Russia

Abstract. In this paper we give a brief exposition of the fermionic version of Dyson's hierarchical model. Action of Wilson's renormalization group (RG) transformation in the space of coupling constants of hierarchical fermionic model is given by the rational map. Global RG-flow in the upper half-plane of the coupling constants is described. Complex behaviour of stable RG-invariant curves leads to the non-trivial picture of critical phenomena in this model. The continuum limit of the hierarchical fermionic model is given by p-adic fermionic model. The connection between coupling constants of p-adic model and its discretized hierarchical version is defined by discretization operator, which is given by non-trivial functional integral. This operator can be considered as a normalizing transformation for RG-map at trivial fixed point. The ultraviolet poles of Feynman amplitudes appear as a resonance values for normalizing transformation. Convergence of functional integral follows from Poincare and Siegel theorems.

1. Introduction

From mathematical point of view Euclidean quantum field theory can be considered as a theory of special random fields over d-dimensional Euclidean space R^d. By convention let us call these theories as models of (R^d, R)-type or (Z^d, R)-type, where Z^d denotes the lattice of integer-valued vectors. Rigorous analysis of these models is extremely complicated task and most of the theorems say only about existence or non-existence of these random fields [1], [2]. At the same time a lot of non-classical and formal mathematical instruments were invented for the computation of physical quantities: functional integrals, Feynman amplitudes, renormalization procedure, $(4 - d)$-expansion, etc [3].

Later it was discovered that if the field takes its values in the Grassmann algebra Γ (fermionic field), then mathematical structure of such model is simpler then usual bosonic one. The Gross–Neveu model in two dimensions was the first renormalizable model in which continuum limit problem was rigorously analyzed

* This work was supported in part by the Russian Foundation for Basic Research (Grants No.99-01-00467 and No.01-01-00610).

V. A. Malyshev and A. M. Vershik (eds.),
Asymptotic Combinatorics with Application to Mathematical Physics, 151–166.

by the methods of renormalization group [4], [5]. Let us call these models as a models of (R^d, Γ)-type.

The change in the domain of definition of the field is a following step. According to the Ostrovski theorem every norm over the field of rational numbers Q is equivalent to the usual absolute value or to p-adic norm for some prime number p. With d-dimensional p-adic space Q_p^d in place of R^d we are led to the models of (Q_p^d, R)-type. P-adic analogue of Z^d is so called hierarchical lattice of p-adic fractional vectors T_p^d. Models of (T_p^d, R)-type (but without any mention about its connection with p-adic numbers) were introduced in statistical physics by F. Dyson [6] and are called hierarchical Dyson's models. We emphasize that only in the framework of the hierarchical bosonic model P. M. Bleher and Ja. G. Sinai gave rigorous justification of the Wilson's approach to the critical phenomena [7], [8]. The suggestion to consider p-adic models in the string theory was proposed by I. V. Volovich [9] (see also [10–13]).

Playing in this game, it is natural to draw a diagram

$$\begin{array}{ccc} (R^d, R) & \longrightarrow & (R^d, \Gamma) \\ \downarrow & & \downarrow \\ (Q_p^d, R) & \longrightarrow & (Q_p^d, \Gamma) \end{array}$$

and to consider models of (T_p^d, Γ) and (Q_p^d, Γ) - types, which are placed in the right lower corner of the diagram. Combination of the Pauli principle with ultrametricity of p-adic distance leads us to the following drastic simplification of the mathematical structure. Problems of statistical physics and quantum field theory are found to be a problems of finite dimensional dynamical system and can be explained in terms of classical mathematics. Action of Wilson's renormalization group transformation in the space of coupling constants of hierarchical fermionic model is given by the rational map. All fixed points and some cycles of renormalization group (RG) map are described explicitly. Global behaviour of stable RG-invariant curves and RG-flow are visualized with the use of computer. The continuum limit of the hierarchical fermionic model is constructed rigorously and is given by p-adic fermionic model. The connection between coupling constants of p-adic model and its discretized hierarchical version is defined by discretization operator, which is given by non-trivial functional integral. On the other hand, this operator can be considered as a normalizing transformation for RG-map at trivial fixed point and therefore this functional integral can be found as a solution of 2-dimensional functional equation. The algorithm of solution of this equation automatically summarizes the subclasses of p-adic Feynman amplitudes. The Gaussian part of our model is given by the homogeneous kernel $|x-y|_p^{-\alpha}$, where $x, y \in Q_p^d$, $|\cdot|_p$ is a p-adic norm, $\alpha \in R$. The ultraviolet poles of Feynman amplitudes in α appear as a resonance values for normalizing transformation. Convergence of functional integral follows from Poincare and Siegel theorems. The Gaussian part, defined by p-adic version of Laplace operator corresponds to $\alpha = 2 + d$, which is resonance value. In this picture renormalization procedure can be interpreted as an inverse transformation to discretization operator.

Note, that some degenerate version of fermionic hierarchical model was proposed by T. Dorlas [14]. This version was described in non-Gibbsian form and therefore critical behavior was not investigated. Values of α and n were fixed and this caused some artefacts. For example, non-existence of thermodynamic limit at nontrivial fixed point, probably, is explained by low value of n in this model.

More detailed discussion of the Dyson hierarchical fermionic model and all proofs can be found in papers [15–22].

2. Renormalization group flow and critical phenomena

We'll begin with a brief consideration of some definitions and results.

Let $N = \{1, 2, \ldots\}$, $V_{k,s} = \{j : j \in N, (k-1)n^s < j \leq kn^s\}$, $k \in N, s \in N$, and let $s(i,j) = \min\{s :$ there is k such that $i \in V_{k,s}, j \in V_{k,s}\}$. The hierarchical distance $d(i,j)$, $j \in N$ is defined as $d(i,j) = n^{s(i,j)}$, if $i \neq j$ and $d(i,i) = 0$. Let us consider the 4–component fermionic field $\psi^*(i) = (\bar{\psi}_1(i), \psi_1(i), \bar{\psi}_2(i), \psi_2(i))$, $i \in N$, where the components are generators of a Grassmann algebra.

The "Gaussian" fermionic field with zero mean and binary correlation function

$$\langle \psi_k(i) \bar{\psi}_l(j) \rangle = \delta_{k,l} b(i,j), \ k,l = 1,2,$$

$$b(i,j) = \frac{1-n^{1-\alpha}}{1-n^{\alpha-2}} d^{\alpha-2}(i,j), \ i \neq j, \quad b(i,i) = \frac{1-n^{-1}}{1-n^{\alpha-2}}, \ \alpha \neq 2$$

is invariant under the renormalization group transformation, which is given as

$$(\psi^*)'(i) = (r_\alpha \psi^*)(i) = n^{-\alpha/2} \sum_{j \in V_{i,1}} \psi^*(j).$$

We denote this "Gaussian" state by $\rho(\alpha)$. Let us redenote $V_{1,N}$ by Λ_N and let A_N be the Grassmann subalgebra, generated by $4 \cdot n^N$ generators, corresponding to this volume. In [15] it was proved, that the restriction of the "Gaussian" state $\rho(\alpha)$ on the volume Λ_N can be represented in the Gibbsian form as

$$\rho(\alpha)\left(F(\psi^*)\right) = Z_{0,N}^{-1} \int F(\psi^*) \exp\{-H_{0,N}(\psi^*;\alpha)\}\, d\psi^*,$$

where the Beresin's anticommuting integration rules are used,

$$H_{0,N}(\bar{\psi},\psi;\alpha) = \sum_{i,j \in \Lambda_N} d_{0,N}(i,j) \bar{\psi}(i) \psi(j),$$

$$d_{0,N}(i,j) = d_0(i,j) - c(N), \quad d_0(i,j) = \frac{1-n^{\alpha-1}}{1-n^{-\alpha}} d^{-\alpha}(i,j), \quad i \neq j,$$

$$d_0(i,i) = \frac{1-n^{-1}}{1-n^{-\alpha}}, \quad c(N) = \frac{(1-n^{\alpha-1})^2}{(1-n^{-\alpha})(1-n^{-1})} n^{-\alpha(N+1)},$$

$$Z_{0,N} = \int \exp\{-H_{0,N}(\psi^*;\alpha)\} \prod_{i \in \Lambda_N} d\psi_1(i)\, d\bar{\psi}_1(i)\, d\psi_2(i)\, d\bar{\psi}_2(i), \ F(\psi^*) \in A_N.$$

The non-Gaussian Gibbs state ("expectation value") $\rho_N(r,g;\alpha)$ on the A_N is defined as

$$(\rho_N(r,g;\alpha))\,(F) = Z_N^{-1}(r,g;\alpha)\rho(\alpha)\,(F\exp\{-H_N\})\,,$$

$$H_N(\psi^*;r,g) = \sum_{i\in\Lambda_N} L(\psi^*(i);r,g), \quad Z_N(r,g;\alpha) = \rho(\alpha)\,(\exp\{-H_N\})\,,$$

$$L(\psi^*(i);r,g) = r(\bar\psi_1(i)\psi_1(i) + \bar\psi_2(i)\psi_2(i)) + g\bar\psi_1(i)\psi_1(i)\bar\psi_2(i)\psi_2(i).$$

If ρ is a state on A_N, then the renormalized state ρ' is defined on A_{N-1} by $\rho'(F) = \rho(F(r_\alpha\psi^*))$. In [15] we proved that $\rho'_N(r,g) = \rho_{N-1}(r',g')$, where

$$r' = n^{\alpha-1}\left(\frac{(r+1)^2-g}{(r+1)^2-g/n}(r+1)-1\right),\; g' = n^{2\alpha-3}\left(\frac{(r+1)^2-g}{(r+1)^2-g/n}\right)^2 g. \quad (1)$$

We'll use also RG-transformation in the space of non-normalized Grassmann-valued "densities" of single spin "distribution" $f(\psi^*;c_0,c_1,c_2) = c_0 + c_1(\bar\psi_1\psi_1 + \bar\psi_2\psi_2) + c_2\bar\psi_1\psi_1\bar\psi_2\psi_2$. Particularly, if $c_0 \neq 0$, we can write the density f in the exponential form $f(\psi^*;c_0,c_1,c_2) = c_0\exp\{-L(\psi^*;r(c),g(c))\}$, where $c = (c_0,c_1,c_2)$, $r(c) = -c_1/c_0$, $g(c) = (c_1^2 - c_0c_2)/c_0^2$. If $c_0 = 0$, as, for example, in the case of Grassmann δ-function $\delta(\psi^*) = \bar\psi_1\psi_1\bar\psi_2\psi_2$, the exponential representation is impossible. We'll denote

$$\rho_N(\alpha;c)(F) = Z_N^{-1}(\alpha;c)\int F\exp\{-H_{0,N}(\psi^*;\alpha)\prod_{i\in\Lambda_N} f(\psi^*(i);c)\,d\psi^*(i),$$

$$Z_N(\alpha;c) = \int \exp\{-H_{0,N}(\psi^*;\alpha)\prod_{i\in\Lambda_N} f(\psi^*(i);c)\,d\psi^*(i),$$

and call c-coordinates by projective coordinates, because $\rho_N(\alpha;c) = \rho_N(\alpha;\lambda c)$, where $\lambda \in R$. The RG-transformation in projective representation have the following form: $\rho'_N(\alpha;c) = \rho_{N-1}(\alpha;c')$, where

$$\begin{aligned} c_0' &= (c_1-c_0)^2 + \frac{1}{n}(c_0c_2 - c_1^2), \\ c_1' &= \lambda_1\left[(c_1-c_0)(c_2-c_1) + \frac{1}{n}(c_0c_2-c_1^2)\right], \\ c_2' &= \lambda_1^2\left[(c_2-c_1)^2 + \frac{1}{n}(c_0c_2-c_1^2)\right], \end{aligned} \quad (2)$$

$\lambda_1 = n^{\alpha-1}$. RG-transformation in (r,g)-space is well defined, if $g \neq n(r+1)^2$. The critical parabola $g = n(r+1)^2$ in c–space is described as a set of $c = (c_0,c_1,c_2)$, such that $c_0' = (c_1-c_0)^2 + (c_0c_2-c_1^2)/n = 0$. But RG-transformation in c–space is well defined in this case also.

We will denote RG-transformation in (r,g)- and c-coordinates, given by formulas (1) and (2) as $R(\alpha)$. The mapping $R(\alpha)$ is correctly defined as the mapping from two-dimensional projective space to itself everywhere except the point $(1,1,1)$

because $R(\alpha)(1,1,1) = (0,0,0)$. In the (r,g) plane this is the point $(-1,0)$, and we call this point the singular point of the RG mapping. RG-transformation in (r,g)–space have three finite fixed points, if $\alpha \neq 1$:

$$r_0(\alpha) \equiv 0, \quad g_0(\alpha) \equiv 0,$$

$$r_+(\alpha) = \frac{n^{1/2} - n^{\alpha-1}}{1 - n^{1/2}}, \quad g_+(\alpha) = \frac{r_+(\alpha)(1 + r_+(\alpha))^2}{1 + r_+(\alpha) + n^{-1/2}}, \ \alpha \neq \frac{1}{2},$$

$$r_-(\alpha) = \frac{-n^{1/2} - n^{\alpha-1}}{1 + n^{1/2}}, \quad g_-(\alpha) = \frac{r_-(\alpha)(1 + r_-(\alpha))^2}{1 + r_-(\alpha) - n^{-1/2}}.$$

For $\alpha = 1$ we have a whole line of fixed points $\{g = 0, r \neq -1\}$. In this paper we'll consider $\alpha > 1$. One can see, that Grassmann Fourier transform transposes the coefficients c_0, c_1, c_2 of the density $f(\bar{\eta}, \eta; c_0, c_1, c_2)$:

$$\begin{aligned} &F_{\eta^* \to \xi^*}(f(\eta^*; c_0, c_1, c_2)) \\ &\quad = \int \exp\{-(\bar{\xi}_1 \eta_1 + \bar{\xi}_2 \eta_2 + \xi_1 \bar{\eta}_1 + \xi_2 \bar{\eta}_2)\} f(\eta^*; c_0, c_1, c_2)\, d\eta_1\, d\bar{\eta}_1\, d\eta_2\, d\bar{\eta}_2 \\ &\qquad = f(\xi^*; c_2, c_1, c_0). \end{aligned}$$

We have also remarkable commutation relation $F\,R(\alpha) = R(2-\alpha)F$. Therefore, if we can describe RG-dynamics for $\alpha > 1$, with using this commutation relation we can describe RG-dynamics for $\alpha < 1$.

Besides three finite fixed points there is fixed point at infinity, which in c-space is described by the vector $(0,0,1)$, determining the Grassmann δ-function density $f(\psi^*; 0,0,1) = \bar{\psi}_1 \psi_1 \bar{\psi}_2 \psi_2$.

"+"-fixed point (FP) $(r_+(\alpha), g_+(\alpha))$ bifurcates from the trivial FP $(0,0)$ at $\alpha = 3/2$ and for $\alpha > 3/2$ belongs to the upper half-plane $\{(r,g) : g > 0\}$. For $\alpha < 3/2$ "+"-FP belongs to the lower half-plane and when $\alpha \to 1/2$ $\ r_+(\alpha) \to -(1 + n^{-1/2})$, $g_+(\alpha) \to -\infty$. In fact, "+"-FP bifurcates from δ–function FP at $\alpha = 1/2$. "−"-FP for all values of α belongs to the upper half-plane.

The stability analysis shows, that trivial FP is repelling FP for $\alpha \geq 3/2$ and saddle point for $1 < \alpha < 3/2$. Fixed point at infinity is attractive FP for $\alpha \geq 3/2$ and saddle point for $1 < \alpha < 3/2$. "+"-FP is saddle point for all values of $\alpha > 1$ and belongs to the upper half-plane $g > 0$ for $\alpha > 3/2$. It is interesting to note, that for sufficiently large n "+"-FP is unstable focus in some range of (α_1, α_2), $1 < \alpha_1 < \alpha_2 < 3/2$. "−"-FP is saddle point for all values of α and belongs to the upper half-plane.

Here we will discuss the RG-dynamics only in the upper half-plane $\{(r,g) : g > 0\}$ for $\alpha > 1$.

The general theory predicts the local existence of stable and unstable RG-invariant curves, passing through different fixed points. It seems essential to emphasize, that in our model we are able to describe the global behaviour of RG-invariant curves. In [17] we obtained the following results.

Theorem 1. *The part of stable RG-invariant curve γ_+, passing through the "+"-FP for $\alpha > 3/2$ is given by equation $g = h_+(r;\alpha)$, $0 \leq r < \infty$, where*

$h_+(r;\alpha)$ is a smooth monotone increasing function (in r). For $1 < \alpha \le 3/2$ the same equation defines stable RG-invariant curve for the trivial FP. The part of stable RG-invariant curve γ_-, passing through the "$-$"-FP for $\alpha > 1$ is given by the equation $g = h_-(r;\alpha)$, $-\infty < r < -1$, where $h_-(r;\alpha)$ is a smooth monotone decreasing function. Moreover $h_+(r;\alpha) \to 0$, when $r \to 0$ and $h_+(r;\alpha) \to \infty$, when $r \to \infty$, $h_-(r;\alpha) \to 0$, when $r \to -1$, $h_-(r;\alpha) \to \infty$, when $r \to -\infty$.

Let $\Omega_+ = \{(r,g) : 0 < r < \infty,\ 0 \le g < h_+(r;\alpha)\}$, $\Omega_- = \{(r,g) : -\infty < r < -1,\ 0 \le g < h_-(r;\alpha)\}$. In [16] we have proved, that domains Ω_+ and Ω_- are RG-invariant and obtained asymptotics of RG-iterations in these domains. Let $(r^{(N)}, g^{(N)}) = R^N(r,g)$. Asymptotic behaviour of RG-flow in the domains Ω_+ and Ω_- is described by the following

Theorem 2. *Let $\alpha > 1$ and $(r,g) \in \Omega_+ \cup \Omega_-$. Then there are constants $b_1(r,g)$ and $b_2(r,g)$ such that*

$$\lim_{N\to\infty} r^{(N)}\lambda_1^{-N} = b_1, \quad \lim_{N\to\infty} g^{(N)}\lambda_2^{-N} = b_2.$$

with $b_1 > 0$ for $(r,g) \in \Omega_+$, $b_1 < 0$ for $(r,g) \in \Omega_-$. Here $\lambda_1 = n^{\alpha-1}$, $\lambda_2 = n^{2\alpha-3}$ are eigenvalues of the differential of $R(\alpha)$ at trivial FP.

We say that thermodynamic limit in our model exists if all correlation functions have a limit when $N \to \infty$. Let $\tilde{\gamma}_+$, $\tilde{\gamma}_-$ denote stable RG-invariant curves for "+" and "$-$" fixed points and let

$$I_\pm(n) = (1 - \Delta_n^\pm, 1 + \Delta_n^\pm), \quad \Delta_n^\pm = \log_n\{a_\pm + (a_\pm^2 - 1)^{1/2}\},$$

$$a_\pm = \frac{(n-1)^2}{4n} \pm \frac{n+1}{2n^{1/2}}, \quad \alpha_\pm(n) = 2 - \log_n \frac{1 \pm 2n^{1/2}}{2 \pm n^{1/2}}.$$

Note that $I_-(n) = \emptyset$ if $n \le 13$. We exclude values $\alpha_\pm(n)$ because statistical sums of the model for "+" and "$-$" fixed points is equal to zero for corresponding values of α.

Theorem 3. *Let $\alpha \in I_+(n) \setminus \{1, \alpha_+(n)\}$ and $(r,g) \in \tilde{\gamma}_1$ or $\alpha \in I_-(n) \setminus \{1, \alpha_-(n)\}$ and $(r,g) \in \tilde{\gamma}_2$. Then hierarchical fermionic model, given by state $\rho_N(\alpha; r, g)$, has a thermodynamic limit.*

The restriction on α in this theorem looks artificial and is explained by technical peculiarities of the proof.

Theorem 4. *Let $\alpha > 1$ and $(r^{(N_0)}, g^{(N_0)}) \in \Omega_+ \cup \Omega_-$ for some $N_0 \ge 0$. Then model, given by state $\rho_N(r,g;\alpha)$, has thermodynamic limit.*

Critical phenomena in our model are related to the limiting behavior of the Grassmann-valued "density" of the properly normalized total spin in the volume Λ_N. Let

$$\psi^*_{N,a} = \frac{1}{n^{aN}} \sum_{i\in\Lambda_N} \psi^*(i),$$

$$\tilde{q}_N^{(a)}(x^*;r,g;\alpha) = \rho(r,g;\alpha)(\delta(\psi^*_{N,a} - x^*)),$$

where $\rho(r,g;\alpha) = \lim_{M\to\infty} \rho_M(r,g;\alpha)$, $x^* = (x_1, \bar{x}_1, x_2, \bar{x}_2)$ and $\delta(\psi^*) = \psi_1\bar{\psi}_1\psi_2\bar{\psi}_2$ is the Grassmann δ-function.

If the spin distribution density admits the exponential representation

$$p(x^*;r,g) = \frac{1}{r^2 - g}\exp\{-L(x^*;r,g)\},$$

then we will say that

$$\lim_{m\to\infty} p(x^*;r_m,g_m) = p(x^*;r,g),$$

if $\lim_{m\to\infty}(r_m,g_m) = (r,g)$.

Theorem 5. *Let $\alpha > 1$. If $(r^{(N_0)}, g^{(N_0)}) \in \Omega_+ \cup \Omega_-$ for some $N_0 \geq 0$, then*

$$\lim_{N\to\infty} q_N^{(1/2)}(x^*;r,g;\alpha) = p(x^*;b_1(r,g),0).$$

If $(r,g) \in \tilde{\gamma}_2$, then

$$\lim_{N\to\infty} q_N^{(\alpha/2)}(x^*;r,g;\alpha) = p_-(x^*;\alpha).$$

If $(r,g) \in \tilde{\gamma}_1$, then for $\alpha > 3/2$

$$\lim_{N\to\infty} q_N^{(\alpha/2)}(x^*;r,g;\alpha) = p_+(x^*;\alpha),$$

for $1 < \alpha \leq 3/2$

$$\lim_{N\to\infty} q_N^{(\alpha/2)}(x^*;r,g) = p(x^*;a_0(\alpha),0).$$

Here non-Gaussian densities $p_+(x^*;\alpha)$ and $p_-(x^*;\alpha)$ are determined by "+"- and "−"-FPs and are described explicitly [18]. The Gaussian density $p(x^*;a_0(\alpha),0)$ corresponds to the trivial Gaussian fixed point, $a_0 = (1-n^{\alpha-2})(1-n^{-1})^{-1}$.

Now let us discuss the RG-dynamics in the domain $\Omega = \{(r,g) : g > 0\} \setminus (\Omega_+ \cup \Omega_-)$ (the domain in the upper half–plane, bounded by the curve γ_+ on the right and by the curve γ_- on the left).

Here we use a computer graphics: every point of the upper half-plane is colored black (white), if after some finite number of RG-iterations this point maps into $\Omega_+(\Omega_-)$. The results of computer experiment in some finite part of (r,g)-plane are shown in the Fig. 1. To obtain the global picture we use the same algorithm in c–space (Fig. 2–4). We identify the projective c-space with hemisphere $S = \{(c_0,c_1,c_2) : c_0^2 + c_1^2 + c_2^2 = 1,\ c_0 \geq 0\}$, where the opposite points of the boundary circle $c_1^2 + c_2^2 = 1$ must be identified also. Moreover, to obtain the planar picture, we use the orthogonal projection of S onto the disk $S_1 = \{(c_1,c_2) : c_1^2 + c_2^2 \leq 1\}$. The grey colored set corresponds to the lower plane $\{(r,g) : g < 0\}$. The black and white zones appear as fractal sets, alternating one with other. We give the following informal classification of the boundary "black-white" points:

1) The point belongs to the "sharp" boundary, if there exists a smooth curve, passing through this point and such that every sufficiently small disk, centered at our point, is divided by the curve on the black and white halves.
2) The point belongs to "fuzzy black" boundary, if there exists a smooth curve, passing through this point and such that every sufficiently small disk, centered at our point, is divided by the curve on the black one-half and mixed one–half, composed from the infinite number of black and white "strips".
3) The point belongs to "fuzzy white" boundary, if there exists a smooth curve, passing through this point and such that every sufficiently small disk, centered at our point, is divided by the curve on the white one-half and mixed one-half, composed from the infinite number of black and white "strips".

The "sharp" and "fuzzy black" boundaries are clearly observable on the figures, but the existence of the "fuzzy white" boundary can be shown after the series of successive magnifications.

From the computer experiments it follows, that:
1) The points of the "sharp" boundary generate stable RG-invariant curve for the FP at infinity (on the disk $c_1^2 + c_2^2 \leq 1$ this FP has coordinates $(0,1) \equiv (0,-1)$).
2) The points of the "fuzzy black" boundary generate stable RG-invariant curve for the "+"-FP, if $\alpha > 3/2$, and stable RG-invariant curve for the trivial FP, if $1 < \alpha \leq 3/2$.
3) The points of the "fuzzy white" boundary generate stable RG-invariant curve for the "$-$"-FP.

From the fact that the opposite points of the circle $c_1^2+c_2^2 = 1$ are identified one can see, that all above–mentioned stable RG-invariant curves are smooth connected (in c–space) curves. These curves selfinteract at the point $(1/\sqrt{3}, 1/\sqrt{3})$ and its RG–inverse images (in (r, g)–space this series of points is described by a sequence $(-\lambda_1^{-k}, 0)$, $k = 0, 1, \ldots, \lambda_1 = n^{\alpha-1}$). Rigorously speaking, we must exclude this series of points from our consideration, because RG–transformation is not defined at the point $(1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})$ in c–space.

All these RG-invariant curves make an infinite number of revolutions in the domain Ω, alternating one with other. This intricate behaviour of RG-flow can be interpreted as an nontrivial critical phenomena in our fermionic statistical system. Theorem 4 and results of computing give the following picture of the critical behavior. We fix the value of $r = r_0$. For definiteness, we set $r_0 > 0$, $\alpha > 3/2$. We consider the half-line (r_0, g), $g > 0$, and let (r_0, g_{cr}) denote the point where this half-line crosses the curve γ_+. Then for $0 < g < g_{cr}$, the total normalized spin in the volume Λ_N with normalization exponent $a = \frac{1}{2}$ has the "Gaussian" distribution with the "variance"

$$\chi(r_0, g) = \int (x_1\bar{x}_1 + \bar{x}_2 x_2) p(x^*; b_1(r_0, g), 0)\, dx^* = \frac{2}{b_1(r, g)}.$$

in the limit $N \to \infty$. We note that $\chi(r, g)$ can also be written as

$$\chi(r, g) = \lim_{N\to\infty} \frac{1}{n^N} \rho(r, g; \alpha) \Big[\sum_{i\in\Lambda_N} \psi_1(i) \sum_{j\in\Lambda_N} \bar{\psi}_1(j) + \sum_{i\in\Lambda_N} \psi_2(i) \sum_{j\in\Lambda_N} \bar{\psi}_2(j) \Big],$$

and this quantity is therefore an anticommutative analogue of the susceptibility.

For $g = g_{cr}$, the total spin density has a nontrivial limit for the normalization exponent $a = \frac{\alpha}{2}$, and this limit is determined by the "+" "non-Gaussian" FP. For large enough g the total spin with the normalization exponent $a = \frac{1}{2}$ has the limiting "Gaussian" distribution with the "variance" $\chi(r_0, g) = \frac{2}{b_1(r_0,g)}$, but in contrast to the case $g < g_{cr}$, this "variance" is negative.

In the interval from g_{cr} to $(r_0 + 1)^2$, computer simulations show that we jump from the attraction domain of the invariant set Ω_+ to the attraction domain of the invariant set Ω_- and back infinitely many times while crossing the stable invariant curves for nontrivial FPs infinitely many times. In this respect, the critical behavior differs from the critical behavior in bosonic models where only one critical temperature exists (in our model, g is an analogue of the reciprocal temperature).

Computer experiments show that if g approaches $\tilde{\gamma}_+$ from below, the variance $\chi(r_0, g)$ tends to $+\infty$, and if g approaches $\tilde{\gamma}_-$ from above, the variance $\chi(r_0, g)$ tends to $-\infty$; if g crosses the stable RG-invariant curve for the FP at infinity, then $\chi(r_0, g) = 0$. The graph of the function $\chi(r_0, g)$ has a complex form with infinitely many poles. The divergence degree (critical exponent) of $\chi(r_0, g)$ at the poles is determined by the spectrum of the RG–differential at the corresponding FP.

3. Continuum limit in the Dyson fermionic model

Let p be a fixed prime number, Q_p be the p–adic number field and $|\cdot|_p$ be the p–adic norm on Q_p. Every p–adic number x can be represented in the form

$$x = c_{-n}p^{-n} + \cdots + c_{-1}p^{-1} + c_0 + c_1 p + \ldots,$$

where the coefficients c are integer numbers from 0 to $(p - 1)$ and $c_n > 0$, n is some integer number. Then $|x|_p = p^n$ and the fractional part of x is defined as

$$\{x\} = c_{-n}p^{-n} + \cdots + c_{-1}p^{-1}.$$

For $x = (x_1, \ldots, x_d) \in Q_p^d$, we set $|x|_p = \max_i |x_i|_p$, $\{x\} = (\{x_1\}, \ldots, \{x_d\})$. Then the discrete set $T_p^d = \{x \in Q_p^d : x = \{x\}\}$ can be viewed as a hierarchical lattice with the elementary cell size $n = p^d$ and with the hierarchical distance $d(i, j) = |i - j|_p$, $i, j \in T_p^d$.

We consider a four–component fermionic field

$$\psi^*(x) = \left(\bar{\psi}_1(x), \psi_1(x), \bar{\psi}_2(x), \psi_2(x)\right), \ x \in Q_p^d.$$

Scaling transformations group is defined by $\left(S_\lambda(\alpha)\psi^*\right)(x) = |\lambda|^{(1-\alpha/2)d}\psi^*(\lambda x)$, where $\lambda \in Q_p$ and α is a real parameter. The Gaussian scaling-invariant fermionic field is defined by the Hamiltonian

$$H_0(\psi^*; \alpha) = c(\alpha) \int |x - y|^{-\alpha d} \left(\bar{\psi}_1(x)\psi_1(y) + \bar{\psi}_2(x)\psi_2(y)\right) dx\, dy, \qquad (3)$$

where x and y are d-dimensional p-adic arguments, dx is the corresponding Haar measure, $c(\alpha)$ is a normalization constant, $c(\alpha) = f_n(\alpha)(f_n(1 - \alpha))^{-1}$, $f_n(\alpha) =$

$(1-n^{-\alpha})^{-1}$, $n=p^d$. The value $\alpha = 1+2/d$ gives rise to a p-adic analogue of the Hamiltonian that is given by the Laplace operator in the real case. We note that in contrast to the real case, the p-adic version with $\alpha = 1+2/d$ describes a model with long-range interaction.

We define the discretization of ψ^* adapted to the lattice T_p^d as the field ξ^* given by

$$\xi^*(j) = \int \psi^*(j+x)\chi(x)\,dx, \quad j \in T_p^d,$$

where $\chi(x)$ is the characteristic function of the ball $Z_p^d = \{x : |x|_p \leq 1\}$. The discretization of the transformation $S_\lambda(\alpha)$, $\lambda = p^{-1} \in Q_p$, is then the standard hierarchical block-spin renormalization group (RG) transformation

$$\left(r(\alpha)\xi^*\right)(j) = n^{-\alpha/2} \sum_{i \in B(j)} \xi^*(j), \qquad B(j) = \{i \in T_p^d : |pi - j| \leq 1\}.$$

The discretization of the Gaussian field with Hamiltonian (3) is determined by the Hamiltonian

$$H_0'(\xi^*;\alpha) = \sum_{i,j} h(i,j;\alpha)\left(\bar{\xi}_1(i)\xi_1(j) + \bar{\xi}_2(i)\xi_2(j)\right),$$

where

$$h(i,j;\alpha) = \frac{f_n(\alpha)}{f_n(1-\alpha)}(1-\delta_{ij})|i-j|^{-\alpha d} + \frac{f_n(\alpha)}{f_n(1)}\delta_{ij}.$$

We now consider the formal fermionic field with the Hamiltonian

$$H(\psi^*;\alpha;r,g) = H_0(\psi^*;\alpha) + \int L(\psi^*(x);r,g)\,dx, \tag{4}$$

where

$$L(\psi^*(x);r,g) = r\left(\bar{\psi}_1(x)\psi_1(x) + \bar{\psi}_2(x)\psi_2(x)\right) + g\bar{\psi}_1(x)\psi_1(x)\bar{\psi}_2(x)\psi_2(x).$$

The scaling transformation $S_\lambda(\alpha)$, $\lambda = p^{-1} \in Q_p$, acts linearly in the space of Hamiltonians (4): $H(\psi^*;\alpha;r,g) \to H(\psi^*;\alpha;n^{\alpha-1}r,n^{2\alpha-3}g)$. The linear transformation $(r,g) \to (n^{\alpha-1}r, n^{2\alpha-3}g)$ in the coupling-constant plane we denote by $S(\alpha)$.

The discretization of the field $\psi^*(x)$ with Hamiltonian (4) is determined by the Hamiltonian

$$\begin{aligned} &H'(\xi^*;\alpha;r,g) \\ &\quad = H_0'(\xi^*;\alpha) - \sum_j \log\left\langle \exp\left\{-\int L(\xi^*(j)+\eta^*(x);r,g)\,dx\right\}\right\rangle_{\mu(d\eta^*)}, \end{aligned}$$

where the average is taken with respect to the Gaussian field $\eta^*(x)$ that has a support contained in the ball Z_p^d, a zero mean, and the binary correlation function

$$\begin{aligned} \langle \bar{\eta}_k(x)\eta_l(y)\rangle &= \delta_{kj}\left(\frac{f_n(2-\alpha)}{f_n(\alpha-1)}|x-y|^{(\alpha-2)d} - \frac{f_n(2-\alpha)}{f_n(1)}\right), \\ \langle \bar{\eta}_k(x)\bar{\eta}_l(y)\rangle &= \langle \eta_k(x)\eta_l(y)\rangle = 0, \qquad x,y \in Z_p^d, \quad k,l = 1,2. \end{aligned}$$

The path integral

$$\log \left\langle \exp \left\{ - \int L(\xi^*(i) + \eta^*(x); r, g)\, dx \right\} \right\rangle_{\mu(d\eta^*)}$$

can be represented as $L(\xi^*(i); u(r,g), v(r,g))$ (see [20]), where the mapping $(r,g) \to (u(r,g), v(r,g))$, denoted by $P(\alpha)$, is a normalizing transformation to the RG mapping $R(\alpha)$ at the origin:

$$R(\alpha)P(\alpha) = P(\alpha)S(\alpha).$$

As it was shown in the section 2, the RG–transformation action in coupling–constant space (r,g) can explicitly determined and is given by the formulas (1). The mapping $S(\alpha)$ is given by the diagonal matrix whose eigenvalues are the eigenvalues of the differential of $R(\alpha)$ at the origin.

For $\alpha > 3/2$, the eigenvalues satisfy $\lambda_1(\alpha) = n^{\alpha-1} > 1$ and $\lambda_2(\alpha) = n^{2\alpha-3} > 1$, and we are in the domain where the classical Poincare theorem [21] applies. According to this theorem, the mapping $P(\alpha)$ can be expanded in a power series in r and g that converges for sufficiently small r and g provided α is a non-resonance value. We recall that α is called a non-resonance value if

$$\lambda_1(\alpha) \neq \lambda_1^{m_1}(\alpha)\lambda_2^{m_2}(\alpha), \qquad \lambda_2(\alpha) \neq \lambda_1^{m_1}(\alpha)\lambda_2^{m_2}(\alpha),$$

where m_1 and m_2 are nonnegative integers such that $m_1 + m_2 \geq 2$. The resonance values in the domain $3/2 < \alpha < 2$ are arranged in the discrete series $\alpha_k = 3/2 + (2(2k-1))^{-1}$, $k = 1, 2, \ldots$ If $1 \leq \alpha \leq 3/2$ and $\lambda_1(\alpha) > 1 > \lambda_2(\alpha)$, we are in the so–called Siegel domain. In that case any rational α is a resonance and the convergence of the mapping $P(\alpha)$ requires, in addition to the non-resonance condition, that the pair $\lambda_1(\alpha), \lambda_2(\alpha)$ be a (C, v)–type set [21], i.e.,

$$|\lambda_1 - \lambda_1^{m_1}\lambda_2^{m_2}| \geq C|m|^{-v}, \qquad |\lambda_2 - \lambda_1^{m_1}\lambda_2^{m_2}| \geq C|m|^{-v},$$

$m = m_1 + m_2 \geq 2,\ m_1 \geq 0, m_2 \geq 0$.

We now turn to the problem of rigorous construction of the continuum limit in this model. Let us consider the sequence of finer and finer hierarchical lattices $T^d_{p,m} = p^m T^d_p$, $m = 0, 1, \ldots$. We introduce the sequence of fields ζ^*_m defined on $T^d_{p,m}$ by the Hamiltonians

$$H'_m(\zeta^*_m; \alpha; \lambda_1^m r^{(-m)}, \lambda_2^m g^{(-m)}) = H'_{m,0}(\zeta^*_m; \alpha) + \\ + \sum_{i \in T^d_{p,m}} p^{-md} L(\zeta^*_m(i); \lambda_1^m r^{(-m)}, \lambda_2^m g^{(-m)}),$$

where $(r^{(-m)}, g^{-m)}) = R^{-m}(\alpha)(r,g)$ and the Gaussian part is

$$H'_{m,0}(\zeta^*_m; \alpha) = \sum_{i,j \in T^d_{p,m}, i \neq j} |i-j|^{-\alpha d} p^{-2md} \left(\bar{\zeta}_{m,1}(i)\zeta_{m,1}(j) + \bar{\zeta}_{m,2}(i)\zeta_{m,2}(j) \right) \\ + \sum_{i \in T^d_{p,m}} p^{md(\alpha-2)} \frac{f_n(\alpha)}{f_n(1)} \left(\bar{\zeta}_{m,1}(i)\zeta_{m,1}(i) + \bar{\zeta}_{m,2}(i)\zeta_{m,2}(i) \right).$$

Note, that inverse map $R^{-1}(\alpha)$ is well defined for almost all (r, g).

We recall that a natural space of test functions on Q_p^d is the space of locally constant finite functions $D(Q_p^d)$. The continuum field ψ^* on Q_p^d is given by the correlation functions

$$\langle(\psi_{k_1}(f_1)\bar{\psi}_{l_1}(g_1)\dots\psi_{k_s}(f_s)\bar{\psi}_{l_s}(g_s)\rangle \tag{5}$$

where $f_1, g_1, \dots, f_s, g_s \in D(Q_p^d)$, $k_i, l_i = 1, 2$ and $i = 1, \dots, s$. Because the functions $f_1, \dots, g_s$ can be represented as finite sums of the characteristic functions of p–adic balls with sufficiently small diameters, correlation functions (5) are defined whenever the correlation functions are defined for all fields ζ_m^*. If thermodynamic limit exists for the field $\xi^* \equiv \zeta_0^*$ determined by the Hamiltonian $H'(\xi_0^*; \alpha; r, g)$ on the lattice T_p^d, then the same is true for the field ζ_m^* determined by the Hamiltonian $H'(\zeta_m^*; \alpha; \lambda_1^m r^{(-m)}, \lambda_2^m g^{(-m)})$. We have shown that the existence domain of the thermodynamic limit in the (r, g)-half-plane comprises a full-measure set in R^2 if $\alpha > 1$. It means that for almost all (r, g) the correlation functions in the hierarchical fermionic model are well defined. Therefore, once the problem of constructing the continuum field is reduced to that of constructing the correlation functions, we have the following:

Theorem 6. *For all points of the coupling–constant upper half-plane (r, g) where the thermodynamic limit exists in the fermionic hierarchical model, there exists a continuum limit.*

Because the Hamiltonians $H_m'(\zeta_m^*; \alpha; \lambda_1^m r^{(-m)}, \lambda_2^m g^{(-m)})$ can be considered as a discrete approximations of the continuum field Hamiltonian

$$H(\psi^*; \alpha; \lambda_1^m r^{(-m)}, \lambda_2^m g^{(-m)}),$$

a natural question arises whether the limit

$$\lim_{m\to\infty}(\lambda_1^m r^{(-m)}, \lambda_2^m g^{(-m)}) = \lim_{m\to\infty} S^m(\alpha)R^{-m}(\alpha)(r, g)$$

exists and if so, how it is related to the coupling constants (r, g) of the inital discrete field. Letting U denote the set of all points (r, g) that tend to the trivial fixed point $(0, 0)$ under iterations of the inverse mapping R^{-1}, we now prove the following

Theorem 7. *Let $\alpha > 2$. Then for all $(r, g) \in U$, the limit*

$$\lim_{m\to\infty} S^m(\alpha)R^{-m}(\alpha)(r, g) = T(\alpha; r, g)$$

exists, and the mapping $T(\alpha)$ satisfies the commutation relation $T(\alpha)R(\alpha) = S(\alpha)T(\alpha)$.

It is easy to see that $T(\alpha)$ is the inverse of $P(\alpha)$, $T(\alpha) = P^{-1}(\alpha)$. As already noted, the mapping $P(\alpha)$ is defined for non-resonance values of $3/2 < \alpha < 2$ and the same is true for $T(\alpha)$.

References

1. Glimm, J. and Jaffe, A. (1981) *Quantum physics: a function integral point of view*, Springer–Verlag, Berlin.
2. Rivasseau, V. (1991) *From perturbative to constructive renormalization*, Princeton University Press.
3. Zinn–Justin, J. (1996) *Quantum field theory and critical phenomena*, Clarendon Press, Oxford.
4. Gawedzki, K. and Kupiainen, A. (1985) *Commun. Math. Phys.*, **102**, p. 1.
5. Feldman, J., Magnen, J., Rivasseau, V. and Seneor, R. (1986) *Commun. Math. Phys.*, **103**, p. 67.
6. Dyson, F. J. (1969) *Commun. Math. Phys.*, **12**, p. 91.
7. Bleher, P. M. and Sinai, Ya. G. (1973) *Commun. Math. Phys.*, **33**, p. 23.
8. Bleher, P. M. and Major, P. (1987) *Ann. Prob.*, **15**, 431–477.
9. Volovich, V. I. (1987) *Class. Quantum Grav.*, **4**, L83–L87.
10. Brekke, L. and Freund, P. G. O. (1993) *Phys. Rep.*, **233**, 2–66.
11. Vladimirov, V. S., Volovich, I. V. and Zelenov, E. I. (1994) *p-Adic analysis and mathematical physics*, World Scientific, Singapore.
12. Lerner, E. Yu. and Missarov, M. D. (1989) *Teor. Mat. Fiz.*, **78**, 248–257 (in Russian).
13. Lerner, E. Yu. and Missarov, M. D. (1989) *Commun. Math. Phys.*, **121**, 35–48.
14. Dorlas, T. C. (1991) *Commun. Math. Phys.*, **136**, p. 169.
15. Lerner, E. Yu. and Missarov, M. D. (1994) *Journal of Stat. Phys.*, **76**, p. 805.
16. Lerner, E. Yu. and Missarov, M. D. (1996) *Teor. Mat. Fiz.*, **107**, 201–212 (in Russian).
17. Missarov, M. D. (1998) *Teor. Mat. Fiz.*, **114**, 323–336 (in Russian).
18. Missarov, M. D. (1998) *Teor. Mat. Fiz.*, **117**, 471–488 (in Russian).
19. Missarov, M. D. (1999) *Teor. Mat. Fiz.*, **118**, 40–50 (in Russian).
20. Missarov, M. D. (1994) *Lett. in Math. Phys.*, **32**, 347–356.
21. Arnold, V. I. (1983) *Geometrical Methods in the Theory of Ordinary Differential Equations*, Springer, Berlin.

Figures

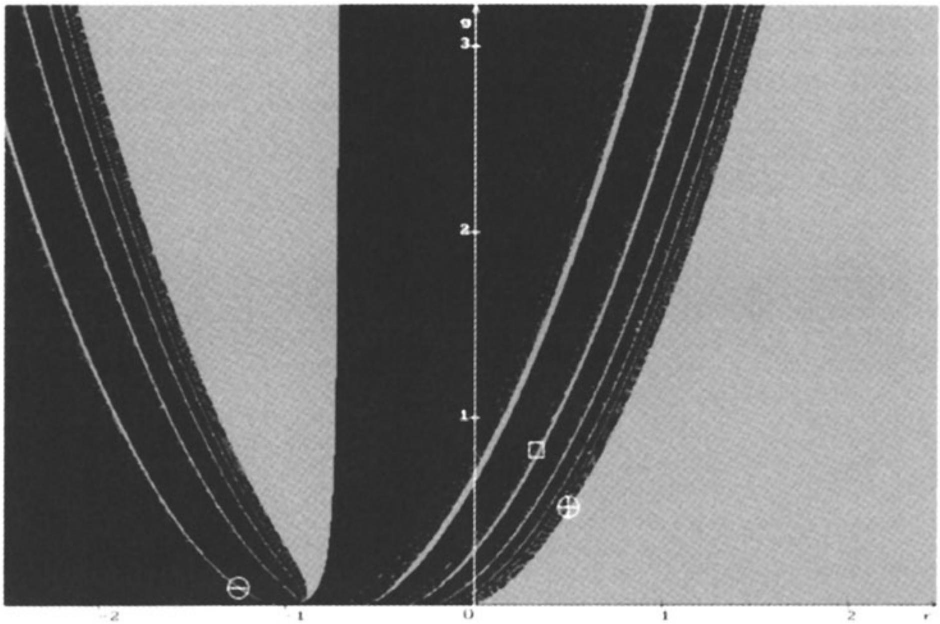


Figure 1. RG-flow in the upper half-plane $\{(r, g) : g > 0\}$, $n = 2$, $\alpha = 1.7$. RG-iterations of the points from "white" ("black") zones go to the right (left), $+$ $(-)$ indicates the location of "$+$" ("$-$") fixed point.

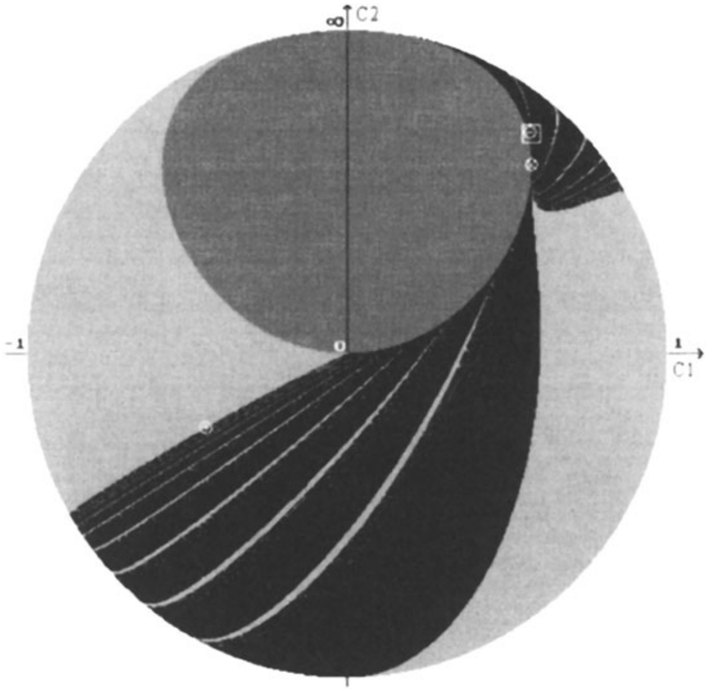


Figure 2. RG-flow in (c_1, c_2)–coordinates, $n = 2$, $\alpha = 1.7$. Signs $+$, $-$, 0, 1 indicate the locations of "+", "−", trivial fixed points and fixed point at infinity correspondingly, $\times$ indicates the location of singular point $(1/\sqrt{3}, 1/\sqrt{3})$. The grey-colored domain corresponds to the lower half-plane in (r, g)-coordinates.

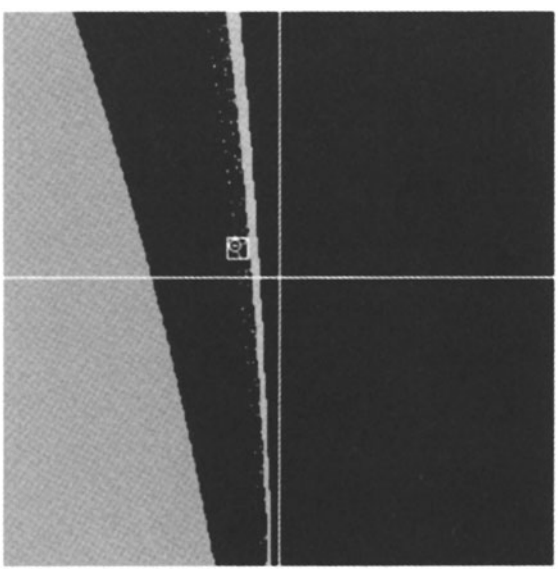


Figure 3. Enlarged picture of the square, selected on the fig. 2.

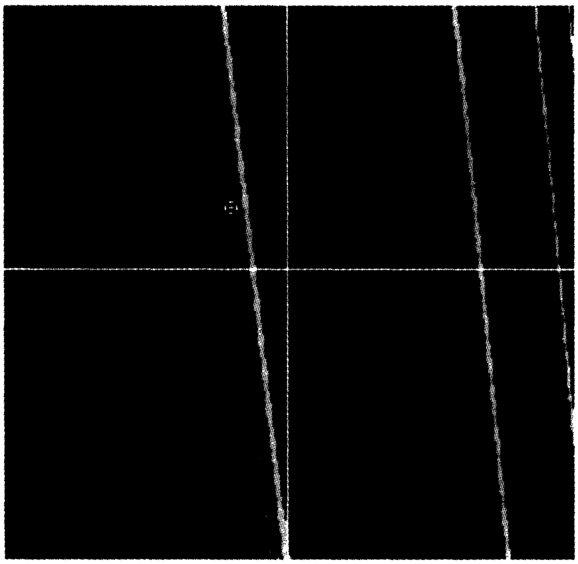


Figure 4. Enlarged picture of the square, selected on the fig. 3. "−"-fixed point lies on the "fuzzy black" boundary.

STATISTICAL MECHANICS AND NUMBER THEORY

H. E. BOOS (boos@mx.ihep.su)
Institute for High Energy Physics
Protvino, 142284, Russia

V. E. KOREPIN (korepin@insti.physics.sunysb.edu)
C.N. Yang Institute for Theoretical Physics
State University of New York at Stony Brook
Stony Brook, NY 11794-3840, USA

Abstract. Riemann zeta function is an important object of number theory. It was also used for description of disordered systems in statistical mechanics. We show that Riemann zeta function is also useful for the description of correlation functions in quantum integrable models.

1. Introduction

Riemann zeta function for $Re(s) > 1$ can be defined as follows:

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}. \tag{1}$$

It also can be represented as a product with respect to all prime numbers p:

$$\zeta(s) = \prod_p (1 - p^{-s})^{-1}. \tag{2}$$

It can be analytically continued in the whole complex plane of s. It has only one pole, at $s = 1$. Riemann zeta function satisfies a functional equation

$$\zeta(s) = 2^s \pi^{s-1} \sin(s\pi/2)\Gamma(1-s)\zeta(1-s). \tag{3}$$

It has 'trivial' zeros at $s = -2n$ ($n > 1$ is an integer). The famous Riemann hypothesis [15] states that nontrivial zeros belong to the straight line $Re(s) = 1/2$. Recently Montgomery and Odlyzko conjectured that for large values of imaginary part of s the distribution of zeros can be described by GUE of random matrices. Forrester and Odlyzko related the problem of distribution of zeros to Painleve differential equation and integrable integral operators [21]. Riemann zeta function is useful for study of distribution of prime numbers on the real axis [14]. The values of Riemann zeta function at special points were studied in [17], [18]. At

V. A. Malyshev and A. M. Vershik (eds.),
Asymptotic Combinatorics with Application to Mathematical Physics, 167–208.

even values of its argument zeta function can be expressed in terms of powers of π and Bernoulli's numbers

$$\zeta(2n) = (-1)^{n+1} 2^{2n-1} \pi^{2n} B_{2n}/(2n)!. \tag{4}$$

The values of Riemann zeta function at odd arguments provide infinitely many different irrational numbers [16]. Riemann zeta function plays an important role, not only in pure mathematics but also theoretical physics. Some Feynman diagrams in quantum field theory can be expressed in terms of $\zeta(n)$, see, for example, [1]. In statistical mechanics Riemann zeta function was used for the description of chaotic systems [19]. One can find more information and citation on the following web-cite `http://www.maths.ex.ac.uk/~mwatkins/`.

We argue that $\zeta(n)$ is also important for exactly solvable models. The most famous integrable models is the Heisenberg XXX spin chain. This model was first suggested by Heisenberg [3] in 1928 and solved by Bethe [4] in 1931. Since that time it found multiple applications in solid state physics and statistical mechanics.

The Hamiltonian of the XXX spin chain can be written like this

$$H = \sum_{i=1}^{N} (\sigma_i^x \sigma_{i+1}^x \ + \ \sigma_i^y \sigma_{i+1}^y \ + \ \sigma_i^z \sigma_{i+1}^z \ - 1 \,). \tag{5}$$

Here N is the length of the lattice and $\sigma_i^x, \sigma_i^y, \sigma_i^z$ are Pauli matrices. We consider the thermodynamical limit, when N goes to infinity. The sign in front of the Hamiltonian indicates that we are considering the anti-ferromagnetic case. We consider periodic boundary conditions. Notice that this Hamiltonian annihilates the ferromagnetic state [all spins up].

The construction of the anti-ferromagnetic ground state wave function $|AFM\rangle$ can be credited to Hulthén [5]. An important correlation function was defined in [8]. It was called the emptiness formation probability

$$P(n) = \langle AFM | \prod_{j=1}^{n} P_j | AFM \rangle$$

where $P_j = (1 + \sigma_j^z)/2$ is a projector on the state with spin up in jth lattice site. Averaging is over the anti-ferromagnetic ground state. It describes the probability of formation of a ferromagnetic string of the length n in the anti-ferromagnetic background $|AFM\rangle$. In this paper we shall first study short strings (n is small), in the end we shall discuss long distance asymptotic (at finite temperature). The four first values of the emptiness-formation probability look as follows:

$$P(1) = \frac{1}{2} = 0.5, \tag{6}$$

$$P(2) = \frac{1}{3}(1 - \ln 2) = 0.102284273, \tag{7}$$

$$P(3) = \frac{1}{4} - \ln 2 + \frac{3}{8}\,\zeta(3) = 0.007624158, \tag{8}$$

$$P(4) = \frac{1}{5} - 2\ln 2 + \frac{173}{60}\,\zeta(3) - \frac{11}{6}\,\zeta(3)\,\ln 2 - \frac{51}{80}\,\zeta^2(3) \\ -\frac{55}{24}\,\zeta(5) + \frac{85}{24}\,\zeta(5)\,\ln 2 = 0.000206270\,. \tag{9}$$

Let us comment. The value of $P(1)$ is evident from the symmetry, $P(2)$ can be extracted from the explicit expression of the ground state energy [5]. $P(3)$ can be extracted from the results of M.Takahashi [9] on the calculation of the nearest neighbor correlation. It was confirmed in paper [10]. One should also mention independent calculation of $P(3)$ in [11]. One can express $P(3)$ in terms of next to the nearest neighbor correlation

$$\langle\, S_i^z S_{i+2}^z \,\rangle \;=\; 2\,P(3) \;-\; 2\,P(2) \;+\; \frac{1}{2}\,P(1). \tag{10}$$

The calculation of $P(3)$ and $P(4)$ is discussed in this paper.

The expression above for $P(4)$ is our main result here.

We briefly announced our results in [2], here we provide the detailed derivation. The plan of the paper is as follows. In the next section we discuss a general procedure of the calculation of $P(n)$. We also show how this scheme works for $P(2)$. In Appendices A and B we describe in detail the calculation of $P(3)$ and $P(4)$ respectively by means of the technique elaborated in Section 2. The main results are summarized in the conclusion.

2. General discussion of the calculation of $P(n)$

There are several different approaches to investigate $P(n)$:

- **Representation of correlation functions as determinants of Fredholm integral operators.** This approach is based on following steps:

  i. Quantum correlation function should be represented as a determinant of a Fredholm integral operators of a special type. We call these operators *integrable* integral operators.

  ii. The determinant can be described by completely integrable equation of Painleve type.

  iii. Asymptotic of correlation function [and the determinant] can be described by Riemann–Hilbert problem.

  This approach was discovered in [23], it is described in detail in the book [6]. It is interesting to mention that this approach was successfully applied also to matrix models [22].

- **Vertex operator approach** was developed in Kyoto by Foda, Jimbo, Miki, Miwa and Nakayashiki. This approach is based on study of representations of infinite dimensional quantum group $U_q\widehat{SL(2)}$, see [7].

We shall use the integral representation obtained in [8]:

$$P(n)=\int_C\frac{d\lambda_1}{2\pi i\lambda_1}\int_C\frac{d\lambda_2}{2\pi i\lambda_2}\cdots\int_C\frac{d\lambda_n}{2\pi i\lambda_n}\prod_{a=1}^{n}(1+\frac{i}{\lambda_a})^{n-a}(\frac{\pi\lambda_a}{\sinh\pi\lambda_a})^n \times\prod_{1\le k<j\le n}\frac{\sinh\pi(\lambda_j-\lambda_k)}{\pi(\lambda_j-\lambda_k-i)}. \qquad (11)$$

The contour C in each integral goes parallel to the real axis with the imaginary part between 0 and $-i$. In the frame of algebraic Bethe Ansatz this formula was derived in [20]. Recently such formula was generalized in paper [12] to the case, where averaging is done over arbitrary Bethe state [with no strings] instead of anti-ferromagnetic state.

Let us describe in general a strategy that may be used for the calculation of $P(n)$. The integral formula (11) can be easily represented as follows:

$$P(n)=\prod_{j=1}^{n}\int_C\frac{d\lambda_j}{2\pi i}\,U_n(\lambda_1,\ldots,\lambda_n)\,T_n(\lambda_1,\ldots,\lambda_n) \qquad (12)$$

where

$$U_n(\lambda_1,\ldots,\lambda_n)\;=\;\pi^{\frac{n(n+1)}{2}}\frac{\prod_{1\le k<j\le n}\sinh\pi(\lambda_j-\lambda_k)}{\prod_{j=1}^{n}\sinh^n\pi\lambda_j} \qquad (13)$$

and

$$T_n(\lambda_1,\ldots,\lambda_n)\;=\;\frac{\prod_{j=1}^{n}\lambda_j^{j-1}(\lambda_j+i)^{n-j}}{\prod_{1\le k<j\le n}(\lambda_j-\lambda_k-i)}. \qquad (14)$$

First of all, let us note that in principle the contour C can be chosen between 0 and $-i$ arbitrary. Let us denote C_α the contour that goes from $i\alpha-\infty$ to $i\alpha+\infty$. In what follows it will be convenient to choose $\alpha=-1/2$, i.e. to integrate over the contour $C_{-1/2}$.

As appeared we can make a lot of simplifications without taking integrals but using some simple observations and properties of the function in the r.h.s. of (12) which has to be integrated.

Let us define a "weak" equality $\sim$. Namely, let us say that two functions $F_n(\lambda_1,\ldots,\lambda_n)$ and $G_n(\lambda_1,\ldots,\lambda_n)$ are "weakly" equivalent

$$F_n(\lambda_1,\ldots,\lambda_n)\;\sim\;G_n(\lambda_1,\ldots,\lambda_n) \qquad (15)$$

if

$$\prod_{j=1}^{n}\int_{C_{-1/2}}\frac{d\lambda_j}{2\pi i}\,U_n(\lambda_1,\ldots,\lambda_n)\,(F_n(\lambda_1,\ldots,\lambda_n)\;-\;G_n(\lambda_1,\ldots,\lambda_n))\;=\;0. \qquad (16)$$

Let us also introduce a "canonical" form of function by the following formula

$$T_n^c(\lambda_1,\ldots,\lambda_n)\;=\;\sum_{j=0}^{[\frac{n}{2}]}P_j^{(n)}\prod_{k=1}^{j}\frac{1}{\lambda_{2k}-\lambda_{2k-1}}. \qquad (17)$$

where $P_j^{(n)}$ are some polynomials of the form

$$P_j^{(n)} \equiv P_j^{(n)}(\lambda_1, \lambda_3, \ldots, \lambda_{2j-1} | \lambda_{2j+1}, \lambda_{2j+2}, \ldots, \lambda_n) =$$

$$= \sum_{\substack{0 \le i_1, i_3, \ldots, i_{2j-1} \le n-2 \\ 0 \le i_{2j+1} < i_{2j+2} < \ldots < i_n \le n-1}} C_{i_1, \quad i_3, \ldots, i_{2j-1}}^{i_{2j+1}, i_{2j+2}, \ldots, i_n} \lambda_1^{i_1} \lambda_3^{i_3} \ldots \lambda_{2j-1}^{i_{2j-1}} \lambda_{2j+1}^{i_{2j+1}} \lambda_{2j+2}^{i_{2j+2}} \ldots \lambda_n^{i_n}, \quad (18)$$

where

$$C_{i_1, \quad i_3, \ldots, i_{2j-1}}^{i_{2j+1}, i_{2j+2}, \ldots, i_n} = i^\beta \hat{C}_{i_1, \quad i_3, \ldots, i_{2j-1}}^{i_{2j+1}, i_{2j+2}, \ldots, i_n}$$

with $\beta = 0$ or 1 in accordance with the equality

$$\beta + i_1 + i_3 + \ldots + i_{2j-1} + i_{2j+1} + i_{2j+2} + \ldots + i_n \equiv j + n \mod 2$$

and some rational numbers $\hat{C}_{i_1, \quad i_3, \ldots, i_{2j-1}}^{i_{2j+1}, i_{2j+2}, \ldots, i_n}$.

This form has some arbitrariness because if we substitute $\lambda_j = x_j - i/2$ where all x_j are real then it is easy to see that the function $\tilde{U}_n(x_1, \ldots, x_n) = U_n(x_1 - i/2, \ldots, x_n - i/2)$ transforms when $\{x_1, \ldots, x_n\} \to \{-x_1, \ldots, -x_n\}$ as follows

$$\tilde{U}_n(-x_1, \ldots, -x_n) = (-1)^{\frac{n(n-1)}{2}} \tilde{U}_n(x_1, \ldots, x_n). \quad (19)$$

Therefore any function $\tilde{F}_n(x_1, \ldots, x_n)$ that satisfies

$$\tilde{F}_n(-x_1, \ldots, -x_n) = (-1)^{\frac{n(n-1)}{2}+1} \tilde{F}_n(x_1, \ldots, x_n) \quad (20)$$

being integrated makes zero contribution

$$\prod_{j=1}^{n} \int_{-\infty}^{\infty} \frac{dx_j}{2\pi i} \tilde{U}_n(x_1, \ldots, x_n) \tilde{F}_n(x_1, \ldots, x_n) = 0.$$

It means that in order to get a nonzero result one should have the function $\tilde{F}_n(x_1, \ldots, x_n)$ of the same parity as of the function $\tilde{U}_n(x_1, \ldots, x_n)$. Then if we re-expand the form (17) in terms of variables x_j instead of λ_j we can fix the arbitrariness by imposing some additional constraints, namely,

$$\tilde{P}_j^{(n)}(x_1, x_3, \ldots, x_{2j-1} | x_{2j+1}, x_{2j+2}, \ldots, x_n)$$
$$= P_j^{(n)}(x_1 - i/2, x_3 - i/2, \ldots, x_{2j-1} - i/2 | x_{2j+1} - i/2, x_{2j+2} - i/2, \ldots, x_n - i/2)$$
$$= \sum_{\substack{0 \le i_1, i_3, \ldots, i_{2j-1} \le n-2 \\ 0 \le i_{2j+1} < i_{2j+2} < \ldots < i_n \le n-1 \\ i_1 + i_3 + \ldots + i_{2j-2} + i_{2j+1} + i_{2j+2} + \ldots + i_n \equiv j+n \bmod 2}} \tilde{C}_{i_1, \quad i_3, \ldots, i_{2j-1}}^{i_{2j+1}, i_{2j+2}, \ldots, i_n} x_1^{i_1} x_3^{i_3} \ldots x_{2j-1}^{i_{2j-1}} x_{2j+1}^{i_{2j+1}} x_{2j+2}^{i_{2j+2}} \ldots x_n^{i_n}. \quad (21)$$

In comparison with the coefficients $C_{i_1, \quad i_3, \ldots, i_{2j-1}}^{i_{2j+1}, i_{2j+2}, \ldots, i_n}$ which can be pure imaginary all the coefficients $\tilde{C}_{i_1, \quad i_3, \ldots, i_{2j-1}}^{i_{2j+1}, i_{2j+2}, \ldots, i_n}$ are real and rational numbers.

So we can expect that the function

$$\tilde{T}_n^c(x_1,\ldots,x_n) \equiv T_n^c(x_1 - i/2,\ldots,x_n - i/2) \tag{22}$$

should satisfy the following property

$$\tilde{T}_n^c(-x_1,\ldots,-x_n) = (-1)^{\frac{n(n-1)}{2}}\,\tilde{T}_n^c(x_1,\ldots,x_n) \tag{23}$$

Below the property (23) will be implied when we will speak about the "canonical" form (17–18). Besides, one can note that the function $\tilde{T}_n^c(x_1,\ldots,x_n)$ should be real for real variables x_j.

Our hypothesis is that for any n one can reduce the function T_n defined by (14) to the canonical form i.e. there exist polynomials P_j in (17) such that

$$T_n(\lambda_1,\ldots,\lambda_n) \;\sim\; T_n^c(\lambda_1,\ldots,\lambda_n). \tag{24}$$

Unfortunately, for the moment we do not have a proof of this statement for any n but we will demonstrate below how it works for $n = 2,3,4$.

In fact, the problem of the calculation of $P(n)$ given by the integral (12) can be reduced to the two steps. The first step corresponds to the obtaining of the "canonical" form for T_n. The second step is the calculation of the integral by means of this "canonical" form.

To do this one needs the following simple facts:

I. Since the function $U_n(\lambda_1,\ldots,\lambda_n)$ is antisymmetric with respect to transposition of any pair of integration variables, say, λ_j and λ_k the following integral

$$\prod_{j=1}^{n}\int_C \frac{d\lambda_j}{2\pi i}\, U(\lambda_1,\ldots,\lambda_n)\, S(\lambda_1,\ldots,\lambda_n) \;=\; 0 \tag{25}$$

if the function S is symmetric for at least one pair of λ-s. Therefore for an arbitrary function $F_n(\lambda_1,\ldots,\lambda_n)$ one can transpose any pair of λ-s taking into consideration appearance of additional sign because of the antisymmetry of $U_n(\lambda_1,\ldots,\lambda_n)$. For example, if one transposes λ_j with λ_k one gets

$$F_n(\ldots,\lambda_j,\ldots,\lambda_k,\ldots) \;\sim\; -F_n(\ldots,\lambda_k,\ldots,\lambda_j,\ldots). \tag{26}$$

II. The reduction of the power of denominator for T_n is based on two relations which can be checked directly

$$\begin{aligned}&\frac{1}{\lambda_k-\lambda_l-i}\,\frac{1}{\lambda_j-\lambda_l-i}\,\frac{1}{\lambda_j-\lambda_k-i}\\ &\qquad = i\,\frac{1}{\lambda_j-\lambda_l-i}\,\frac{1}{\lambda_j-\lambda_k-i} + i\,\frac{1}{\lambda_k-\lambda_l-i}\,\frac{1}{\lambda_j-\lambda_l-i}\\ &\qquad\qquad - i\,\frac{1}{\lambda_k-\lambda_l-i}\,\frac{1}{\lambda_j-\lambda_k-i}\end{aligned} \tag{27}$$

$$\prod_{k=1}^{j-1}\frac{1}{\lambda_j-\lambda_k-i} = \sum_{k=1}^{j-1}\frac{1}{\lambda_j-\lambda_k-i}\prod_{\substack{l=1\\ l\neq k}}^{j-1}\frac{1}{\lambda_k-\lambda_l}. \tag{28}$$

In Appendices A and B we will show how the reduction can be performed for $n=3$ and $n=4$. Unfortunately, so far we have not succeeded in finding a result for general n.

III. The ratio

$$\frac{T_{n+1}(\lambda_1,\ldots,\lambda_{n+1})}{T_n(\lambda_1,\ldots,\lambda_n)} = \frac{\prod_{j=1}^{n}(\lambda_j+i)\,\lambda_{n+1}^n}{\prod_{j=1}^{n}(\lambda_{n+1}-\lambda_j-i)} \tag{29}$$

is symmetric with respect to any permutation of $\lambda_1,\ldots,\lambda_n$. Therefore the relation (29) allows us to use the result for T_n also for derivation of T_{n+1} if this result was obtained by applying relations (26–28) from I and II.

IV.

Proposition 1. *Let the function $f(\lambda_1,\ldots,\lambda_n)$ have only poles of form $1/(\lambda_j-\lambda_k+ia)$ with a an integer i.e. the product $U_n(\lambda_1,\ldots,\lambda_n)f(\lambda_1,\ldots,\lambda_n)$ does not have poles of that kind. Then*

$$\lambda_j^m\, f(\ldots,\lambda_j,\ldots) \;\sim\; -(\lambda_j+i)^m\, f(\ldots,\lambda_j+i,\ldots) \tag{30}$$

where m is an integer and $m\geq n$.

Proof. Let us suppose that all variables $\lambda_1,\ldots,\lambda_{j-1},\lambda_{j+1},\ldots,\lambda_n$ are fixed. Extracting from $U_n(\lambda_1,\ldots,\lambda_n)$ the function which depends on λ_j one gets

$$\begin{aligned}
&\int_{C_{-1/2}}\frac{d\lambda_j}{2\pi i}\frac{\prod_{k\neq j}\sinh\pi(\lambda_j-\lambda_k)}{\sinh^n\pi\lambda_j}\lambda_j^m\, w(\lambda_j)\\
&\quad= -\int_{C_{-3/2}}\frac{d\lambda_j}{2\pi i}\frac{\prod_{k\neq j}\sinh\pi(\lambda_j-\lambda_k)}{\sinh^n\pi\lambda_j}(\lambda_j+i)^m\, w(\lambda_j+i)\\
&= -\Big(\int_{C_{-3/2}}-\int_{C_{-1/2}}+\int_{C_{-1/2}}\Big)\frac{d\lambda_j}{2\pi i}\frac{\prod_{k\neq j}\sinh\pi(\lambda_j-\lambda_k)}{\sinh^n\pi\lambda_j}(\lambda_j+i)^m\, w(\lambda_j+i)
\end{aligned}$$

where $w(\lambda_j)=f(\ldots,\lambda_j,\ldots)$. The first step here was to shift integration variable $\lambda_j\to\lambda_j+i$ and to use the fact that $\sinh\pi(x+i) = -\sinh\pi x$. The two first integrals in the last expression are equal to a contour integral around the point $\lambda_j=-i$ in a complex plane of the variable λ_j. Since, $m\geq n$ the term $(\lambda_j+i)^m$ which is in the numerator and corresponds to a zero of order m compensates the pole from the term $\sinh^n\pi\lambda_j$ in the denominator. Therefore the contribution of those two integrals is zero and we immediately come to the statement (30). □

One can get two useful corollaries from Proposition 1.

Corollary 2.

$$\lambda_j^m\, g(\lambda_1,\ldots,\hat{\lambda}_j,\ldots,\lambda_n) \;\sim\; \frac{(-i)}{2}\sum_{k=0}^{m-1}\lambda_j^k(\lambda_j+i)^{m-1-k}\, g(\lambda_1,\ldots,\hat{\lambda}_j,\ldots,\lambda_n) \tag{31}$$

where the function $g(\lambda_1,\ldots,\hat{\lambda}_j,\ldots,\lambda_n)$ does not depend on λ_j and as above it is implied that $m\geq n$.

Proof. The relation (31) is easy to derive using the relation $(\lambda_j^m + (\lambda_j + i)^m) \times g(\lambda_1, \ldots, \hat{\lambda}_j, \ldots, \lambda_n) \sim 0$ or equivalently $\lambda_j^m\, g(\lambda_1, \ldots, \hat{\lambda}_j, \ldots, \lambda_n) \sim 1/2\, (\lambda_j^m - (\lambda_j + i)^m)\, g(\lambda_1, \ldots, \hat{\lambda}_j, \ldots, \lambda_n)$. □

Corollary 3.

$$\frac{\lambda_j^{m-1}}{\lambda_k - \lambda_j}\, g(\lambda_1, \ldots, \hat{\lambda}_k, \ldots, \hat{\lambda}_j, \ldots, \lambda_n)$$
$$\sim \frac{i}{m}\Big(\sum_{l=2}^{m} \frac{\binom{m}{l} i^l \lambda_j^{m-l}}{\lambda_k - \lambda_j} + \sum_{l=0}^{m-1} \lambda_k^l (\lambda_j + i)^{m-1-l}\Big) g(\lambda_1, \ldots, \hat{\lambda}_k, \ldots, \hat{\lambda}_j, \ldots, \lambda_n) \tag{32}$$

where

$$\binom{m}{l} = \frac{m!}{l!\,(m-l)!}$$

is binomial coefficient and the function $g(\lambda_1, \ldots, \hat{\lambda}_k, \ldots, \hat{\lambda}_j, \ldots, \lambda_n)$ does not depend on λ_k and λ_j and $m \geq n$.

Proof. Using Proposition 1 we get

$$\begin{aligned}
&\frac{\lambda_j^m}{\lambda_k - \lambda_j}\, g(\lambda_1, \ldots, \hat{\lambda}_k, \ldots, \hat{\lambda}_j, \ldots, \lambda_n) \\
&\qquad \sim -\frac{(\lambda_j + i)^m}{\lambda_k - \lambda_j - i}\, g(\lambda_1, \ldots, \hat{\lambda}_k, \ldots, \hat{\lambda}_j, \ldots, \lambda_n) \\
&\quad = \Big(-\frac{\lambda_k^m}{\lambda_k - \lambda_j - i} + \frac{\lambda_k^m - (\lambda_j + i)^m}{\lambda_k - \lambda_j - i}\Big)\, g(\lambda_1, \ldots, \hat{\lambda}_k, \ldots, \hat{\lambda}_j, \ldots, \lambda_n) \\
&\quad \sim \Big(\frac{(\lambda_k + i)^m}{\lambda_k - \lambda_j} + \frac{\lambda_k^m - (\lambda_j + i)^m}{\lambda_k - \lambda_j - i}\Big)\, g(\lambda_1, \ldots, \hat{\lambda}_k, \ldots, \hat{\lambda}_j, \ldots, \lambda_n) \\
&\qquad \sim \Big(\frac{(\lambda_j + i)^m}{\lambda_k - \lambda_j} + \frac{\lambda_k^m - (\lambda_j + i)^m}{\lambda_k - \lambda_j - i}\Big)\, g(\lambda_1, \ldots, \hat{\lambda}_k, \ldots, \hat{\lambda}_j, \ldots, \lambda_n)
\end{aligned}$$

or

$$\Big(\frac{(\lambda_j + i)^m - \lambda_j^m}{\lambda_k - \lambda_j} + \frac{\lambda_k^m - (\lambda_j + i)^m}{\lambda_k - \lambda_j - i}\Big)\, g(\lambda_1, \ldots, \hat{\lambda}_k, \ldots, \hat{\lambda}_j, \ldots, \lambda_n) \sim 0.$$

Then expanding both numerators according to formulae

$$(\lambda_j + i)^m - \lambda_j^m = \sum_{l=1}^{m} \binom{m}{l} i^l \lambda_j^{m-l}$$

$$\lambda_k^m - (\lambda_j + i)^m = (\lambda_k - \lambda_j - i) \sum_{l=0}^{m-1} \lambda_k^l (\lambda_j + i)^{m-1-l}$$

we arrive at formula (32). □

With the help of Corollaries 2 and 3 one can effectively reduce the power of the numerator in T_n.

V. For the calculation of integrals we need the following

Proposition 4. *Let the integral*

$$\int_{C_{-1/2}} \frac{d\lambda}{2\pi i}\, e^{i\delta\lambda}\, \frac{f(\lambda+iN)}{\sinh^n \pi\lambda}$$

be convergent for any real number $\delta \geq 0$ and integer N. Further, let the function $f(\lambda)$ be analytic in the whole complex plane λ and satisfy the following two conditions

$$\lim_{R\to\infty} |e^{-n\pi R}\, f(iy - i/2 \pm R)| = 0, \tag{33}$$

$$\lim_{N\to\infty} |e^{-\delta' N}\, \frac{f(x - i/2 \pm iN)}{\cosh^n \pi x}| = 0 \tag{34}$$

where the first limit is uniform in y, when $y \in [0, N]$. The second limit is uniform in x for any real x. The value δ' is a fixed real positive number ($\delta' > 0$). Then

$$\int_{C_{-1/2}} \frac{d\lambda}{2\pi i}\, \frac{f(\lambda)}{\sinh^n \pi\lambda} = \lim_{\delta\to 0+} d^{(n)}(\epsilon)_{\epsilon\to 0} \sum_{l=0}^{\infty} (-1)^{ln}\, e^{-\delta l}\, f(il+\epsilon) \tag{35}$$

$$= -\lim_{\delta\to 0+} d^{(n)}(\epsilon)_{\epsilon\to 0} \sum_{l=1}^{\infty} (-1)^{ln}\, e^{-\delta l}\, f(-il+\epsilon) \tag{36}$$

where a differential operator $d^{(n)}(\epsilon)$ looks as follows

$$d^{(n)}(\epsilon) = \frac{1}{\pi^n (n-1)!} \sum_{l=0}^{n-1} \binom{n-1}{l} \left(\frac{\partial}{\partial\alpha}\right)^{n-1-l}_{\alpha\to 0} \left(\sum_{0\leq 2k<n} \frac{(\pi\alpha)^{2k}}{(2k+1)!}\right)^{-n} \left(\frac{\partial}{\partial\epsilon}\right)^{l}. \tag{37}$$

In particular, for $n = 2, 3, 4$

$$d^{(2)}(\epsilon) = \frac{1}{\pi^2}\frac{\partial}{\partial\epsilon} \tag{38}$$

$$d^{(3)}(\epsilon) = -\frac{1}{2\pi}\,(1 - \frac{1}{\pi^2}\frac{\partial^2}{\partial\epsilon^2}) \tag{39}$$

$$d^{(4)}(\epsilon) = -\frac{2}{3\pi^2}\,(\frac{\partial}{\partial\epsilon} - \frac{1}{4\pi^2}\frac{\partial^3}{\partial\epsilon^3}). \tag{40}$$

Proof. Let

$$F(\lambda, N) = \int_{C_N} \frac{d\lambda}{2\pi i}\, e^{i\delta\lambda} \frac{f(\lambda)}{\sinh^n \pi\lambda} \tag{41}$$

where $\delta > 0$ and C_N is a rectangular contour shown in Fig. 1.

The contours $C_{-1/2}$ and $-C_{N-1/2}$ correspond to lower and upper horizontal parts of the contour C_N respectively (sign $-$ is because the contour $C_{N-1/2}$ should

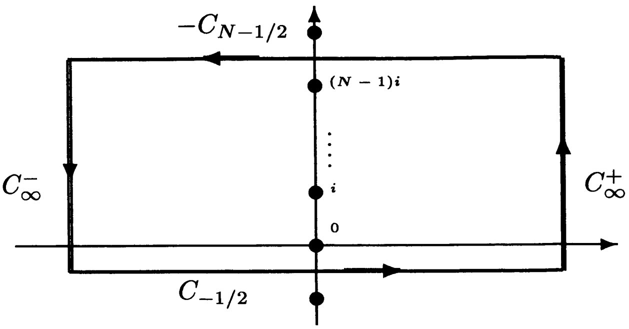


Figure 1. The contour C_N

be taken in the opposite direction). The contours C_{∞}^{+} and C_{∞}^{-} correspond to the right and left vertical parts of the contour C_N and have real parts $+\infty$ and $-\infty$ respectively. Due to the Cauchy theorem one has

$$F(\delta, N) = F_{-1/2} - F_{N-1/2} + F_{+} + F_{-} = d^{(n)}(\epsilon)_{\epsilon \to 0} \sum_{l=0}^{N-1} (-1)^{ln} e^{-\delta l + i\delta\epsilon} f(il + \epsilon) \tag{42}$$

where

$$F_{-1/2} = \int_{C_{-1/2}} \frac{d\lambda}{2\pi i} e^{i\delta\lambda} \frac{f(\lambda)}{\sinh^n \pi\lambda} \tag{43}$$

$$F_{N-1/2} = \int_{C_{N-1/2}} \frac{d\lambda}{2\pi i} e^{i\delta\lambda} \frac{f(\lambda)}{\sinh^n \pi\lambda} \tag{44}$$

$$F_{\pm} = \int_{C_{\infty}^{\pm}} \frac{d\lambda}{2\pi i} e^{i\delta\lambda} \frac{f(\lambda)}{\sinh^n \pi\lambda} \tag{45}$$

The r.h.s. of the formula (42) is a result of the calculation of residues corresponding to zeros of the denominator $\sinh^n \pi\lambda$ which are placed inside the contour C_N.

The first step is to prove that the integrals over the contours $C_{\infty}^{\pm}$

$$F_{\pm} = 0. \tag{46}$$

Actually one has

$$\begin{aligned} F_{\pm} &= \lim_{R \to \infty} \int_{C_R^{\pm}} \frac{d\lambda}{2\pi i} e^{i\delta\lambda} \frac{f(\lambda)}{\sinh^n \pi\lambda} \\ &= \frac{1}{2\pi i} \lim_{R \to \infty} \pm \int_0^N \frac{dy}{(-i)^n \cosh^n \pi(\pm R + iy)} e^{i\delta(\pm R + iy - i/2)} f(\pm R + iy - i/2) \end{aligned} \tag{47}$$

where vertical contours $C_R^{\pm}$ are defined as follows: $\{\lambda = \pm R + iy - i/2;\quad y \in [0, N]\}$. Then

$$
\begin{aligned}
|F_{\pm}| &\leq \frac{1}{2\pi} \lim_{R\to\infty} \int_0^N \frac{dy}{\sinh^n \pi R} e^{-\delta y} |f(\pm R + iy - i/2)| \\
&= \frac{2^{n-1}}{\pi} \lim_{R\to\infty} \int_0^N \frac{dy e^{-\delta y}}{(1 - e^{-2\pi R})^n} e^{-n\pi R} |f(\pm R + iy - i/2)| \qquad (48)
\end{aligned}
$$

where we have used a simple fact that for $R > 0$

$$|\cosh \pi(\pm R + iy)| \geq \sinh \pi R.$$

The uniform character of limit (33) allows to interchange the order of the integration over y and the limiting procedure $R \to \infty$. Indeed, condition (33) means that for any small real number ϵ there exists a real R_ϵ which is independent of y such that for any $R > R_\epsilon$

$$e^{-n\pi R} |f(\pm R + iy - i/2)| < \epsilon.$$

Therefore for $R > R_\epsilon$

$$
\begin{aligned}
\int_0^N \frac{dy\, e^{-\delta y}}{(1 - e^{-2\pi R})^n} e^{-n\pi R} |f(\pm R + iy - i/2)| &< \frac{\epsilon}{(1 - e^{-2\pi R})^n} \int_0^N dy\, e^{-\delta y} \\
= \frac{\epsilon}{(1 - e^{-2\pi R})^n} \frac{1 - e^{-N\delta}}{\delta} &< \frac{\epsilon}{(1 - e^{-2\pi R_\epsilon})^n} \frac{1 - e^{-N\delta}}{\delta} \qquad (49)
\end{aligned}
$$

Hence we have got that

$$\lim_{R\to\infty} \int_0^N \frac{dy e^{-\delta y}}{(1 - e^{-2\pi R})^n} e^{-n\pi R} |f(\pm R + iy - i/2)| = 0$$

and we come to the statement (46).

The next our step is to prove that for a fixed real $\delta > 0$

$$\lim_{N\to\infty} F_{N-1/2} = 0 \qquad (50)$$

Indeed,

$$
\begin{aligned}
|F_{N-1/2}| &= \left| \int_{C_{N-1/2}} \frac{d\lambda}{2\pi i} e^{i\delta\lambda} \frac{f(\lambda)}{\sinh^n \pi\lambda} \right| \\
&= \left| (-1)^{Nn} \int_{C_{-1/2}} \frac{d\lambda}{2\pi i} e^{i\delta(\lambda + iN)} \frac{f(\lambda + iN)}{\sinh^n \pi\lambda} \right| \\
&= \frac{e^{-\delta N}}{2\pi} \left| \int_{C_{-1/2}} \frac{d\lambda}{\sinh^n \pi\lambda} e^{i\delta\lambda} f(\lambda + iN) \right| \\
&\leq \frac{e^{-\delta N}}{2\pi} \int_{-\infty}^{\infty} \frac{dx}{\cosh^n \pi x} |f(x - i/2 + iN)| \qquad (51)
\end{aligned}
$$

As above due to the uniform character of limit (34) one can interchange the integration over x and the limit $N \to \infty$. Therefore the last expression in (51) tends to zero when $N \to \infty$ and we come to statement (50).

So for a fixed $\delta > 0$ we conclude from (42) that

$$\lim_{N\to\infty} F(\delta, N) = \int_{C_{-1/2}} \frac{d\lambda}{2\pi i} e^{i\delta\lambda} \frac{f(\lambda)}{\sinh^n \pi\lambda}$$

$$= d^{(n)}(\epsilon)_{\epsilon\to 0} \sum_{l=0}^{\infty} (-1)^{ln} e^{-\delta l + i\delta\epsilon} f(il + \epsilon).$$

The convergence of the sum in r.h.s. of the last expression is guaranteed by the convergence of the integral in the l.h.s. Finally, after taking the limit $\lim_{\delta\to 0^+}$ we come to formula (35). Let us note that generally speaking we can not interchange the order of the limits $N \to \infty$ and $\delta \to 0^+$.

Another form (36) can be proved in a similar way if considering a contour which is analogous to C_N but placed in a lower half-plane and a real number δ should be taken negative. It completes the proof of Proposition 4. $\square$

Using Proposition 4 we can get the following

Proposition 5.

$$\prod_{j=1}^{n} \int_{C_{-1/2}} \frac{d\lambda_j}{2\pi i} U_n(\lambda_1, \ldots, \lambda_n)\, F_n(\lambda_1, \ldots, \lambda_n) \;=\; D^{(n)}\, \tilde{F}_n(\epsilon_1, \ldots, \epsilon_n) \qquad (52)$$

where the multiple integral converges and the product $U_n(\lambda_1, \ldots, \lambda_n) F_n(\lambda_1, \ldots, \lambda_n)$ does not have any other poles besides the poles of the denominator $\prod_{j=1}^{n} \sinh^n \pi\lambda_j$ of the function $U_n(\lambda_1, \ldots, \lambda_n)$. The function

$$G_n^{(j)}(\lambda_1, \ldots, \lambda_n) \;=\; \frac{\prod_{1\le k<l\le n} \sinh \pi(\lambda_l - \lambda_k)}{\prod_{k\ne j} \sinh^n \pi\lambda_k} F_n(\lambda_1, \ldots, \lambda_n) \qquad (53)$$

should satisfy conditions which generalize (33) and (34)

$$\lim_{R\to\infty} \left| e^{-n\pi R} G_n^{(j)}(\lambda_1, \ldots, \lambda_{j-1}, ix_j - i/2 \pm R, \lambda_{j+1}, \ldots, \lambda_n) \right| = 0, \qquad j = 1, \ldots, n \quad (54)$$

$$\lim_{N\to\infty} \left| e^{-\delta N} G_n^{(j)}(\lambda_1, \ldots, \lambda_{j-1}, x_j - i/2 \pm iN, \lambda_{j+1}, \ldots, \lambda_n) \right| / \cosh^n \pi x_j = 0, \qquad j = 1, \ldots, n \quad (55)$$

where for each j both limits are uniform on real numbers $x_1, \ldots, x_n$ with $\lambda_k = x_k + im_k/2$, $k \ne j$ and some integers m_k. For the first limit $x_j \in [0, N]$ while for the second one x_j is any real number. $D^{(n)}$ is a differential operator

$$D^{(n)} = \pi^{\frac{n(n+1)}{2}} \prod_{j=1}^{n} d^{(n)}(\epsilon_j)_{\epsilon_j\to 0} \prod_{1\le k<j\le n} \sinh \pi(\epsilon_j - \epsilon_k) \qquad (56)$$

and

$$\tilde{F}_n(\epsilon_1,\dots,\epsilon_n)$$
$$= \lim_{\delta_1\to 0+}\sum_{l_1=0}^{\infty}(-1)^{l_1}e^{-\delta_1 l_1}\dots\lim_{\delta_n\to 0+}\sum_{l_n=0}^{\infty}(-1)^{l_n}e^{-\delta_n l_n}F_n(i\,l_1\,+\,\epsilon_1,\dots,i\,l_n\,+\,\epsilon_n) \tag{57}$$

where each sum $\sum_{l_j=0}^{\infty}(-1)^{l_j}e^{-\delta_j l_j}F_n(\lambda_1,\dots,\lambda_{j-1},i\,l_j\,+\,\epsilon_j,\lambda_{j+1},\dots,\lambda_n)$ *should be convergent uniformly in other arguments* $\lambda_1,\dots,\lambda_{j-1},\lambda_{j+1},\dots,\lambda_n$ $(\lambda_k = x_k + im_k/2,\quad k\neq j)$.

Proof. Formula (52) can be obtained by the recursive application of formula (35) to each integral in the l.h.s. of (52) and by taking into account the manifest form (13) of the function $U_n(\lambda_1,\dots,\lambda_n)$. After $j-1$-th application one can interchange the integration over λ_j with the summation over $l_1,\dots,l_{j-1}$ due to the uniform convergence of these sums. Due to conditions (54) and (55) one can apply Proposition 4 when integrating over the variable λ_j.

Let us note that each sum $\sum_{l_j=0}^{\infty}(-1)^{l_j}e^{-\delta_j l_j}F_n(\dots,i\lambda_j\,+\,\epsilon_j,\dots)$ in formula (57) can be substituted by $-\sum_{l_j=1}^{\infty}(-1)^{l_j}e^{-\delta_j l_j}F_n(\dots,-i\lambda_j+\epsilon_j,\dots)$ corresponding to the choice of the contour in the lower half-plane. In what follows we will use this fact depending on a convenience. □

Let us consider a special class of functions $\tilde{F}_n(\epsilon_1,\dots,\epsilon_n)$ that does not have a singularity when $\epsilon_j\to 0$ for $j=1,\dots,n$. In this case one can expand $\tilde{F}_n(\epsilon_1,\dots,\epsilon_n)$ into the infinite series on powers of ϵ-s. We have checked that for $n\le 4$ the differential operator $D^{(n)}$ given by (56) when acting on some monomial $\epsilon_1^{i_1}\dots\epsilon_n^{i_n}$ makes non-zero contribution only for monomial of the form $\epsilon_{\sigma(1)}^{0}\epsilon_{\sigma(2)}^{1}\cdots\epsilon_{\sigma(n)}^{n-1}$ where σ is some element of the permutation group S_n of n elements. More precisely, we can write

$$D^{(n)} = \frac{1}{\prod_{j=1}^{n-1}j!}\prod_{1\le k<j\le n}\left(\frac{\partial}{\partial\epsilon_k}-\frac{\partial}{\partial\epsilon_j}\right)_{\vec{\epsilon}\to 0} \tag{58}$$

We proved it for $n\le 4$ and we use this formula only in this case. We believe that this relation is valid for any n but this fact is still to be proven.

VI. Now let us discuss the integrals of a special form, namely, when the function $F_n(\lambda_1,\dots,\lambda_n)$ from Proposition 5 is rational on it's arguments and the function

$$\prod_{1\le k<j\le n}\sinh\pi(\lambda_j-\lambda_k)F_n(\lambda_1,\dots,\lambda_n)$$

is analytic. It means that the function $F_n(\lambda_1,\dots,\lambda_n)$ can have only simple poles when $\lambda_j\to\lambda_k + mi$ with an integer m. As it is seen from definitions (14), (17) both the function $T_n(\lambda_1,\dots,\lambda_n)$ and the function $T_n^c(\lambda_1,\dots,\lambda_n)$ are of that form.

Let us show that the conditions of applicability of Propositions 4 and 5 are fulfilled for such a function $F_n(\lambda_1,\dots,\lambda_n)$. To show that the first condition (54) is satisfied let us fix without loss of generality $j=1$ and take into account that

$0 \le x_1 \le N$. First let us consider the case when the rational function F_n does not have poles at all. Then if $m_2, \dots, m_n$ have the same parity

$$
\begin{aligned}
& e^{-n\pi R} \, |G_n^{(1)}(ix_1 - i/2 \pm R, x_2 + im_2/2, \dots, x_n + im_n/2)| \\
&= e^{-n\pi R} \prod_{k=2}^{n} |\sinh \pi(x_k \mp R - ix_1)| \frac{\prod_{1<k<l\le n} |\sinh \pi(x_l - x_k)|}{\prod_{k=2}^{n} \cosh^n \pi x_k} \\
&\qquad \times |F_n(ix_1 - i/2 \pm R, x_2 + im_2/2, \dots, x_n + im_n/2)| \\
&\le e^{-n\pi R} \, 2^{\frac{(n-2)(n-1)}{2}} \prod_{k=2}^{n} \frac{|\sinh \pi(x_k \mp R - ix_1)|}{\cosh^2 \pi x_k} \\
&\qquad \times |F_n(ix_1 - i/2 \pm R, x_2 + im_2/2, \dots, x_n + im_n/2)| \\
&\le e^{-n\pi R} \, 2^{\frac{(n-2)(n-1)}{2}} \prod_{k=2}^{n} \frac{\cosh \pi(x_k \mp R)}{\cosh^2 \pi x_k} \\
&\qquad \times |F_n(ix_1 - i/2 \pm R, x_2 + im_2/2, \dots, x_n + im_n/2)| \\
&\le e^{-n\pi R} \, 2^{\frac{(n-2)(n-1)}{2}} (\cosh \pi R + \sinh \pi R)^{n-1} \prod_{k=2}^{n} \frac{1}{\cosh \pi x_k} \\
&\qquad \times |F_n(ix_1 - i/2 \pm R, x_2 + im_2/2, \dots, x_n + im_n/2)| \\
&= 2^{\frac{(n-2)(n-1)}{2}} e^{-\pi R} |F_n(ix_1 - i/2 \pm R, x_2 + im_2/2, \dots, x_n + im_n/2)| \qquad (59)
\end{aligned}
$$

where we have used an inequality

$$
\frac{\prod\limits_{1<k<l\le n} |\sinh \pi(x_l - x_k)|}{\prod\limits_{k=2}^{n} \cosh^n \pi x_k} \le 2^{\frac{(n-2)(n-1)}{2}} \prod_{k=2}^{n} \frac{1}{\cosh^2 \pi x_k} \qquad (60)
$$

which can be checked directly. For a set of arbitrary integers m_k, $k = 2, \dots, n$ one can repeat the derivation above also.

Since $0 \le x_1 \le N$ and F_n is rational without poles one can find real numbers $R^\star$ and $M(N) > 0$ which is independent of R such that for $R > R^\star$

$$
\prod_{k=2}^{n} \frac{1}{\cosh \pi x_k} |F_n(ix_1 - i/2 \pm R, x_2 + im_2/2, \dots, x_n + im_n/2)| \le R^s M(N)
$$

with some power s. Hence, we get that expression (59) can not exceed

$$
2^{\frac{(n-2)(n-1)}{2}} \, e^{-\pi R} \, R^s M(N)
$$

which tends to zero when $R \to \infty$ independently of the variables $x_1, \dots, x_n$ and we get the uniform character of the limit (54) on these variables.

Let us prove the validity of the second limit (55). Indeed, for a given integer $j = 1, \dots n$ and $\lambda_k = x_k + im_k/2$, $\quad k \ne j$ with integers m_k of the same parity the function

$$
|e^{-\delta N} \, G_n^{(j)}(\lambda_1, \dots, \lambda_{j-1}, x_j - i/2 \pm iN, \lambda_{j+1}, \dots, \lambda_n)| / \cosh^n \pi x_j
$$

is bounded for any real numbers $x_k \in (-\infty, \infty)$ and a positive integer N. Therefore one can use again the inequality (60) in order to get

$$\prod_{k=2}^{n} |\sinh \pi(x_k - x_1 - iN)| \frac{\prod_{1<k<l\leq n} |\sinh \pi(x_l - x_k)|}{\prod_{k=1}^{n} \cosh^n \pi x_k}$$
$$= \frac{\prod_{1\leq k<l\leq n} |\sinh \pi(x_l - x_k)|}{\prod_{k=1}^{n} \cosh^n \pi x_k} \leq 2^{\frac{n(n-1)}{2}} \frac{1}{\prod_{k=1}^{n} \cosh \pi x_k}$$

Again one can repeat this for arbitrary integers $m_2, \ldots, m_n$.

Since F_n is rational the maximum over the real variables $x_1, \ldots, x_n$ of the function

$$\frac{1}{\prod_{k=1}^{n} \cosh \pi x_k} |F_n(x_1 - i/2 + iN, x_2 + im_2/2, \ldots, x_n + im_n/2)|$$

can be some power of N, say, $N^{s'}$ multiplied with some constant which is independent of N when $N > N^\star$ with $N^\star$ is a big enough integer. Therefore we get the limit

$$\lim_{N\to\infty} e^{-\delta N} N^{s'} = 0$$

for any real $\delta > 0$ independently of variables $x_1, \ldots, x_n$ and we come to the uniform limit (55).

Suppose the function $F_n(\lambda_1, \ldots, \lambda_n)$ has a simple pole of a type $1/(\lambda_k - \lambda_l + ia_{kl})$ with an integer a_{kl}. Let us restrict ourself only with the case $a_{kl} = 0$ because only such poles can appear in the expression for a canonical form (17). Then we can write

$$\left| \frac{\sinh \pi(\lambda_k - \lambda_l)}{\lambda_k - \lambda_l} \right| \leq \begin{cases} \pi \sinh 1 & \text{if } |\lambda_k - \lambda_l| \leq \frac{1}{\pi} \\ \pi |\sinh \pi(\lambda_k - \lambda_l)| & \text{if } |\lambda_k - \lambda_l| > \frac{1}{\pi} \end{cases}$$

and use again the technique described above.

Let us comment on a question of uniform convergence of a sum

$$\sum_{l_j=0}^{\infty} (-1)^{l_j} e^{-\delta_j l_j} F_n(\lambda_1, \ldots, \lambda_{j-1}, i\, l_j + \epsilon_j, \lambda_{j+1}, \ldots, \lambda_n)$$

as well as how to proceed further by considering two typical examples. Let us take for simplicity $n = 2$. The integral we need looks as follows

$$J_2 = \int_{C_{-1/2}} \frac{d\lambda_2}{2\pi i} \int_{C_{-1/2}} \frac{d\lambda_1}{2\pi i} \frac{\sinh \pi(\lambda_2 - \lambda_1)}{\sinh^2 \pi\lambda_1 \sinh^2 \pi\lambda_2} F_2(\lambda_1, \lambda_2). \tag{61}$$

(i) As a first example let us consider again the case when F_2 does not have poles at all i.e. $F_2(\lambda_1, \lambda_2)$ is some polynomial on λ_1 and λ_2. First let us integrate over λ_1 using formula (35)

$$J_2 = \int_{C_{-1/2}} \frac{d\lambda_2}{2\pi i \sinh^2 \pi\lambda_2} \lim_{\delta_1 \to 0+} d^{(2)}(\epsilon_1)_{\epsilon_1 \to 0} \sinh \pi(\lambda_2 - \epsilon_1) \times \sum_{l_1=0}^{\infty} (-1)^{l_1} e^{-\delta_1 l_1} F_2(il_1 + \epsilon_1, \lambda_2). \quad (62)$$

Since $F_2(\lambda_1, \lambda_2)$ is a polynomial then $F_2(il_1 + \epsilon_1, \lambda_2)$ is a polynomial on l_1 as well. Let us pick out some monomial on l_1 from it, say,

$$l_1^a F_1(\lambda_2) \quad (63)$$

where a is a non-negative integer and $F_1(\lambda_2)$ is a polynomial on λ_2. Actually it has a factorized form. Therefore in this case we do not have any problems with the uniform convergence and the corresponding contribution into J_2 is as follows

$$\begin{aligned} &\int_{C_{-1/2}} \frac{d\lambda_2}{2\pi i \sinh^2 \pi\lambda_2} F_1(\lambda_2) \lim_{\delta_1 \to 0+} d^{(2)}(\epsilon_1)_{\epsilon_1 \to 0} \sinh \pi(\lambda_2 - \epsilon_1) \times \sum_{l_1=0}^{\infty} (-1)^{l_1} e^{-\delta_1 l_1} l_1^a \\ &= \int_{C_{-1/2}} \frac{d\lambda_2}{2\pi i \sinh^2 \pi\lambda_2} F_1(\lambda_2) d^{(2)}(\epsilon_1)_{\epsilon_1 \to 0} \sinh \pi(\lambda_2 - \epsilon_1) \rho(a) \end{aligned} \quad (64)$$

where

$$\rho(a) = \lim_{\delta \to 0+} \sum_{l_1=0}^{\infty} (-e^{-\delta})^{l_1} l_1^a \quad (65)$$

Let us adduce a number of first values of $\rho(a)$

$$\rho(0) = \frac{1}{2},\ \rho(1) = -\frac{1}{4},\ \rho(2) = 0,\ \rho(3) = \frac{1}{8},\ \rho(4) = 0,\ \rho(5) = -\frac{1}{4} \quad (66)$$

Since $F_1(\lambda_2)$ is a polynomial we can treat the integral in (64) in a similar way as the integral over λ_1. In the very end we should calculate the limits $\epsilon_1 \to 0$ and $\epsilon_2 \to 0$ and get the final answer for J_2.

(*ii*) The second example corresponds to an existing of a simple pole, namely, when

$$F_2(\lambda_1, \lambda_2) = \frac{Q(\lambda_1, \lambda_2)}{\lambda_2 - \lambda_1 - ia_{12}} \quad (67)$$

where $Q(\lambda_1, \lambda_2)$ is a polynomial on λ_1, λ_2. As above we shall consider only the case $a_{12} = 0$. So in this case doing the first integration one gets

$$J_2 = \int_{C_{-1/2}} \frac{d\lambda_2}{2\pi i \sinh^2 \pi\lambda_2} \lim_{\delta_1 \to 0+} d^{(2)}(\epsilon_1)_{\epsilon_1 \to 0} \sinh \pi(\lambda_2 - \epsilon_1) \times \sum_{l_1=0}^{\infty} (-1)^{l_1} e^{-\delta_1 l_1} \frac{Q(il_1 + \epsilon_1, \lambda_2)}{\lambda_2 - il_1 - \epsilon_1}. \quad (68)$$

Since $Q(il_1+\epsilon_1,\lambda_2)$ is a polynomial on l_1 also again let us pick out some monomial from it

$$l_1^{a'}Q'(\lambda_2)$$

with an integer $a' \geq 0$ and a polynomial $Q'(\lambda_2)$. Then the corresponding contribution into the expression (68) for J_2 looks as follows

$$\int_{C_{-1/2}} \frac{d\lambda_2}{2\pi i\,\sinh^2\pi\lambda_2} Q'(\lambda_2) \lim_{\delta_1\to 0+} d^{(2)}(\epsilon_1)_{\epsilon_1\to 0} \sinh\pi(\lambda_2-\epsilon_1) \times \sum_{l_1=0}^{\infty}(-1)^{l_1}e^{-\delta_1 l_1}\frac{l_1^{a'}}{\lambda_2-il_1-\epsilon_1}. \tag{69}$$

Since $\lambda_2 = x_2 - i/2$ with a real number $x_2 \in (-\infty,\infty)$ the denominator

$$\frac{1}{\lambda_2-il_1-\epsilon_1} = \frac{i}{l_1+1/2+ix_2-i\epsilon_1}$$

and $Re(l_1+1/2+ix_2-i\epsilon_1) \geq 1/2$ because $l_1 \geq 0$. Therefore one can use an evident integral representation

$$\frac{1}{\alpha} = \int_0^1 \frac{ds}{s}s^{\alpha} \tag{70}$$

which is valid if $Re(\alpha) > 0$. Hence the sum in expression (69) is

$$i\sum_{l_1=0}^{\infty}(-1)^{l_1}e^{-\delta_1 l_1}l_1^{a'}\int_0^1\frac{ds}{s}s^{l_1+1/2+ix_2-i\epsilon_1} \tag{71}$$

Since for $\delta_1 > 0$ the sum

$$\sum_{l_1=0}^{N'}(-1)^{l_1}e^{-\delta_1 l_1}l_1^{a'}s^{l_1} = \left(-\frac{\partial}{\partial\delta_1}\right)^{a'}\frac{1-(-e^{-\delta_1}s)^{N'+1}}{1+e^{-\delta_1}s}$$

where N' is a positive integer converges uniformly in $s\in[0,1]$ when $N'\to\infty$ then one can interchange the sum over l_1 and the integration over s. The result for (71) is as follows

$$i\int_0^1\frac{ds}{s}s^{i\lambda_2-i\epsilon_1}\left(-\frac{\partial}{\partial\delta_1}\right)^{a'}\frac{1}{1+e^{-\delta_1}s} = i\int_0^1\frac{ds}{s}s^{i\lambda_2-i\epsilon_1}\sum_{k=1}^{a'+1}C(a',k)\frac{1}{(1+e^{-\delta_1}s)^k} \tag{72}$$

where $C(a',k)$ are some rational coefficients. So the contribution of k-th term into the expression (69) is

$$\int_{C_{-1/2}}\frac{d\lambda_2}{2\pi i\,\sinh^2\pi\lambda_2}Q'(\lambda_2) \times \lim_{\delta_1\to 0+} d^{(2)}(\epsilon_1)_{\epsilon_1\to 0} i\int_0^1\frac{ds}{s}s^{i\lambda_2-i\epsilon_1}C(a',k)\frac{1}{(1+e^{-\delta_1}s)^k}. \tag{73}$$

It is not very difficult to check that the function

$$Q'(\lambda_2)\sinh\pi(\lambda_2-\epsilon_1)s^{i\lambda_2}$$

satisfies conditions (33) and (34) of Proposition 4 with $n=2$. Moreover the integral

$$\int_{C_{-1/2}}\frac{d\lambda_2}{2\pi i}e^{i\delta_2\lambda_2}\frac{\sinh\pi(\lambda_2-\epsilon_1)}{\sinh^2\pi\lambda_2}Q'(\lambda_2)\,s^{i\lambda_2}$$

converges uniformly in $s\in[0,1]$. Since it, actually, does not depend on δ_1 also and $d^{(2)}(\epsilon_1)$ given by (38) is the second derivative on ϵ_1 one can calculate the integral over λ_2 first using formulae (35–36) of Proposition 4 and then calculate the integral over s and take the limits on ϵ_1 and δ_1.

Therefore we have shown that the formula (52) is correct for $n=2$. Below we will use this formula for the concrete calculation of $P(2)$.

A generalization of our discussion to the arbitrary n is straightforward and we will not do it here.

The most efficient way of taking the integrals is as follows. Fist we apply the formula (52) using either (35) or (36) in such a way that the denominators like

$$\frac{1}{\lambda_2-\lambda_1}\qquad\text{become}\qquad\frac{i}{l_1+l_2+i(\epsilon_2-\epsilon_1)}$$

with $l_1\geq 0$ and $l_2\geq 1$. In this case we shall not face a singularities like $1/(\epsilon_2-\epsilon_1)$ and the whole expression will be analytic on $\epsilon_1,\ldots,\epsilon_n$. Hence, we can use the differential operator (58). Then after using the formula like (70) one can get rid of all such denominators and expand over ϵ-s until the order which still makes a non-zero contribution after applying the differential operator (58). It can be applied after taking all summations in (57). As the last step one should take integrals on auxiliary variables like s appeared above.

Below we shall also take all $\delta_1,\ldots,\delta_n$ to be zero at once implying the limiting procedure $\lim_{\delta_j\to 0^+}$ described above.

Now let us illustrate how the whole procedure works for a simple case $P(2)$

$$P(2)\;=\;\pi^3\int_{C_{-1/2}}\frac{d\lambda_1}{2\pi i}\int_{C_{-1/2}}\frac{d\lambda_2}{2\pi i}\,\frac{\sinh\pi(\lambda_2-\lambda_1)}{\sinh^2\pi\lambda_1\sinh^2\pi\lambda_2}\,T_2(\lambda_1,\lambda_2)\tag{74}$$

In this case it is very simple to perform the first step, namely, to get the "canonical" form (17) described in the beginning of this Section because we do not need to reduce a power of denominator in this case. Indeed,

$$T_2(\lambda_1,\lambda_2)\;=\;\frac{(\lambda_1+i)\lambda_2}{\lambda_2-\lambda_1-i}\;=\;\lambda_1+i\;+\;\frac{(\lambda_1+i)^2}{\lambda_2-\lambda_1-i}\;\sim\;\lambda_1\;-\;\frac{\lambda_1^2}{\lambda_2-\lambda_1}\tag{75}$$

where we have used property I and formula (30) from Proposition 1 of item IV. Then using the formula (32) of Corollary 3 for $m = 2,3$ one gets

$$
\begin{aligned}
-\frac{\lambda_1^2}{\lambda_2-\lambda_1} &\sim -\frac{i}{3}\left(\frac{3i^2\lambda_1+i^3}{\lambda_2-\lambda_1}+\lambda_2^2+\lambda_2(\lambda_1+i)+(\lambda_1+i)^2\right)\\
&\sim \frac{i\lambda_1}{\lambda_2-\lambda_1}-\frac{1}{3}\frac{1}{\lambda_2-\lambda_1}-\frac{i}{3}(i\lambda_1)\sim\frac{1}{2}\frac{1}{\lambda_2-\lambda_1}-\frac{1}{3}\frac{1}{\lambda_2-\lambda_1}+\frac{1}{3}\lambda_1\\
&=\frac{1}{3}\lambda_1+\frac{1}{6}\frac{1}{\lambda_2-\lambda_1}.
\end{aligned}
$$

Substituting it to the formula (75) we get

$$
T_2(\lambda_1,\lambda_2)\ \sim\ T_2^c(\lambda_1,\lambda_2) \tag{76}
$$

where

$$
T_2^c(\lambda_1,\lambda_2)\ =\ \frac{4}{3}\lambda_1\ +\ \frac{1}{6}\frac{1}{\lambda_2-\lambda_1} \tag{77}
$$

and this is the "canonical" form for T_2 i.e. the polynomials $P_0^{(2)}$ and $P_1^{(2)}$ from (17-18) are equal to $4/3\lambda_1$ and $1/6$ respectively.

Let us take the integral from the first term using the formula (57), our comments about the limiting procedure like in formulae (35-36) of Proposition 4

$$
\begin{aligned}
J_0^{(2)} &= \pi^3\int_{C_{-1/2}}\frac{d\lambda_1}{2\pi i}\int_{C_{-1/2}}\frac{d\lambda_2}{2\pi i}\frac{\sinh\pi(\lambda_2-\lambda_1)}{\sinh^2\pi\lambda_1\sinh^2\pi\lambda_2}\frac{4}{3}\lambda_1\\
&= D^{(2)}\frac{4}{3}\sum_{l_1=0}^{\infty}(-1)^{l_1}(il_1+\epsilon_1)\sum_{l_2=0}^{\infty}(-1)^{l_2}=D^{(2)}\frac{4}{3}(i\rho(1)+\epsilon_1\rho(0))\rho(0)\\
&= D^{(2)}\frac{4}{3}(-\frac{i}{4}+\frac{\epsilon_1}{2})\frac{1}{2}=\left(\frac{\partial}{\partial\epsilon_1}-\frac{\partial}{\partial\epsilon_2}\right)_{\epsilon_1,\epsilon_2\to 0}\frac{2}{3}(-\frac{i}{4}+\frac{\epsilon_1}{2})=\frac{1}{3}
\end{aligned}
\tag{78}
$$

where we have used formulae (66) for $\rho(b)$ given by (65) implying the limiting procedure as it was explained above.

The second integral is treated as it was described in item VI with the help of the integral representation (70)

$$
\begin{aligned}
J_1^{(2)} &= \pi^3\int_{C_{-1/2}}\frac{d\lambda_1}{2\pi i}\int_{C_{-1/2}}\frac{d\lambda_2}{2\pi i}\frac{\sinh\pi(\lambda_2-\lambda_1)}{\sinh^2\pi\lambda_1\sinh^2\pi\lambda_2}\frac{1}{6}\frac{1}{\lambda_2-\lambda_1}\\
&= D^{(2)}\frac{-1}{6}\sum_{l_1=1}^{\infty}\sum_{l_2=0}^{\infty}\frac{(-1)^{l_1+l_2}}{il_2+il_1+\epsilon_2-\epsilon_1}=D^{(2)}\frac{i}{6}\sum_{l_1=1}^{\infty}\sum_{l_2=0}^{\infty}\frac{(-1)^{l_1+l_2}}{l_1+l_2+i(\epsilon_1-\epsilon_2)}\\
&= D^{(2)}\frac{i}{6}\int_0^1\frac{ds}{s}\sum_{l_1=1}^{\infty}\sum_{l_2=0}^{\infty}(-1)^{l_1+l_2}s^{l_1+l_2+i(\epsilon_1-\epsilon_2)}=D^{(2)}\frac{-i}{6}\int_0^1 ds\,\frac{s^{i(\epsilon_1-\epsilon_2)}}{(1+s)^2}\\
&= (\frac{\partial}{\partial\epsilon_1}-\frac{\partial}{\partial\epsilon_2})_{\epsilon_1,\epsilon_2\to 0}\frac{-i}{6}\int_0^1 ds\,\frac{s^{i(\epsilon_1-\epsilon_2)}}{(1+s)^2}=\frac{1}{3}\int_0^1 ds\,\frac{\ln s}{(1+s)^2}=-\frac{\ln 2}{3}.
\end{aligned}
\tag{79}
$$

Summing up two answers (78) and (79) we get the result

$$P(2) = J_0^{(2)} + J_1^{(2)} = \frac{1}{3} - \frac{\ln(2)}{3}. \tag{80}$$

which coincides with the formula (7).

In Appendices A and B we shall derive formulae (8) and (9) for $P(3)$ and $P(4)$ respectively. In the end of this Section let us note that both results (8) and (9) are expressed in terms of the logarithmic function and the Riemann zeta function of odd arguments and do not depend on polylogarithms like, for example, $\mathrm{Li}_4(1/2)$. All coefficients before those functions in (6–9) are rational. Also they do not contain any powers of π which could be considered as Riemann zeta functions of even arguments, see formula (4) from Introduction.

Our conjecture is that the final answer for any $P(n)$ will also be expressed in terms of logarithm $\ln 2$ and Riemann zeta functions $\zeta(k)$ with odd integers k and with rational coefficients.

In fact, this conjecture is intimately connected with our hypothesis from the beginning of this Section that the function T_n (14) can be reduced to the "canonical" form. Looking at the "canonical" form (17) one can conclude that only Riemann zeta functions and their products can enter into the final answer because all the denominators in the r.h.s. of (17) are split out. It means that after applying the formula (57) the multiple summation can be performed by pairs, say, $\sum_{l_{2k-1}}$ and $\sum_{l_{2k}}$. Each pair of these summations results in some combination of zeta functions.

3. Conclusion

We want to emphasize an interesting connection between integrable and disordered models. In order to describe correlations in integrable models one can use integrable integral operators [23]. On the other hand Tracy and Widom showed that in matrix models the distribution of eigenvalues and level spacing can be described by the integral operators, belonging to the same integrable class [22].

Our current work supports this link between integrable models and chaotic models. Riemann zeta function appears in the description of both kind of models.

Let us repeat that the main result of this paper is the calculation of $P(3)$ and $P(4)$ (8–9) by means of the multi-integral representation (11). The fact that only the logarithm $\ln 2$ and Riemann zeta function with odd arguments participate in the answers for $P(1), \ldots, P(4)$ and with rational coefficients before these functions allows us to suppose that this is the general property of $P(n)$. One could compare the calculation of $P(n)$ with the many-loop calculation of self-energy diagrams in the renormalizable quantum field theory which can also be expressed in terms of ζ functions of odd arguments [1].

Unfortunately, so far we have not got even a conjecture for $P(n)$ but we believe that it is not an unsolvable problem. May be already after calculation of $P(5)$ one could guess the right formula for a generic case $P(n)$. It would give an answer to the question discussed in the previous section, namely, the question about the law of decay of $P(n)$ when n tends to infinity.

Also it would be interesting to generalize above results to the XXZ spin chain. Some interesting conjectures were recently invented by Razumov and Stroganov [13] for the special case of the XXZ model with $\Delta = -1/2$. These conjectures would be supported if it were possible to get $P(n)$ from the general integral representation obtained by the RIMS group [7].

4. Acknowledgements

The authors would like to thank B. McCoy, A. Razumov, M. Shiroishi, Yu. Stroganov, M. Takahashi, L. Takhtajan for useful discussions. This research has been supported by the NSF grant PHY-9988566 and by INTAS Grant No. 01-561.

Appendix A

Here we discuss in detail the calculation of $P(3)$ performed by means of the general procedure described above.

As was pointed out in the beginning of Section 2 the first step should be a reduction of the function $T_3(\lambda_1, \lambda_2, \lambda_3)$ to the form (17) which we have called the "canonical" form. In comparison with the case $P(2)$ here we should reduce the power of the denominator in $T_3(\lambda_1, \lambda_2, \lambda_3)$. To do this we will use the formula (27) from item II. Namely,

$$T_3(\lambda_1, \lambda_2, \lambda_3) = \frac{(\lambda_1 + i)^2(\lambda_2 + i)\lambda_2\lambda_3^2}{(\lambda_2 - \lambda_1 - i)(\lambda_3 - \lambda_1 - i)(\lambda_3 - \lambda_2 - i)} = I_1^{(3)} + I_2^{(3)} + I_3^{(3)} \tag{81}$$

where

$$I_1^{(3)} = i\frac{(\lambda_1 + i)^2(\lambda_2 + i)\lambda_2\lambda_3^2}{(\lambda_3 - \lambda_1 - i)(\lambda_3 - \lambda_2 - i)}, \tag{82}$$

$$I_2^{(3)} = i\frac{(\lambda_1 + i)^2(\lambda_2 + i)\lambda_2\lambda_3^2}{(\lambda_2 - \lambda_1 - i)(\lambda_3 - \lambda_1 - i)}, \tag{83}$$

$$I_3^{(3)} = -i\frac{(\lambda_1 + i)^2(\lambda_2 + i)\lambda_2\lambda_3^2}{(\lambda_2 - \lambda_1 - i)(\lambda_3 - \lambda_2 - i)}. \tag{84}$$

Due to the $1 \leftrightarrow 2$ symmetry of the denominator the first term $I_1^{(3)}$ can be simplified as follows

$$\begin{aligned} I_1^{(3)} &\sim -\frac{(\lambda_1 + i)(\lambda_2 + i)\lambda_2\lambda_3^2}{(\lambda_3 - \lambda_1 - i)(\lambda_3 - \lambda_2 - i)} \sim \frac{(\lambda_1 + i)(\lambda_2 + i)\lambda_3^2}{\lambda_3 - \lambda_1 - i} \\ &= (\lambda_1 + i)(\lambda_2 + i)(\lambda_1 + \lambda_3 + i) + \frac{(\lambda_1 + i)^3(\lambda_2 + i)}{\lambda_3 - \lambda_1 - i} \sim \lambda_1^2\lambda_2 - \frac{(\lambda_1 + i)^3(\lambda_3 + i)}{\lambda_2 - \lambda_1 - i}. \end{aligned} \tag{85}$$

The denominator of the second term (83) has the symmetry under the transposition $2 \leftrightarrow 3$. Therefore it can also be simplified

$$I_2^{(3)} \sim -\frac{(\lambda_1+i)^2\lambda_2\lambda_3^2}{(\lambda_2-\lambda_1-i)(\lambda_3-\lambda_1-i)} \sim \frac{(\lambda_1+i)^2\lambda_2\lambda_3}{\lambda_2-\lambda_1-i}$$
$$= -(\lambda_1+i)^2\lambda_3 - \frac{(\lambda_1+i)^3\lambda_3}{\lambda_2-\lambda_1-i} \sim \lambda_1^2\lambda_2 - \frac{(\lambda_1+i)^3\lambda_3}{\lambda_2-\lambda_1-i} \quad (86)$$

The third term (84) is treated as follows

$$I_3^{(3)} = -i(\lambda_1+i)^2(\lambda_2+i)(\lambda_3+\lambda_2+i) - -i\frac{(\lambda_1+i)^2(\lambda_2+i)^3}{\lambda_3-\lambda_2-i}$$
$$+i\frac{(\lambda_1+i)^3\lambda_3^2}{\lambda_2-\lambda_1-i} - i\frac{(\lambda_1+i)^3\lambda_3^3}{(\lambda_2-\lambda_1-i)(\lambda_3-\lambda_2-i)}$$
$$\sim -\lambda_1^2\lambda_2 - \frac{(\lambda_1+i)^3(\lambda_3+i)^2}{\lambda_2-\lambda_1-i}$$
$$+i\frac{(\lambda_1+i)^3\lambda_3^2}{\lambda_2-\lambda_1-i} - i\frac{(\lambda_1+i)^3\lambda_3^3}{(\lambda_2-\lambda_1-i)(\lambda_3-\lambda_2-i)} \quad (87)$$

Now adding up all the three terms together we get

$$T_3(\lambda_1,\lambda_2,\lambda_3) \sim \lambda_1^2\lambda_2 - i\frac{(\lambda_1+i)^3\lambda_3^3}{(\lambda_2-\lambda_1-i)(\lambda_3-\lambda_2-i)}$$
$$\sim -\lambda_2\lambda_3^2 - i\frac{(\lambda_1+i)^3\lambda_3^3}{(\lambda_2-\lambda_1-i)(\lambda_3-\lambda_2-i)}. \quad (88)$$

Let us note that up to this moment we have used only the symmetry property (26) from item I, formula (27) from item II and a simple algebra.

Now we would like to use the formula (30) of the Proposition 1 for $m=3$ and again apply the transposition formula (26)

$$T_3(\lambda_1,\lambda_2,\lambda_3) \sim -\lambda_2\lambda_3^2 - i\frac{\lambda_1^3(\lambda_3+i)^3}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_2)}$$
$$\sim -\lambda_2\lambda_3^2 + i\frac{(\lambda_1+i)^3\lambda_3^3}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_2)} = -\lambda_2\lambda_3^2 + i\frac{\lambda_1^3\lambda_3^3+3i\lambda_1^2\lambda_3^3-3\lambda_1\lambda_3^3-i\lambda_3^3}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_2)}$$
$$\sim -\lambda_2\lambda_3^2 - \frac{3\lambda_1^2\lambda_3^2+3i\lambda_1\lambda_3(\lambda_3+\lambda_2)-\lambda_3^2-\lambda_3\lambda_2-\lambda_2^2}{\lambda_2-\lambda_1}$$
$$\sim -\lambda_2\lambda_3^2 - \frac{3\lambda_1^2\lambda_3^2+3i\lambda_1\lambda_3^2+3i\lambda_1^2\lambda_3-\lambda_3^2-\lambda_3\lambda_1-\lambda_1^2}{\lambda_2-\lambda_1}. \quad (89)$$

Now we can reduce the power of λ_1 in the numerator of the second term (89) by applying formula (32) of Corollary 3 from item IV. Doing this we finally get the "canonical form" (17–18) of T_3

$$T_3(\lambda_1,\lambda_2,\lambda_3) \sim T_3^c(\lambda_1,\lambda_2,\lambda_3) = P_0^{(3)} + \frac{P_1^{(3)}}{\lambda_2-\lambda_1} \quad (90)$$

where the polynomials $P_0^{(3)}$ and $P_1^{(3)}$ are as follows

$$P_0^{(3)} = -2\lambda_2\lambda_3^2, \quad P_1^{(3)} = \frac{1}{3} - i\lambda_1 - i\lambda_3 - 2\lambda_1\lambda_3. \tag{91}$$

Let us note that if we express variables λ_j through the real variables x_j via $\lambda_j = x_j - i/2$ in order to get the polynomials $\tilde{P}_j^{(3)}$ (see the formula (21)) we get especially simple formulae, namely,

$$\tilde{P}_0^{(3)} = -2x_2x_3^2, \quad \tilde{P}_1^{(3)} = -\frac{1}{6} - 2x_1x_3. \tag{92}$$

So, the function

$$\tilde{T}_3^c(x_1,x_2,x_3) = \tilde{P}_0^{(3)} + \frac{\tilde{P}_1^{(3)}}{x_2 - x_1}. \tag{93}$$

is odd i.e.

$$\tilde{T}_3^c(-x_1,-x_2,-x_3) = -\tilde{T}_3^c(x_1,x_2,x_3)$$

as it should be according to the formula (23) from the beginning of Section 2.

Now we are ready to calculate the integral in order to get the result for $P(3)$:

$$P(3) = \prod_{j=1}^{3} \int_{C_{-1/2}} \frac{d\lambda_j}{2\pi i}\, U_3(\lambda_1,\lambda_2,\lambda_3)\, T_3(\lambda_1,\lambda_2,\lambda_3) = J_0^{(3)} + J_1^{(3)} \tag{94}$$

where

$$J_0^{(3)} = \prod_{j=1}^{3} \int_{C_{-1/2}} \frac{d\lambda_j}{2\pi i}\, U_3(\lambda_1,\lambda_2,\lambda_3)\, P_0^{(3)},$$

$$J_1^{(3)} = \prod_{j=1}^{3} \int_{C_{-1/2}} \frac{d\lambda_j}{2\pi i}\, U_3(\lambda_1,\lambda_2,\lambda_3)\, \frac{P_1^{(3)}}{\lambda_2 - \lambda_1}. \tag{95}$$

Using formulae (91), (57), (58), (66) we get

$$\begin{aligned}
J_0^{(3)} &= \prod_{j=1}^{3} \int_{C_{-1/2}} \frac{d\lambda_j}{2\pi i}\, U_3(\lambda_1,\lambda_2,\lambda_3)\,(-2)\lambda_2\lambda_3^2 \\
&= D^{(3)} \sum_{l_1=0}^{\infty}(-1)^{l_1} \sum_{l_2=0}^{\infty}(-1)^{l_2} \sum_{l_3=0}^{\infty}(-1)^{l_3}(-2)(il_2+\epsilon_2)(il_3+\epsilon_3)^2 \\
&= D^{(3)} \sum_{l_1=0}^{\infty}(-1)^{l_1} \sum_{l_2=0}^{\infty}(-1)^{l_2} \sum_{l_3=0}^{\infty}(-1)^{l_3}(-2)\epsilon_2\epsilon_3^2 \\
&= \frac{1}{8} \prod_{1\le k<j\le 3} \left(\frac{\partial}{\partial\epsilon_k} - \frac{\partial}{\partial\epsilon_j}\right)_{\vec{\epsilon}\to 0} (-2)\epsilon_2\epsilon_3^2 = \frac{1}{4}.
\end{aligned} \tag{96}$$

To calculate the second term $J_1^{(3)}$ we should also use the integral representation (70)

$$
\begin{aligned}
J_1^{(3)} &= \prod_{j=1}^{3} \int_{C_{-1/2}} \frac{d\lambda_j}{2\pi i} U_3(\lambda_1, \lambda_2, \lambda_3) \frac{\frac{1}{3} - i\lambda_1 - i\lambda_3 - 2\lambda_1\lambda_3}{\lambda_2 - \lambda_1} \\
&= -D^{(3)} \sum_{l_1=1}^{\infty} (-1)^{l_1} \sum_{l_2=0}^{\infty} (-1)^{l_2} \sum_{l_3=0}^{\infty} (-1)^{l_3} \\
&\quad \times \frac{\frac{1}{3} - i(-il_1 + \epsilon_1) - i(il_3 + \epsilon_3) - 2(-il_1 + \epsilon_1)(il_3 + \epsilon_3)}{il_2 + il_1 + \epsilon_2 - \epsilon_1} \\
&= i\, D^{(3)} \sum_{l_1=1}^{\infty} (-1)^{l_1} \sum_{l_2=0}^{\infty} (-1)^{l_2} \sum_{l_3=0}^{\infty} (-1)^{l_3} \\
&\quad \times \frac{\frac{1}{3} - l_1 + l_3 - i\epsilon_1 - i\epsilon_3 - 2(l_1 + i\epsilon_1)(l_3 - i\epsilon_3)}{l_1 + l_2 + i(\epsilon_1 - \epsilon_2)} \\
&= i\, D^{(3)} \sum_{l_1=1}^{\infty} (-1)^{l_1} \sum_{l_2=0}^{\infty} (-1)^{l_2} \frac{-\frac{1}{12} - i\epsilon_3/2 + l_1 i\epsilon_3 - \epsilon_1\epsilon_3}{l_1 + l_2 + i(\epsilon_1 - \epsilon_2)} \\
&= i\, D^{(3)} \int_0^1 \frac{ds}{s} s^{i(\epsilon_1 - \epsilon_2)} \sum_{l_1=1}^{\infty} \sum_{l_2=0}^{\infty} \left(-\frac{1}{12} - i\epsilon_3/2 + l_1 i\epsilon_3 - \epsilon_1\epsilon_3\right)(-s)^{l_1+l_2} \\
&= i\, D^{(3)} \int_0^1 \frac{ds}{s} s^{i(\epsilon_1 - \epsilon_2)} \left(\frac{(-s)}{(1+s)^2} \left(-\frac{1}{12} - i\epsilon_3/2 - \epsilon_1\epsilon_3\right) + \frac{(-s)}{(1+s)^3} i\epsilon_3 \right) \\
&= \frac{i}{2} \prod_{1 \le k < j \le 3} \left(\frac{\partial}{\partial \epsilon_k} - \frac{\partial}{\partial \epsilon_j} \right)_{\vec{\epsilon} \to 0} \int_0^1 \frac{ds}{(1+s)^2} \\
&\quad \times \left(\frac{1}{12} (-3\epsilon_1^2 \epsilon_2 + 3\epsilon_1 \epsilon_2^2)(-i) \frac{\ln^3 s}{6} + i \ln s\, \epsilon_1^2 \epsilon_3 \right) \\
&= \int_0^1 ds \frac{\ln s}{(1+s)^2} - \frac{1}{12} \int_0^1 ds \frac{\ln^3 s}{(1+s)^2} = -\ln 2 + \frac{3}{8}\zeta(3).
\end{aligned}
\tag{97}
$$

Summing up $J_0^{(3)}$ and $J_1^{(3)}$ we get the final answer (8)

$$
P(3) = J_0^{(3)} + J_1^{(3)} = \frac{1}{4} - \ln 2 + \frac{3}{8}\zeta(3). \tag{98}
$$

Appendix B

Here we will calculate $P(4)$. This case is much more complicated than the previous ones. But using our general technique we will try to simplify our discussion as much as possible.

As above the first step of our scheme is to get the "canonical" form for the function $T_4(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$.

But before start doing this let us list some useful formulae which can be derived from Proposition 1 and Corollary 3 in the case $n = 4$

$$\frac{\lambda_1^3}{\lambda_2-\lambda_1} \sim \frac{-\frac{3}{2}i\lambda_1^2 + \lambda_1 + \frac{i}{4}}{\lambda_2-\lambda_1} + \frac{1}{2}\lambda_2^2 + \frac{i}{2}\lambda_2 \tag{99}$$

$$\frac{\lambda_1^4}{\lambda_2-\lambda_1} \sim \frac{-\lambda_1^2 - i\lambda_1 + \frac{3}{10}}{\lambda_2-\lambda_1} + \frac{3}{5}\lambda_2^3 + \frac{1}{5}\lambda_1\lambda_2^2 + \frac{2}{5}\lambda_2 \tag{100}$$

$$\frac{\lambda_1^5}{\lambda_2-\lambda_1} \sim \frac{-\frac{\lambda_1}{6} - \frac{i}{12}}{\lambda_2-\lambda_1} + \frac{1}{3}\lambda_1\lambda_2^3 - \frac{4}{3}i\lambda_2^3 + \frac{13}{6}\lambda_2^2 + \frac{4}{3}i\lambda_2 - \frac{1}{3} \tag{101}$$

$$\begin{aligned}\frac{\lambda_1^6}{\lambda_2-\lambda_1} &\sim \frac{-\frac{1}{2}\lambda_1^2 - \frac{1}{2}i\lambda_1 - \frac{1}{7}}{\lambda_2-\lambda_1} \\ &+ \frac{1}{7}\lambda_1^2\lambda_2^3 - \frac{6}{7}i\lambda_1\lambda_2^3 + \frac{2}{7}\lambda_2^3 + \frac{10}{7}\lambda_1\lambda_2^2 - \frac{25}{14}i\lambda_2^2 + \frac{31}{14}\lambda_2\end{aligned} \tag{102}$$

$$\frac{(\lambda_1+i)^3}{\lambda_2-\lambda_1} \sim \frac{\frac{3}{2}i\lambda_1^2 - 2\lambda_1 - \frac{3}{4}i}{\lambda_2-\lambda_1} + \frac{1}{2}\lambda_2^2 + \frac{i}{2}\lambda_2 \tag{103}$$

$$\frac{(\lambda_1+i)^4}{\lambda_2-\lambda_1} \sim \frac{-\lambda_1^2 - i\lambda_1 + \frac{3}{10}}{\lambda_2-\lambda_1} + \frac{3}{5}\lambda_2^3 + \frac{1}{5}\lambda_1\lambda_2^2 + 2i\lambda_2^2 - \frac{8}{5}\lambda_2 \tag{104}$$

$$\begin{aligned}\frac{(\lambda_1+i)^5}{\lambda_2-\lambda_1} &\sim \frac{-\frac{\lambda_1}{6} - \frac{i}{12}}{\lambda_2-\lambda_1} \\ &+ \frac{1}{3}\lambda_1\lambda_2^3 + \frac{5}{3}i\lambda_2^3 + i\lambda_1\lambda_2^2 - \frac{17}{6}\lambda_2^2 - \frac{5}{3}i\lambda_2 - \frac{1}{3}\end{aligned} \tag{105}$$

Here we have omitted an arbitrary function $g(\lambda_3, \lambda_4)$ that can be multiplied on the r.h.s. and the l.h.s. of (99-105) as in formula (32).

Now we can start our derivation. Fortunately, due to the formula (29) of the observation III which in this case looks as follows

$$\frac{T_4(\lambda_1,\lambda_2,\lambda_3,\lambda_4)}{T_3(\lambda_1,\lambda_2,\lambda_3)} = \frac{(\lambda_1+i)(\lambda_2+i)(\lambda_3+i)\lambda_4^3}{(\lambda_4-\lambda_1-i)(\lambda_4-\lambda_2-i)(\lambda_4-\lambda_3-i)} \tag{106}$$

and has a symmetry under any permutation of the variables $\lambda_1, \lambda_2, \lambda_3$ we can use the result (88) that was obtained by means of the symmetry and simple algebra. So, after the application of the formulae (88) and (106) we can start with the following expression

$$\begin{aligned}T_4(\lambda_1,\lambda_2,\lambda_3,\lambda_4) &\sim -\frac{(\lambda_1+i)(\lambda_2+i)\lambda_2(\lambda_3+i)\lambda_3^2\lambda_4^3}{(\lambda_4-\lambda_1-i)(\lambda_4-\lambda_2-i)(\lambda_4-\lambda_3-i)} \\ &- i\frac{(\lambda_1+i)^4(\lambda_2+i)(\lambda_3+i)\lambda_3^3\lambda_4^3}{(\lambda_2-\lambda_1-i)(\lambda_3-\lambda_2-i)(\lambda_4-\lambda_1-i)(\lambda_4-\lambda_2-i)(\lambda_4-\lambda_3-i)}\end{aligned} \tag{107}$$

The denominator of the first term in the r.h.s. of (107) is symmetric under permutation of $\lambda_1, \lambda_2, \lambda_3$. Therefore it can be simplified as follows

$$
\begin{aligned}
&-\frac{(\lambda_1+i)(\lambda_2+i)\lambda_2(\lambda_3+i)\lambda_3^2\lambda_4^3}{(\lambda_4-\lambda_1-i)(\lambda_4-\lambda_2-i)(\lambda_4-\lambda_3-i)}\\
&\sim \frac{(\lambda_1+i)(\lambda_2+i)\lambda_2(\lambda_3+i)(\lambda_3+\lambda_4+i)\lambda_4^3}{(\lambda_4-\lambda_1-i)(\lambda_4-\lambda_2-i)}\\
&\sim -\frac{(\lambda_1+i)(\lambda_2+i)(\lambda_3+i)(\lambda_3+\lambda_4+i)\lambda_4^3}{\lambda_4-\lambda_1-i}\\
&\sim -\frac{(\lambda_1+i)(\lambda_2+i)(\lambda_3+i)\lambda_3\lambda_4^3}{\lambda_4-\lambda_1-i}\\
&= (\lambda_2+i)(\lambda_3+i)\lambda_3\lambda_4^3 - \frac{(\lambda_2+i)(\lambda_3+i)\lambda_3\lambda_4^4}{\lambda_4-\lambda_1-i}\\
&\sim \lambda_2\lambda_3^2\lambda_4^3 - \frac{\lambda_1^4(\lambda_3+i)(\lambda_4+i)\lambda_4}{\lambda_2-\lambda_1+i}\\
&\sim \lambda_2\lambda_3^2\lambda_4^3 - \frac{\lambda_1^4(\lambda_3\lambda_4^2+i\lambda_4^2-\lambda_4)}{\lambda_2-\lambda_1+i}\\
&\sim \lambda_2\lambda_3^2\lambda_4^3 + \frac{(\lambda_1+i)^4(\lambda_3\lambda_4^2+i\lambda_4^2-\lambda_4)}{\lambda_2-\lambda_1}\\
&\sim \frac{8}{5}\lambda_2\lambda_3^2\lambda_4^3 - \frac{\left(\lambda_1^2+i\lambda_1-\frac{3}{10}\right)(\lambda_3\lambda_4^2+i\lambda_4^2-\lambda_4)}{\lambda_2-\lambda_1}
\end{aligned}
\tag{108}
$$

The latter formula was obtained with the help of formula (104). In fact, it is nothing but the "canonical" form (17) for the first term in (107).

Now let us treat the second term. Using formulae (27) and (28) we can write it down as follows

$$
\begin{aligned}
&-i\frac{(\lambda_1+i)^4(\lambda_2+i)(\lambda_3+i)\lambda_3^3\lambda_4^3}{(\lambda_2-\lambda_1-i)(\lambda_3-\lambda_2-i)(\lambda_4-\lambda_1-i)(\lambda_4-\lambda_2-i)(\lambda_4-\lambda_3-i)}\\
&\quad = -i(\lambda_1+i)^4(\lambda_2+i)(\lambda_3+i)\lambda_3^3\lambda_4^3\Bigg\{\\
&\qquad \frac{1}{2}\frac{1}{(\lambda_3-\lambda_2-i)(\lambda_4-\lambda_1-i)(\lambda_4-\lambda_3-i)}\\
&\qquad -\frac{1}{2}\frac{1}{(\lambda_2-\lambda_1-i)(\lambda_4-\lambda_1-i)(\lambda_4-\lambda_3-i)}\\
&\qquad -\frac{1}{2}\frac{1}{(\lambda_2-\lambda_1-i)(\lambda_3-\lambda_2-i)(\lambda_4-\lambda_1-i)}\\
&\qquad -\frac{1}{2}\frac{1}{(\lambda_2-\lambda_1-i)(\lambda_3-\lambda_2-i)(\lambda_4-\lambda_3-i)}\\
&\qquad -\frac{1}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)(\lambda_4-\lambda_1-i)}\\
&\qquad -\frac{1}{(\lambda_3-\lambda_2-i)(\lambda_4-\lambda_1-i)(\lambda_4-\lambda_2-i)}
\end{aligned}
$$

$$- \frac{1}{(\lambda_3-\lambda_1)(\lambda_3-\lambda_2)(\lambda_4-\lambda_3-i)}$$
$$+ \frac{1}{(\lambda_2-\lambda_1-i)(\lambda_4-\lambda_2-i)(\lambda_4-\lambda_3-i)}$$
$$+ \frac{1}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_2)(\lambda_4-\lambda_2-i)}$$
$$\left. + \frac{1}{(\lambda_2-\lambda_1-i)(\lambda_3-\lambda_2-i)(\lambda_4-\lambda_2-i)} \right\}$$

Let us enumerate all ten terms as they enter here and make some appropriate transformations for each of them:

$$I_1^{(4)} = -\frac{i(\lambda_1+i)^4(\lambda_2+i)(\lambda_3+i)\lambda_3^3\lambda_4^3}{2(\lambda_3-\lambda_2-i)(\lambda_4-\lambda_1-i)(\lambda_4-\lambda_3-i)}$$
$$\underset{\substack{1\to4\\2\to2\\3\to1\\4\to3}}{\sim} -\frac{i(\lambda_1+i)\lambda_1^3(\lambda_2+i)\lambda_3^3(\lambda_4+i)^4}{2(\lambda_2-\lambda_1+i)(\lambda_3-\lambda_1-i)(\lambda_4-\lambda_3+i)}; \tag{109}$$

$$I_2^{(4)} = \frac{i}{2}\frac{(\lambda_1+i)^4(\lambda_2+i)(\lambda_3+i)\lambda_3^3\lambda_4^3}{(\lambda_2-\lambda_1-i)(\lambda_4-\lambda_1-i)(\lambda_4-\lambda_3-i)}$$
$$\underset{3\leftrightarrow4}{\sim} \frac{i}{2}\frac{(\lambda_1+i)^4(\lambda_2+i)\lambda_3^3(\lambda_4+i)\lambda_4^3}{(\lambda_2-\lambda_1-i)(\lambda_3-\lambda_1-i)(\lambda_4-\lambda_3+i)}; \tag{110}$$

$$I_3^{(4)} = \frac{i}{2}\frac{(\lambda_1+i)^4(\lambda_2+i)(\lambda_3+i)\lambda_3^3\lambda_4^3}{(\lambda_2-\lambda_1-i)(\lambda_3-\lambda_2-i)(\lambda_4-\lambda_1-i)}$$
$$\underset{\substack{1\to1\\2\to3\\3\to4\\4\to2}}{\sim} \frac{i}{2}\frac{(\lambda_1+i)^4\lambda_2^3(\lambda_3+i)(\lambda_4+i)\lambda_4^3}{(\lambda_2-\lambda_1-i)(\lambda_3-\lambda_1-i)(\lambda_4-\lambda_3-i)}; \tag{111}$$

$$I_4^{(4)} = \frac{i}{2}\frac{(\lambda_1+i)^4(\lambda_2+i)(\lambda_3+i)\lambda_3^3\lambda_4^3}{(\lambda_2-\lambda_1-i)(\lambda_3-\lambda_2-i)(\lambda_4-\lambda_3-i)}$$
$$\underset{1\leftrightarrow2}{\sim} \frac{i}{2}\frac{(\lambda_1+i)(\lambda_2+i)^4(\lambda_3+i)\lambda_3^3\lambda_4^3}{(\lambda_2-\lambda_1+i)(\lambda_3-\lambda_1-i)(\lambda_4-\lambda_3-i)}; \tag{112}$$

$$I_5^{(4)} = i\frac{(\lambda_1+i)^4(\lambda_2+i)(\lambda_3+i)\lambda_3^3\lambda_4^3}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)(\lambda_4-\lambda_1-i)}; \tag{113}$$

$$I_6^{(4)} = i\frac{(\lambda_1+i)^4(\lambda_2+i)(\lambda_3+i)\lambda_3^3\lambda_4^3}{(\lambda_3-\lambda_2-i)(\lambda_4-\lambda_1-i)(\lambda_4-\lambda_2-i)}$$
$$\underset{\substack{1\to4\\2\to1\\3\to2\\4\to3}}{\sim} i\frac{(\lambda_1+i)(\lambda_2+i)\lambda_2^3\lambda_3^3(\lambda_4+i)^4}{(\lambda_2-\lambda_1-i)(\lambda_3-\lambda_1-i)(\lambda_4-\lambda_3+i)}; \tag{114}$$

$$I_7^{(4)} = i\frac{(\lambda_1+i)^4(\lambda_2+i)(\lambda_3+i)\lambda_3^3\lambda_4^3}{(\lambda_3-\lambda_1)(\lambda_3-\lambda_2)(\lambda_4-\lambda_3-i)}$$
$$\underset{1\leftrightarrow 3}{\sim} -i\frac{(\lambda_1+i)\lambda_1^3(\lambda_2+i)(\lambda_3+i)^4\lambda_4^3}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)(\lambda_4-\lambda_1-i)}; \quad (115)$$

$$I_8^{(4)} = -\frac{i(\lambda_1+i)^4(\lambda_2+i)(\lambda_3+i)\lambda_3^3\lambda_4^3}{(\lambda_2-\lambda_1-i)(\lambda_4-\lambda_2-i)(\lambda_4-\lambda_3-i)}$$
$$\underset{\substack{1\leftrightarrow 2\\3\leftrightarrow 4}}{\sim} -\frac{i(\lambda_1+i)(\lambda_2+i)^4\lambda_3^3(\lambda_4+i)\lambda_4^3}{(\lambda_2-\lambda_1+i)(\lambda_3-\lambda_1-i)(\lambda_4-\lambda_3+i)}; \quad (116)$$

$$I_9^{(4)} = -i\frac{(\lambda_1+i)^4(\lambda_2+i)(\lambda_3+i)\lambda_3^3\lambda_4^3}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_2)(\lambda_4-\lambda_2-i)}$$
$$\underset{\substack{1\to 3\\2\to 1\\3\to 2\\4\to 4}}{\sim} i\frac{(\lambda_1+i)(\lambda_2+i)\lambda_2^3(\lambda_3+i)^4\lambda_4^3}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)(\lambda_4-\lambda_1-i)}; \quad (117)$$

$$I_{10}^{(4)} = -\frac{i(\lambda_1+i)^4(\lambda_2+i)(\lambda_3+i)\lambda_3^3\lambda_4^3}{(\lambda_2-\lambda_1-i)(\lambda_3-\lambda_2-i)(\lambda_4-\lambda_2-i)}$$
$$\underset{1\leftrightarrow 2}{\sim} -\frac{i(\lambda_1+i)(\lambda_2+i)^4(\lambda_3+i)\lambda_3^3\lambda_4^3}{(\lambda_2-\lambda_1+i)(\lambda_3-\lambda_1-i)(\lambda_4-\lambda_1-i)}. \quad (118)$$

Let us note that after these transformations $I_j^{(4)}$ have two kinds of the denominators, namely, $I_5^{(4)}, I_7^{(4)}, I_9^{(4)}, I_{10}^{(4)}$ have denominators of the form

$$\frac{1}{(\lambda_2-\lambda_1-ia)(\lambda_3-\lambda_1-ib)(\lambda_4-\lambda_1-ic)}$$

with a set of integers a, b, c, while the denominators of the rest of them are of the form

$$\frac{1}{(\lambda_2-\lambda_1-ia')(\lambda_3-\lambda_1-ib')(\lambda_4-\lambda_3-ic')}$$

with some other set of integers a', b', c'. Moreover, some of the terms $I_j^{(4)}$ have just coinciding denominators like, for example, $I_2^{(4)}$ and $I_6^{(4)}$. Nevertheless sometimes it will be more convenient to treat them separately.

Let us start with the first group which is easier to treat. Since, the denominator of $I_5^{(4)}$ has the $2 \leftrightarrow 3$ symmetry we can simplify it as follows

$$I_5^{(4)} \sim i\frac{(\lambda_1+i)^4(\lambda_2+i)(\lambda_3+i)(\lambda_3^2+\lambda_3\lambda_1+\lambda_1^2)\lambda_4^3}{(\lambda_2-\lambda_1)(\lambda_4-\lambda_1-i)} \quad (119)$$

where we have used the trivial identity

$$\lambda_3^3 = (\lambda_3-\lambda_1)(\lambda_3^2+\lambda_3\lambda_1+\lambda_1^2) + \lambda_1^3 \quad (120)$$

and the fact that the second term in the r.h.s. of (120) does not contribute into $I_5^{(4)}$.

Then summing up (115) and (117) we get

$$I_7^{(4)} + I_9^{(4)} \sim i\frac{(\lambda_1+i)(\lambda_2+i)(\lambda_1^2+\lambda_1\lambda_2+\lambda_2^2)(\lambda_3+i)^4\lambda_4^3}{(\lambda_3-\lambda_1)(\lambda_4-\lambda_1-i)}$$
$$\underset{1\leftrightarrow 2}{\sim} -i\frac{(\lambda_1+i)(\lambda_2+i)^4(\lambda_3+i)(\lambda_1^2+\lambda_1\lambda_3+\lambda_3^2)\lambda_4^3}{(\lambda_2-\lambda_1)(\lambda_4-\lambda_1-i)} \qquad (121)$$

Now adding the r.h.s. of (119) to (121) we get

$$\begin{aligned}
&I_5^{(4)} + I_7^{(4)} + I_9^{(4)} \\
&\sim -i(\lambda_1+i)(\lambda_2+i)(\lambda_3+i) \\
&\qquad\times\frac{((\lambda_1+i)^2+(\lambda_1+i)(\lambda_2+i)+(\lambda_2+i)^2)(\lambda_3^2+\lambda_3\lambda_1+\lambda_1^2)\lambda_4^3}{\lambda_4-\lambda_1-i} \\
&\sim 3\,\frac{(\lambda_1+i)(\lambda_2+i)(\lambda_3+i)(\lambda_1+\lambda_2+i)(\lambda_3^2+\lambda_3\lambda_1+\lambda_1^2)\lambda_4^3}{\lambda_4-\lambda_1-i} \\
&= -3\,(\lambda_2+i)(\lambda_3+i)(\lambda_1+\lambda_2+i)(\lambda_3^2+\lambda_3\lambda_1+\lambda_1^2)\lambda_4^3 \\
&\qquad + 3\,\frac{(\lambda_2+i)(\lambda_3+i)(\lambda_1+\lambda_2+i)(\lambda_3^2+\lambda_3\lambda_1+\lambda_1^2)\lambda_4^4}{\lambda_4-\lambda_1-i} \\
&\sim 3\,\frac{(\lambda_2+i)(\lambda_3+i)(\lambda_1+\lambda_2+i)(\lambda_3^2+\lambda_3\lambda_1+\lambda_1^2)\lambda_4^4}{\lambda_4-\lambda_1-i} \\
&\sim 3\,\frac{(\lambda_1+i)(\lambda_2+i)(\lambda_3+i)^2\lambda_3\lambda_4^4}{\lambda_4-\lambda_1-i} \\
&\qquad + 3\,\frac{(\lambda_2+i)(\lambda_3+i)\lambda_3(\lambda_2\lambda_3+1)\lambda_4^4}{\lambda_4-\lambda_1-i} \\
&\sim -3\lambda_2\lambda_3^2\lambda_4^3 + 3\,\frac{(\lambda_2+i)(\lambda_3+i)^2\lambda_3\lambda_4^5}{\lambda_4-\lambda_1-i} \\
&\qquad + 3\,\frac{(\lambda_2+i)(\lambda_3+i)\lambda_3(\lambda_2\lambda_3+1)\lambda_4^4}{\lambda_4-\lambda_1-i} \\
&\sim -3\lambda_2\lambda_3^2\lambda_4^3 + 3\,\frac{(\lambda_2+i)(\lambda_3+i)^2\lambda_3(\lambda_4+i)^5}{\lambda_4-\lambda_1} \\
&\qquad + 3\,\frac{(\lambda_2+i)(\lambda_3+i)\lambda_3(\lambda_2\lambda_3+1)(\lambda_4+i)^4}{\lambda_4-\lambda_1} \\
&\underset{\substack{1\to 2\\ 2\to 3\\ 3\to 4\\ 4\to 1}}{\sim} -3\lambda_2\lambda_3^2\lambda_4^3 - 3\,\frac{(\lambda_1+i)^5(\lambda_3+i)(\lambda_4+i)^2\lambda_4}{\lambda_2-\lambda_1} \\
&\qquad - 3\,\frac{(\lambda_1+i)^4(\lambda_3+i)(\lambda_4+i)\lambda_4(\lambda_3\lambda_4+1)}{\lambda_2-\lambda_1} \\
&\sim -\frac{17}{10}\lambda_2\lambda_3^2\lambda_4^3 + \frac{(\frac{\lambda_1}{2}+\frac{i}{4})(\lambda_3+i)(\lambda_4+i)^2\lambda_4}{\lambda_2-\lambda_1} \\
&\qquad + \frac{(3\lambda_1^2+3i\lambda_1-\frac{9}{10})(\lambda_3+i)(\lambda_4+i)\lambda_4(\lambda_3\lambda_4+1)}{\lambda_2-\lambda_1}
\end{aligned}$$

$$
\sim -\frac{17}{10}\lambda_2\lambda_3^2\lambda_4^3 + \frac{(\frac{\lambda_1}{2} + \frac{i}{4})(\lambda_3\lambda_4^3 + i\lambda_4^3 + 2i\lambda_3\lambda_4^2 - 2\lambda_4^2 - i\lambda_4)}{\lambda_2 - \lambda_1} + \frac{(3\lambda_1^2 + 3i\lambda_1 - \frac{9}{10})(\lambda_3^2\lambda_4^3 + i\lambda_3\lambda_4^3 + i\lambda_4^2 - \lambda_4)}{\lambda_2 - \lambda_1} \tag{122}
$$

In fact, it is the "canonical" form. To get this expression we have used symmetries, formula (30) for $m = 4$ and $m = 5$ of Proposition 1 and relations (104) and (105).

Now let us treat $I_{10}^{(4)}$

$$
\begin{aligned}
I_{10}^{(4)} &= -i\frac{(\lambda_1 + i)(\lambda_2 + i)^4(\lambda_3 + i)\lambda_3^3\lambda_4^3}{(\lambda_2 - \lambda_1 + i)(\lambda_3 - \lambda_1 - i)(\lambda_4 - \lambda_1 - i)} \\
&\sim -i\frac{(\lambda_1 + i)(\lambda_2 + i)^4\lambda_3^3\lambda_4^3}{(\lambda_2 - \lambda_1 + i)(\lambda_4 - \lambda_1 - i)} \\
&= i\frac{(\lambda_2 + i)^4\lambda_3^3\lambda_4^3}{\lambda_2 - \lambda_1 + i} - i\frac{(\lambda_2 + i)^4\lambda_3^3\lambda_4^4}{(\lambda_2 - \lambda_1 + i)(\lambda_4 - \lambda_1 - i)} \\
&\sim -i\frac{(\lambda_2 + i)^4\lambda_3^3\lambda_4^4}{(\lambda_2 - \lambda_1 + i)(\lambda_4 - \lambda_1 - i)} \sim -i\frac{\lambda_2^4\lambda_3^3(\lambda_4 + i)^4}{(\lambda_2 - \lambda_1)(\lambda_4 - \lambda_1)} \\
&= i\frac{\lambda_3^3(\lambda_2^4\lambda_4^4 + 4i\lambda_2^3\lambda_4^4 - 6\lambda_2^2\lambda_4^4 - 4i\lambda_2\lambda_4^4 + \lambda_4^4)}{(\lambda_2 - \lambda_1)(\lambda_4 - \lambda_1)} \\
&\sim i\frac{\lambda_3^3(4i\lambda_2^3\lambda_4^3 - 6\lambda_2^2\lambda_4^2(\lambda_4 + \lambda_1))}{\lambda_2 - \lambda_1} \\
&\qquad - i\frac{\lambda_3^3(4i\lambda_2\lambda_4(\lambda_4^2 + \lambda_4\lambda_1 + \lambda_1^2) + \lambda_4^3 + \lambda_4^2\lambda_1 + \lambda_4\lambda_1^2 + \lambda_1^3)}{\lambda_2 - \lambda_1} \\
&\sim i\frac{\lambda_3^3(-6\lambda_1\lambda_2^2\lambda_4^2 - 4i\lambda_1\lambda_2\lambda_4^2 - 4i\lambda_1^2\lambda_2\lambda_4 + \lambda_1\lambda_4^2 + \lambda_1^2\lambda_4 + \lambda_1^3)}{\lambda_2 - \lambda_1} \\
&= i\lambda_3^3(-6\lambda_1^2\lambda_4^2 - 4i\lambda_1\lambda_4^2 - 4i\lambda_1^2\lambda_4) \\
&\qquad + i\frac{\lambda_3^3(-6\lambda_1^3\lambda_4^2 - 4i\lambda_1^2\lambda_4^2 - 4i\lambda_1^3\lambda_4 + \lambda_1\lambda_4^2 + \lambda_1^2\lambda_4 + \lambda_1^3)}{\lambda_2 - \lambda_1} \\
&\sim i\frac{\lambda_3^3(\lambda_4^2(5i\lambda_1^2 - 5\lambda_1 - \frac{3}{2}i) + \lambda_4(-5\lambda_1^2 + 4i\lambda_1 + 1) - \frac{3}{2}i\lambda_1^2 + \lambda_1 + \frac{i}{4})}{\lambda_2 - \lambda_1} \\
&\qquad + \lambda_3^3(-3)\lambda_4^2 i\lambda_2 + i\lambda_3^3(-4)i\lambda_4\frac{1}{2}\lambda_2^2 \\
&\sim -\lambda_2\lambda_3^2\lambda_4^3 \\
&\quad + \frac{(5\lambda_1^2 + 5i\lambda_1 - \frac{3}{2})\lambda_3^2\lambda_4^3 + (5i\lambda_1^2 - 4\lambda_1 - i)\lambda_3\lambda_4^3 + (-\frac{3}{2}\lambda_1^2 - i\lambda_1 + \frac{1}{4})\lambda_4^3}{\lambda_2 - \lambda_1}
\end{aligned} \tag{123}
$$

Here the symmetry, formula (30) for $m = 4$ and relation (99) were used.

Now we shall treat the other six terms $I_1^{(4)}, I_2^{(4)}, I_3^{(4)}, I_4^{(4)}, I_6^{(4)}$ and $I_8^{(4)}$ given by the expressions (109), (110), (111), (112), (114) and (116) respectively. Here we shall proceed in two steps. As a first step we will reduce all these six terms to a following form

$$I_j^{(4)} \sim \frac{A_j}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)(\lambda_4-\lambda_3)} + \frac{B_j}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)} + \frac{C_j}{(\lambda_2-\lambda_1)(\lambda_4-\lambda_3)} + \frac{D_j}{\lambda_2-\lambda_1} \quad (124)$$

where $j = 1,2,3,4,6,8$ and A_j, B_j, C_j, D_j are some polynomials. Then we shall sum up all the six results and get the "canonical" form for the sum obtained.

Let us start with expression (109) for term $I_1^{(4)}$:

$$\begin{aligned}
I_1^{(4)} &= -\frac{i}{2}\frac{(\lambda_1+i)\lambda_1^3(\lambda_2+i)\lambda_3^3(\lambda_4+i)^4}{(\lambda_2-\lambda_1+i)(\lambda_3-\lambda_1-i)(\lambda_4-\lambda_3+i)} \\
&= -\frac{i}{2}\frac{(\lambda_1^2+\lambda_1(\lambda_2+i)+(\lambda_2+i)^2)(\lambda_2+i)\lambda_3^3(\lambda_4+i)^4}{\lambda_4-\lambda_3+i} \\
&\quad + \frac{i}{2}\frac{(\lambda_1^2+\lambda_1(\lambda_2+i)+(\lambda_2+i)^2)(\lambda_2+i)\lambda_3^4(\lambda_4+i)^4}{(\lambda_3-\lambda_1-i)(\lambda_4-\lambda_3+i)} \\
&\quad + \frac{i}{2}\frac{(\lambda_2+i)^4\lambda_3^3(\lambda_4+i)^3}{\lambda_2-\lambda_1+i} - \frac{i}{2}\frac{(\lambda_2+i)^4\lambda_3^4(\lambda_4+i)^3}{(\lambda_2-\lambda_1-i)(\lambda_3-\lambda_1+i)} \\
&\quad - \frac{i}{2}\frac{(\lambda_1+i)(\lambda_2+i)^4\lambda_3^4(\lambda_4+i)^3}{(\lambda_2-\lambda_1+i)(\lambda_3-\lambda_1-i)(\lambda_4-\lambda_3+i)} \\
&\sim \frac{i}{2}\frac{(\lambda_1^2+\lambda_1(\lambda_2+i)+(\lambda_2+i)^2)(\lambda_2+i)\lambda_3^3\lambda_4^4}{\lambda_4-\lambda_3} \\
&\quad - \frac{i}{2}\frac{(\lambda_1^2+\lambda_1(\lambda_2+i)+(\lambda_2+i)^2)(\lambda_2+i)(\lambda_3+i)^4(\lambda_4+i)^4}{(\lambda_3-\lambda_1)(\lambda_4-\lambda_3)} \\
&\quad - \frac{i}{2}\frac{\lambda_2^4\lambda_3^3(\lambda_4+i)^3}{\lambda_2-\lambda_1} - \frac{i}{2}\frac{\lambda_2^4(\lambda_3+i)^4(\lambda_4+i)^3}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)} \\
&\quad - \frac{i}{2}\frac{(\lambda_1+i)\lambda_2^4(\lambda_3+i)^4(\lambda_4+i)^3}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)(\lambda_4-\lambda_3)} \\
&\sim \frac{i}{2}\frac{\lambda_1^3\lambda_2^4(\lambda_3^2+\lambda_3(\lambda_4+i)+(\lambda_4+i)^2)(\lambda_4+i)}{\lambda_2-\lambda_1} \\
&\quad + \frac{i}{2}\frac{(\lambda_1+i)^4(\lambda_2+i)^4(\lambda_3^2+\lambda_3(\lambda_4+i)+(\lambda_4+i)^2)(\lambda_4+i)}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)} \\
&\quad - \frac{i}{2}\frac{\lambda_2^4\lambda_3^3(\lambda_4+i)^3}{\lambda_2-\lambda_1} - \frac{i}{2}\frac{\lambda_2^4(\lambda_3+i)^4(\lambda_4+i)^3}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)} \\
&\quad - \frac{i}{2}\frac{(\lambda_1+i)\lambda_2^4(\lambda_3+i)^4(\lambda_4+i)^3}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)(\lambda_4-\lambda_3)} \qquad (125)
\end{aligned}$$

where we have used symmetries and the formula (30) with $m = 4$.

For the next term $I_2^{(4)}$ we start with the expression (110):

$$I_2^{(4)} \sim \frac{i}{2}\frac{(\lambda_1+i)^4(\lambda_2+i)\lambda_3^3(\lambda_4+i)\lambda_4^3}{(\lambda_2-\lambda_1-i)(\lambda_3-\lambda_1-i)(\lambda_4-\lambda_3+i)}$$

$$
\begin{aligned}
&= -\frac{i}{2}\frac{(\lambda_1+i)^4(\lambda_2+i)(\lambda_3^2+\lambda_3(\lambda_4+i)+(\lambda_4+i)^2)(\lambda_4+i)\lambda_4^3}{(\lambda_2-\lambda_1-i)(\lambda_3-\lambda_1-i)} \\
&\quad + \frac{i}{2}\frac{(\lambda_1+i)^4(\lambda_2+i)(\lambda_4+i)^4\lambda_4^3}{(\lambda_2-\lambda_1-i)(\lambda_3-\lambda_1-i)(\lambda_4-\lambda_3+i)} \\
&\sim \frac{i}{2}\frac{\lambda_1^4(\lambda_2+i)(\lambda_3^2+\lambda_3(\lambda_4+i)+(\lambda_4+i)^2)(\lambda_4+i)\lambda_4^3}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)} \\
&\quad + \frac{i}{2}\frac{\lambda_1^4(\lambda_2+i)\lambda_4^4(\lambda_4-i)^3}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)(\lambda_4-\lambda_3)}.
\end{aligned}
\tag{126}
$$

The term $I_3^{(4)}$ given by (111) is also simple to treat:

$$
\begin{aligned}
I_3^{(4)} &= \frac{i}{2}\frac{(\lambda_1+i)^4\lambda_2^3(\lambda_3+i)(\lambda_4+i)\lambda_4^3}{(\lambda_2-\lambda_1-i)(\lambda_3-\lambda_1-i)(\lambda_4-\lambda_3-i)} \\
&= -\frac{i}{2}\frac{(\lambda_1+i)^4\lambda_2^3(\lambda_4+i)\lambda_4^3}{(\lambda_2-\lambda_1-i)(\lambda_3-\lambda_1-i)} \\
&\quad + \frac{i}{2}\frac{(\lambda_1+i)^4\lambda_2^3(\lambda_4+i)\lambda_4^4}{(\lambda_2-\lambda_1-i)(\lambda_3-\lambda_1-i)(\lambda_4-\lambda_3-i)} \\
&\sim \frac{i}{2}\frac{\lambda_1^4\lambda_2^3(\lambda_4+i)\lambda_4^3}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)} + \frac{i}{2}\frac{\lambda_1^4\lambda_2^3(\lambda_4+2i)(\lambda_4+i)^4}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)(\lambda_4-\lambda_3)}.
\end{aligned}
\tag{127}
$$

The fourth term $I_4^{(4)}$ (112) demands more work in order to reduce it to the form (124):

$$
\begin{aligned}
I_4^{(4)} &\sim \frac{i}{2}\frac{(\lambda_1+i)(\lambda_2+i)^4(\lambda_3+i)\lambda_3^3\lambda_4^3}{(\lambda_2-\lambda_1+i)(\lambda_3-\lambda_1-i)(\lambda_4-\lambda_3-i)} \\
&= -\frac{i}{2}\frac{(\lambda_1+i)(\lambda_2+i)^4\lambda_3^3\lambda_4^3}{(\lambda_2-\lambda_1+i)(\lambda_3-\lambda_1-i)} \\
&\quad + \frac{i}{2}\frac{(\lambda_1+i)(\lambda_2+i)^4\lambda_3^3\lambda_4^4}{(\lambda_2-\lambda_1+i)(\lambda_3-\lambda_1-i)(\lambda_4-\lambda_3-i)} \\
&= \frac{i}{2}\frac{(\lambda_2+i)^4\lambda_3^3\lambda_4^3}{(\lambda_2-\lambda_1+i)} - \frac{i}{2}\frac{(\lambda_2+i)^4\lambda_3^4\lambda_4^3}{(\lambda_2-\lambda_1+i)(\lambda_3-\lambda_1-i)} \\
&\quad + \frac{i(\lambda_1+i)(\lambda_2+i)^4(\lambda_3^2+\lambda_3(\lambda_1+i)+(\lambda_1+i)^2)\lambda_4^4}{2(\lambda_2-\lambda_1+i)(\lambda_4-\lambda_3-i)} \\
&\quad + \frac{i(\lambda_1+i)^4(\lambda_2+i)^4\lambda_4^4}{2(\lambda_2-\lambda_1+i)(\lambda_3-\lambda_1-i)(\lambda_4-\lambda_3-i)} \\
&\sim -\frac{i}{2}\frac{\lambda_2^4(\lambda_3+i)^4\lambda_4^3}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)} \\
&\quad + \frac{i}{2}\frac{(\lambda_1+i)\lambda_2^4(\lambda_3^2+\lambda_3(\lambda_1+i)+(\lambda_1+i)^2)(\lambda_4+i)^4}{(\lambda_2-\lambda_1)(\lambda_4-\lambda_3)} \\
&\quad - \frac{i}{2}\frac{\lambda_1^4\lambda_2^4(\lambda_4+i)^4}{(\lambda_2-\lambda_1+i)(\lambda_3-\lambda_1)(\lambda_4-\lambda_3)}
\end{aligned}
$$

$$
\begin{aligned}
&\sim -\frac{i}{2}\frac{\lambda_2^4(\lambda_3+i)^4\lambda_4^3}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)} \\
&\qquad +\frac{i}{2}\frac{(\lambda_1+i)\lambda_2^4(\lambda_3^2+\lambda_3(\lambda_1+i)+(\lambda_1+i)^2)(\lambda_4+i)^4}{(\lambda_2-\lambda_1)(\lambda_4-\lambda_3)} \\
&\qquad +\frac{i(\lambda_1^3+\lambda_1^2(\lambda_2+i)+\lambda_1(\lambda_2+i)^2+(\lambda_2+i)^3)\lambda_2^4(\lambda_4+i)^4}{2(\lambda_3-\lambda_1)(\lambda_4-\lambda_3)} \\
&\qquad -\frac{i(\lambda_2+i)^4\lambda_2^4(\lambda_4+i)^4}{2(\lambda_2-\lambda_1+i)(\lambda_3-\lambda_1)(\lambda_4-\lambda_3)} \\
&\sim -\frac{i}{2}\frac{\lambda_2^4(\lambda_3+i)^4\lambda_4^3}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)} \\
&\qquad +\frac{i}{2}\frac{(\lambda_1+i)\lambda_2^4(\lambda_3^2+\lambda_3(\lambda_1+i)+(\lambda_1+i)^2)(\lambda_4+i)^4}{(\lambda_2-\lambda_1)(\lambda_4-\lambda_3)} \\
&\qquad -\frac{i}{2}\frac{(\lambda_2+i)^4\lambda_4^4(\lambda_3^3+\lambda_3^2(\lambda_4+i)+\lambda_3(\lambda_4+i)^2+(\lambda_4+i)^3)}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)} \\
&\qquad +\frac{i}{2}\frac{\lambda_2^4(\lambda_2-i)^4(\lambda_4+i)^4}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)(\lambda_4-\lambda_3)}.
\end{aligned} \tag{128}
$$

Now we treat $I_6^{(4)}$ given by (114) as follows:

$$
\begin{aligned}
I_6^{(4)} &\sim i\,\frac{(\lambda_1+i)(\lambda_2+i)\lambda_2^3\lambda_3^3(\lambda_4+i)^4}{(\lambda_2-\lambda_1-i)(\lambda_3-\lambda_1-i)(\lambda_4-\lambda_3+i)} \\
&= i\,\frac{(\lambda_1+i)(\lambda_2+i)(\lambda_2^2+\lambda_2(\lambda_1+i)+(\lambda_1+i)^2)\lambda_3^3(\lambda_4+i)^4}{(\lambda_3-\lambda_1-i)(\lambda_4-\lambda_3+i)} \\
&\qquad + i\,\frac{(\lambda_1+i)^4(\lambda_2+i)\lambda_3^3(\lambda_4+i)^4}{(\lambda_2-\lambda_1-i)(\lambda_3-\lambda_1-i)(\lambda_4-\lambda_3+i)} \\
&= -i\,\frac{(\lambda_2+i)(\lambda_2^2+\lambda_2(\lambda_1+i)+(\lambda_1+i)^2)\lambda_3^3(\lambda_4+i)^4}{\lambda_4-\lambda_3+i} \\
&\qquad + i\,\frac{(\lambda_2+i)(\lambda_2^2+\lambda_2(\lambda_1+i)+(\lambda_1+i)^2)\lambda_3^4(\lambda_4+i)^4}{(\lambda_3-\lambda_1-i)(\lambda_4-\lambda_3+i)} \\
&\qquad + i\,\frac{(\lambda_1+i)^4(\lambda_2+i)\lambda_3^3(\lambda_4+i)^4}{(\lambda_2-\lambda_1-i)(\lambda_3-\lambda_1-i)(\lambda_4-\lambda_3+i)} \\
&= i\,\frac{(\lambda_2+i)(\lambda_2^2+\lambda_2(\lambda_1+i)+(\lambda_1+i)^2)\lambda_3^3\lambda_4^4}{\lambda_4-\lambda_3} \\
&\qquad + i\,\frac{(\lambda_2+i)(\lambda_2^2+\lambda_2(\lambda_1+i)+(\lambda_1+i)^2)(\lambda_3+i)^4(\lambda_4+i)^4}{(\lambda_3-\lambda_1)(\lambda_4-\lambda_3)} \\
&\qquad + i\,\frac{\lambda_1^4(\lambda_2+i)\lambda_3^3\lambda_4^4}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)(\lambda_4-\lambda_3)}
\end{aligned}
$$

$$\sim\ -i\,\frac{\lambda_1^3\lambda_2^4(\lambda_3+i)(\lambda_3^2+\lambda_3(\lambda_4+i)+(\lambda_4+i)^2)}{\lambda_2-\lambda_1}$$
$$-\,i\,\frac{(\lambda_1+i)^4(\lambda_2+i)^4(\lambda_4^2+\lambda_4(\lambda_3+i)+(\lambda_3+i)^2)(\lambda_4+i)}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)}$$
$$+\,i\,\frac{\lambda_1^4(\lambda_2+i)\lambda_3^3\lambda_4^4}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)(\lambda_4-\lambda_3)}. \tag{129}$$

The last term $I_8^{(4)}$ (116) is more simple

$$\begin{aligned}
I_8^{(4)} &\sim -i\,\frac{(\lambda_1+i)(\lambda_2+i)^4\lambda_3^3(\lambda_4+i)\lambda_4^3}{(\lambda_2-\lambda_1+i)(\lambda_3-\lambda_1-i)(\lambda_4-\lambda_3+i)}\\
&= i\,\frac{(\lambda_2+i)^4\lambda_3^3(\lambda_4+i)\lambda_4^3}{(\lambda_2-\lambda_1+i)(\lambda_4-\lambda_3+i)}\\
&\quad -i\,\frac{(\lambda_2+i)^4\lambda_3^4(\lambda_4+i)\lambda_4^3}{(\lambda_2-\lambda_1+i)(\lambda_3-\lambda_1-i)(\lambda_4-\lambda_3+i)}\\
&= i\,\frac{(\lambda_2+i)^4\lambda_3^3\lambda_4^3}{\lambda_2-\lambda_1+i} + i\,\frac{(\lambda_2+i)^4\lambda_3^4\lambda_4^3}{(\lambda_2-\lambda_1+i)(\lambda_4-\lambda_3+i)}\\
&\quad -i\,\frac{(\lambda_2+i)^4\lambda_3^4(\lambda_4+i)\lambda_4^3}{(\lambda_2-\lambda_1+i)(\lambda_3-\lambda_1-i)(\lambda_4-\lambda_3+i)}\\
&\sim\ i\,\frac{\lambda_2^4(\lambda_3+i)^4\lambda_4^3}{(\lambda_2-\lambda_1)(\lambda_4-\lambda_3)} - i\,\frac{\lambda_2^4(\lambda_3+i)^4(\lambda_4+i)\lambda_4^3}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)(\lambda_4-\lambda_3)}.
\end{aligned} \tag{130}$$

Now we are prepared to perform the next our step. Namely, we will gather all the six results (125–130) into the form like (124)

$$I_1^{(4)} + I_2^{(4)} + I_3^{(4)} + I_4^{(4)} + I_6^{(4)} + I_8^{(4)}$$
$$\sim \frac{A}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)(\lambda_4-\lambda_3)} + \frac{B}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)}$$
$$+ \frac{C}{(\lambda_2-\lambda_1)(\lambda_4-\lambda_3)} + \frac{D}{\lambda_2-\lambda_1} \tag{131}$$

where

$$\begin{aligned}
A \;=\; &-\frac{i}{2}(\lambda_1+i)\lambda_2^4(\lambda_3+i)^4(\lambda_4+i)^3 + \frac{i}{2}\lambda_1^4(\lambda_2+i)\lambda_4^4(\lambda_4-i)^3\\
&+ \frac{i}{2}\lambda_1^4\lambda_2^3(\lambda_4+2i)(\lambda_4+i)^4 + \frac{i}{2}\lambda_2^4(\lambda_2-i)(\lambda_4-i)^4\\
&+ i\lambda_1^4(\lambda_2+i)\lambda_3^3\lambda_4^4 - i\lambda_2^4(\lambda_3+i)^4(\lambda_4+i)\lambda_4^3;
\end{aligned} \tag{132}$$

$$\begin{aligned}
B \;=\; &\frac{i}{2}(\lambda_1+i)^4(\lambda_2+i)^4(\lambda_3^2+\lambda_3(\lambda_4+i)+(\lambda_4+i)^2)(\lambda_4+i)\\
&- i\lambda_2^4(\lambda_3+i)^4(\lambda_4+i)^3\\
&+ \frac{i}{2}\lambda_1^4(\lambda_2+i)(\lambda_3^2+\lambda_3(\lambda_4+i)+(\lambda_4+i)^2)(\lambda_4+i)\lambda_4^3 + i\lambda_1^4\lambda_2^3(\lambda_4+i)\lambda_4^3\\
&+ i(\lambda_1+i)^4(\lambda_2+i)^4(\lambda_4^2+\lambda_4(\lambda_3+i)+(\lambda_3+i)^2)(\lambda_4+i);
\end{aligned} \tag{133}$$

$$C = \frac{i}{2}(\lambda_1+i)\lambda_2^4(\lambda_3^2+\lambda_3(\lambda_1+i)+(\lambda_1+i)^2)(\lambda_4+i)^4 + i\lambda_2^4(\lambda_3+i)^4\lambda_4^3; \quad (134)$$

$$D = \frac{i}{2}\lambda_1^3\lambda_2^4(\lambda_3^2+\lambda_3(\lambda_4+i)+(\lambda_4+i)^2)(\lambda_4+i) - i\lambda_2^4\lambda_3^3(\lambda_4+i)^3 - i\lambda_1^3\lambda_2^4(\lambda_3^2+\lambda_3(\lambda_4+i)+(\lambda_4+i)^2)(\lambda_3+i). \quad (135)$$

Now we want to get the "canonical" form (17) for the expression (131). To do this we actively used the program *MATHEMATICA* because the calculations are straightforward but become more cumbersome. Let us outline our further actions.

It is more convenient to start with the first term in the r.h.s. of the formula (131)

$$\frac{A}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)(\lambda_4-\lambda_3)}. \quad (136)$$

We shall use the fact that the denominator is antisymmetric under the substitution

$$\lambda_1 \leftrightarrow \lambda_3$$
$$\lambda_2 \leftrightarrow \lambda_4$$

Since A is a polynomial given by (132) then the simplification procedure of the term (136) is as follows. If in the expression (132) one faces a monomial

$$\lambda_1^{i_1}\lambda_2^{i_2}\lambda_3^{i_3}\lambda_4^{i_4} \quad (137)$$

where without loss of generality $i_4 \geq i_2$ one can apply the evident identity

$$\lambda_4^{i_4} = \lambda_4^{i_2}\lambda_3^{i_4-i_2} + \begin{cases} (\lambda_4-\lambda_3)\sum_{k=0}^{i_4-i_2-1}\lambda_4^{i_2+k}\lambda_3^{i_4-i_2-1-k}, & i_4 > i_2; \\ 0, & i_4 = i_2. \end{cases} \quad (138)$$

Therefore if $i_4 > i_2$ then

$$\frac{\lambda_1^{i_1}\lambda_2^{i_2}\lambda_3^{i_3}\lambda_4^{i_4}}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)(\lambda_4-\lambda_3)} = \frac{\lambda_1^{i_1}\lambda_2^{i_2}\lambda_3^{i_4-i_2+i_3}\lambda_4^{i_2}}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)(\lambda_4-\lambda_3)} + \sum_{k=0}^{i_4-i_2-1}\frac{\lambda_1^{i_1}\lambda_2^{i_2}\lambda_3^{i_4-i_2+i_3-1-k}\lambda_4^{i_2+k}}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)}. \quad (139)$$

Let us note that the second term in (139) gives rise into the second term "B" in the r.h.s. of the formula (131). If $i_4 = i_2$ then only the first term in (139) survives. The first term in (139) is symmetric under the transposition $\lambda_2 \leftrightarrow \lambda_4$. If $i_1+i_2 = i_3+i_4$ then it is also symmetric under the transposition $\lambda_1 \leftrightarrow \lambda_3$ and the first term is "weakly" equivalent to zero according to the formula (25). If $i_1+i_2 < i_3+i_4$ the following "weak" equality is valid

$$\frac{\lambda_1^{i_1}\lambda_2^{i_2}\lambda_3^{i_4-i_2+i_3}\lambda_4^{i_2}}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)(\lambda_4-\lambda_3)} \sim \frac{1}{2}\frac{(\lambda_1^{i_1}\lambda_3^{i_3+i_4-i_2}-\lambda_3^{i_1}\lambda_1^{i_3+i_4-i_2})\lambda_2^{i_2}\lambda_4^{i_2}}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)(\lambda_4-\lambda_3)} = \frac{1}{2}\sum_{k=0}^{i_3+i_4-i_2-i_1-1}\frac{\lambda_1^{i_1+k}\lambda_2^{i_2}\lambda_3^{i_3+i_4-i_2-1-k}\lambda_4^{i_2}}{(\lambda_2-\lambda_1)(\lambda_4-\lambda_3)}. \quad (140)$$

For the case $i_1 + i_2 > i_3 + i_4$ the sum in (140) should be substituted by

$$-\frac{1}{2} \sum_{k=0}^{i_1+i_2-i_3-i_4-1} \frac{\lambda_1^{i_1-1-k} \lambda_2^{i_2} \lambda_3^{i_3+i_4-i_2+k} \lambda_4^{i_2}}{(\lambda_2-\lambda_1)(\lambda_4-\lambda_3)}. \tag{141}$$

In both cases, namely, if $i_1 + i_2 \neq i_3 + i_4$ the sum in (140) or (141) gives rise into the third term "C" in the r.h.s. of (131).

Performing this procedure for all monomials of the form (137) participating in the polynomial A given by the formula (132) one can arrive at the formula

$$\begin{aligned} I_1^{(4)} &+ I_2^{(4)} + I_3^{(4)} + I_4^{(4)} + I_6^{(4)} + I_8^{(4)} \\ &\sim \frac{B'}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)} + \frac{C'}{(\lambda_2-\lambda_1)(\lambda_4-\lambda_3)} + \frac{D}{\lambda_2-\lambda_1} \end{aligned} \tag{142}$$

with some other polynomials B' and C'.

Due to the $2 \leftrightarrow 3$ symmetry of the denominator of the first term in (142) one can treat any monomial $\lambda_1^{j_1}\lambda_2^{j_2}\lambda_3^{j_3}\lambda_4^{j_4}$ participating in B' as follows

$$\frac{\lambda_1^{j_1}\lambda_2^{j_2}\lambda_3^{j_3}\lambda_4^{j_4}}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)} \sim \sum_{k=0}^{j_3-j_2-1} \frac{\lambda_1^{j_1}\lambda_2^{j_2+k}\lambda_3^{j_3-k-1}\lambda_4^{j_4}}{\lambda_2-\lambda_1} \tag{143}$$

where without loss of generality it is implied that $j_3 > j_2$ because if $j_2 = j_3$ then due to the $2 \leftrightarrow 3$ symmetry and the formula (25) this term would make zero contribution. So the r.h.s. of the (143) gives rise into the third term in (142). Using this one can treat the whole first term in (142).

For any monomial $\lambda_1^{k_1}\lambda_2^{k_2}\lambda_3^{k_3}\lambda_4^{k_4}$ participating in C' of the second term of the expression (142) one can write

$$\begin{aligned} \frac{\lambda_1^{k_1}\lambda_2^{k_2}\lambda_3^{k_3}\lambda_4^{k_4}}{(\lambda_2-\lambda_1)(\lambda_4-\lambda_3)} &= \sum_{l=0}^{k_2-1} \frac{\lambda_1^{k_1+l}\lambda_2^{k_2}\lambda_3^{k_3-1-l}\lambda_4^{k_4}}{\lambda_4-\lambda_3} \\ &+ \sum_{l=0}^{k_4-1} \frac{\lambda_1^{k_1+k_2}\lambda_3^{k_3+l}\lambda_4^{k_4-1-l}}{\lambda_2-\lambda_1} + \frac{\lambda_1^{k_1+k_2}\lambda_3^{k_3+k_4}}{(\lambda_2-\lambda_1)(\lambda_4-\lambda_3)} \end{aligned} \tag{144}$$

$$\begin{aligned} \sim &\sum_{l=0}^{k_2-1} \frac{\lambda_1^{k_1}\lambda_2^{k_4}\lambda_3^{k_1+l}\lambda_4^{k_2-1-l}}{\lambda_2-\lambda_1} \\ &+ \sum_{l=0}^{k_4-1} \frac{\lambda_1^{k_1+k_2}\lambda_3^{k_3+l}\lambda_4^{k_4-1-l}}{\lambda_2-\lambda_1} + \frac{\lambda_1^{k_1+k_2}\lambda_3^{k_3+k_4}}{(\lambda_2-\lambda_1)(\lambda_4-\lambda_3)} \end{aligned} \tag{145}$$

where for the first sum of (144) we have applied transformation

$$\lambda_1 \leftrightarrow \lambda_3,$$
$$\lambda_2 \leftrightarrow \lambda_4.$$

So the both the first term and the second term in (145) give rise into the third "D" term in (142) while the third term in (145) gives rise to the second term "C" in (142). Proceeding this way one can treat the whole expression

$$\frac{C'}{(\lambda_2-\lambda_1)(\lambda_4-\lambda_3)}.$$

As a result of performing this scheme one can arrive at the expression

$$I_1^{(4)} + I_2^{(4)} + I_3^{(4)} + I_4^{(4)} + I_6^{(4)} + I_8^{(4)} \sim \frac{C''}{(\lambda_2-\lambda_1)(\lambda_4-\lambda_3)} + \frac{D'}{\lambda_2-\lambda_1} \tag{146}$$

where C'' is a polynomial of two variable λ_1 and λ_3.

Now with the help of the identity

$$\lambda_1^{i_1}\lambda_2^{i_2} = \lambda_1^{i_1+i_2} + \begin{cases} (\lambda_2-\lambda_1)\sum_{k=0}^{i_2-1}\lambda_1^{i_1+k}\lambda_2^{i_2-1-k}, & i_2>0; \\ 0, & i_2=0 \end{cases} \tag{147}$$

one can reduce the second term in (146) to the form

$$\frac{D'}{\lambda_2-\lambda_1} \sim \frac{D''}{\lambda_2-\lambda_1} + E$$

where D'' is a polynomial of λ_1, λ_3 and λ_4 and E is some polynomial.

What is left now is to reduce the power of the polynomials C'', D'' and E with the help of formulae (30), (31), (32) or (99–102). We should also use the fact that if we do the substitution $\lambda_j \to x_j - i/2$ there is a restriction that the function should be even under the transformation $\{x_1,x_2,x_3,x_4\} \to \{-x_1,-x_2,-x_3,-x_4\}$ according to the formulae (19), (20) and (21).

As a result of all these actions described above we get the "canonical" form for the sum

$$I_1^{(4)} + I_2^{(4)} + I_3^{(4)} + I_4^{(4)} + I_6^{(4)} + I_8^{(4)}$$
$$\sim \frac{C'''}{(\lambda_2-\lambda_1)(\lambda_4-\lambda_3)} + \frac{D'''}{\lambda_2-\lambda_1} + E' \tag{148}$$

where

$$C''' = 2\lambda_1^2\lambda_3^2 + 4i\lambda_1\lambda_3^2 - \frac{3}{2}\lambda_3^2 - \frac{3}{2}\lambda_1\lambda_3 - i\lambda_3 + \frac{1}{5}; \tag{149}$$

$$\begin{aligned} D''' = {} & \lambda_1^2\,(22\lambda_3^2\lambda_4^3 + 22i\lambda_3\lambda_4^3 - \frac{29}{2}\lambda_4^3 + 19\lambda_3\lambda_4^2 - \frac{5}{4}i\lambda_4^2) \\ & + \lambda_1\,(22i\lambda_3^2\lambda_4^3 - \frac{47}{2}\lambda_3\lambda_4^3 - \frac{61}{4}i\lambda_4^3 + \frac{67}{4}i\lambda_3\lambda_4^2 - \frac{3}{2}\lambda_4^2 + \frac{23}{4}i\lambda_4) \\ & - \frac{88}{5}\lambda_3^2\lambda_4^3 - \frac{367}{20}i\lambda_3\lambda_4^3 + \frac{427}{40}\lambda_4^3 - \frac{303}{40}\lambda_3\lambda_4^2 + \frac{37}{40}i\lambda_4^2 - \frac{97}{20}\lambda_4); \end{aligned} \tag{150}$$

$$E' = -\frac{57}{10}\lambda_2\lambda_3^2\lambda_4^3. \tag{151}$$

Now we have to sum up four contributions (108), (122), (123) and (148) and get the "canonical" form for $T_4(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$

$$T_4(\lambda_1, \lambda_2, \lambda_3, \lambda_4) \sim P_0^{(4)} + \frac{P_1^{(4)}}{\lambda_2 - \lambda_1} + \frac{P_2^{(4)}}{(\lambda_2 - \lambda_1)(\lambda_4 - \lambda_3)} \tag{152}$$

where

$$P_0^{(4)} = -\frac{34}{5}\lambda_2\lambda_3^2\lambda_4^3; \tag{153}$$

$$\begin{aligned} P_1^{(4)} = {} & \lambda_1^2\,(30\,\lambda_3^2\lambda_4^3 + 30\,i\lambda_3\lambda_4^3 - 16\,\lambda_4^3 + 18\,\lambda_3\lambda_4^2 + 8\,\lambda_4) \\ & + \lambda_1\,(30\,i\lambda_3^2\lambda_4^3 + 30\,\lambda_3\lambda_4^3 - 16\,i\lambda_4^3 + 18\,i\lambda_3\lambda_4^2 - 4\,\lambda_4^2 + 4\,i\lambda_4) \\ & - 20\,\lambda_3^2\lambda_4^3 - 20\,i\lambda_3\lambda_4^3 + \frac{54}{5}\lambda_4^3 - \frac{42}{5}\lambda_3\lambda_4^2 - \frac{43}{10}i\lambda_4; \end{aligned} \tag{154}$$

$$P_2^{(4)} = 2\lambda_1^2\lambda_3^2 + 4i\lambda_1\lambda_3^2 - \frac{3}{2}\lambda_3^2 - \frac{3}{2}\lambda_1\lambda_3 - i\lambda_3 + \frac{1}{5}. \tag{155}$$

Let us note that in terms of the real variables x_j the polynomials $\tilde{P}_j^{(4)}$ (see (21)) look a little bit simpler

$$\tilde{P}_0^{(4)} = -\frac{34}{5}x_2x_3^2x_4^3; \tag{156}$$

$$\begin{aligned} \tilde{P}_1^{(4)} = {} & x_1^2\,(30\,x_3^2x_4^3 - \frac{17}{2}x_4^3 + \frac{81}{2}x_3x_4^2 + \frac{79}{8}x_4) - 4\,x_1x_4^2 \\ & - \frac{25}{2}x_3^2x_4^3 + \frac{147}{40}x_4^3 - \frac{531}{40}x_3x_4^2 - \frac{653}{160}x_4; \end{aligned} \tag{157}$$

$$\tilde{P}_2^{(4)} = 2\,x_1^2x_3^2 + \frac{1}{2}x_1x_3 - \frac{1}{2}x_3^2 + \frac{3}{40}. \tag{158}$$

So, the function

$$\tilde{T}_4^c(x_1, x_2, x_3, x_4) = \tilde{P}_0^{(4)} + \frac{\tilde{P}_1^{(4)}}{x_2 - x_1} + \frac{\tilde{P}_2^{(4)}}{(x_2 - x_1)(x_4 - x_3)} \tag{159}$$

is even i.e.

$$\tilde{T}_4^c(-x_1, -x_2, -x_3, -x_4) = \tilde{T}_4^c(x_1, x_2, x_3, x_4)$$

as it should be according to the formula (23) with $n = 4$.

Now let us start the second step of our general scheme, namely, the performing of the integration of the "canonical" form (152)

$$P(4) = \prod_{j=1}^{4} \int_{C_{-1/2}} \frac{d\lambda_j}{2\pi i} U_4(\lambda_1, \lambda_2, \lambda_3, \lambda_4)\, T_4(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$$

$$= J_0^{(4)} + J_1^{(4)} + J_2^{(4)} \qquad (160)$$

where

$$J_0^{(4)} = \prod_{j=1}^{4} \int_{C_{-1/2}} \frac{d\lambda_j}{2\pi i} U_4(\lambda_1, \lambda_2, \lambda_3, \lambda_4)\, P_0^{(4)} \qquad (161)$$

$$J_1^{(4)} = \prod_{j=1}^{4} \int_{C_{-1/2}} \frac{d\lambda_j}{2\pi i} U_4(\lambda_1, \lambda_2, \lambda_3, \lambda_4)\, \frac{P_1^{(4)}}{\lambda_2 - \lambda_1} \qquad (162)$$

$$J_1^{(4)} = \prod_{j=1}^{4} \int_{C_{-1/2}} \frac{d\lambda_j}{2\pi i} U_4(\lambda_1, \lambda_2, \lambda_3, \lambda_4)\, \frac{P_2^{(4)}}{(\lambda_2 - \lambda_1)(\lambda_4 - \lambda_3)}. \qquad (163)$$

Using the formulae (161–163), (153–155), (57), (58), (66) we get

$$\begin{aligned}
J_0^{(4)} &= \prod_{j=1}^{4} \int_{C_{-1/2}} \frac{d\lambda_j}{2\pi i} U_4(\lambda_1, \lambda_2, \lambda_3, \lambda_4) \left(-\frac{34}{5}\right) \lambda_2 \lambda_3^2 \lambda_4^3 \\
&= D^{(4)} \sum_{l_1=0}^{\infty} (-1)^{l_1} \sum_{l_2=0}^{\infty} (-1)^{l_2} \sum_{l_3=0}^{\infty} (-1)^{l_3} \sum_{l_4=0}^{\infty} (-1)^{l_4} \\
&\qquad \times \left(-\frac{34}{5}\right)(il_2 + \epsilon_2)(il_3 + \epsilon_3)^2 (il_4 + \epsilon_4)^2 \\
&= D^{(4)} \sum_{l_1=0}^{\infty} (-1)^{l_1} \sum_{l_2=0}^{\infty} (-1)^{l_2} \sum_{l_3=0}^{\infty} (-1)^{l_3} \sum_{l_4=0}^{\infty} (-1)^{l_4} \left(-\frac{34}{5}\right) \epsilon_2 \epsilon_3^2 \epsilon_4^3 \\
&= \frac{1}{12} \prod_{0 \le k < j \le 4} \left(\frac{\partial}{\partial \epsilon_k} - \frac{\partial}{\partial \epsilon_j}\right)_{\vec{\epsilon} \to 0} \left(-\frac{34}{5}\right) \epsilon_2 \epsilon_3^2 \epsilon_4^3 = -\frac{17}{40}.
\end{aligned} \qquad (164)$$

Restoring for convenience the dependence of the polynomials $P_1^{(4)}$ and $P_2^{(4)}$ on λ-s as in the formula (18) we can get for the term $J_1^{(4)}$

$$\begin{aligned}
J_1^{(4)} &= \prod_{j=1}^{4} \int_{C_{-1/2}} \frac{d\lambda_j}{2\pi i} U_4(\lambda_1, \lambda_2, \lambda_3, \lambda_4)\, \frac{P_1^{(4)}(\lambda_1 | \lambda_3, \lambda_4)}{\lambda_2 - \lambda_1} \\
&= D^{(4)} \sum_{l_1=1}^{\infty} (-1)^{l_1} \sum_{l_2=0}^{\infty} (-1)^{l_2} \sum_{l_3=1}^{\infty} (-1)^{l_3} \sum_{l_4=0}^{\infty} (-1)^{l_4} \\
&\qquad \times \frac{P_1^{(4)}(-il_1 + \epsilon_1 | -il_3 + \epsilon_3, il_4 + \epsilon_4)}{il_2 + il_1 + \epsilon_2 - \epsilon_1}
\end{aligned}$$

$$= -iD^{(4)}\int_0^1 \frac{ds}{s}\sum_{l_1=1}^{\infty}(-1)^{l_1}\sum_{l_2=0}^{\infty}(-1)^{l_2}\sum_{l_3=1}^{\infty}(-1)^{l_3}\sum_{l_4=0}^{\infty}(-1)^{l_4}$$
$$\times P_1^{(4)}(-il_1+\epsilon_1|-il_3+\epsilon_3, il_4+\epsilon_4)s^{l_1+l_2+i(\epsilon_1-\epsilon_2)}$$
$$= \int_0^1 ds\Big(-\frac{15}{2}\frac{(s-1)}{(1+s)^3} + \frac{3}{4}\frac{(7-26s+7s^2)}{(1+s)^4}\ln s$$
$$+ \frac{21}{8}\frac{(s-1)}{(1+s)^3}\ln^2 s - \frac{7}{10}\frac{(2-s+2s^2)}{(1+s)^4}\ln^3 s$$
$$+ \frac{5}{48}\frac{(s-1)}{(1+s)^3}\ln^4 s + \frac{1}{240}\frac{(1+22s+s^2)}{(1+s)^4}\ln^5 s\Big)$$
$$= \frac{5}{8} - 2\ln 2 + \frac{61}{20}\zeta(3) - \frac{65}{32}\zeta(5). \tag{165}$$

The last term $J_2^{(4)}$ can be calculated as follows

$$J_2^{(4)} = \prod_{j=1}^{4}\int_{C_{-1/2}}\frac{d\lambda_j}{2\pi i}U_4(\lambda_1,\lambda_2,\lambda_3,\lambda_4)\frac{P_2^{(4)}(\lambda_1,\lambda_3)}{(\lambda_2-\lambda_1)(\lambda_4-\lambda_3)}$$
$$= D^{(4)}\sum_{l_1=1}^{\infty}(-1)^{l_1}\sum_{l_2=0}^{\infty}(-1)^{l_2}\sum_{l_3=1}^{\infty}(-1)^{l_3}\sum_{l_4=0}^{\infty}(-1)^{l_4}$$
$$\times\frac{P_2^{(4)}(-il_1+\epsilon_1,-il_3+\epsilon_3)}{(il_2+il_1+\epsilon_2-\epsilon_1)(il_4+il_3+\epsilon_4-\epsilon_3)}$$
$$= -D^{(4)}\int_0^1\frac{ds}{s}\int_0^1\frac{dt}{t}\sum_{l_1=1}^{\infty}(-1)^{l_1}\sum_{l_2=0}^{\infty}(-1)^{l_2}\sum_{l_3=1}^{\infty}(-1)^{l_3}\sum_{l_4=0}^{\infty}(-1)^{l_4}$$
$$\times P_2^{(4)}(-il_1+\epsilon_1,-il_3+\epsilon_3)s^{l_1+l_2+i(\epsilon_1-\epsilon_2)}t^{l_3+l_4+i(\epsilon_3-\epsilon_4)}$$

$$= \int_0^1 ds\int_0^1 dt\Big\{4\frac{\ln s(\ln s-\ln t)}{(1+s)^2(1+t)^2} + \Big(-\frac{4}{3}(-3-s+t+3st)\ln^3 s$$
$$+ 4(-1+3s-3t+st)\ln^2 s\ln t - \frac{3}{4}(3-13s+3t+3st)\ln^2 s\ln^2 t$$
$$+ \frac{5}{12}(3-5s-5t+3st)\ln^4 s\Big)\frac{1}{(1+s)^3(1+t)^3}$$
$$-\frac{1}{6}\frac{(-1+5s-2s^2+2t-5st+s^2t)}{(1+s)^4(1+t)^3}\ln s(\ln^2 s-\ln^2 t)(\ln^2 s-5\ln^2 t)$$
$$+\Big(\frac{1}{3}(3+30s-5s^2-34t-4st-34s^2t$$
$$-5t^2+30st^2+3s^2t^2)\ln^3 s\ln t$$
$$+\frac{1}{30}(4-17s+9s^2-17t+66st-17s^2t$$
$$+9t^2-17st^2+4s^2t^2)\ln^5 s\ln t$$

$$- \frac{1}{15}(2 - 31s + 7s^2 + 14t + 33st - 21s^2t$$
$$+ 2t^2 + 4st^2 + 2s^2t^2) \ln^3 s \ln^3 t \Big) \frac{1}{(1+s)^4(1+t)^4} \Big\}.$$

Taking the integral we come to an answer for $J_2^{(4)}$

$$J_2^{(4)} = -\frac{1}{6}\zeta(3) - \frac{11}{6}\zeta(3)\ln 2 - \frac{51}{80}\zeta(3)^2 - \frac{25}{96}\zeta(5) + \frac{85}{24}\zeta(5)\ln 2. \quad (166)$$

Summing up all the three results (164), (165) and (166) we finally get our main formula (9)

$$\begin{aligned} P(4) &= J_0^{(4)} + J_1^{(4)} + J_2^{(4)} \\ &= \frac{1}{5} - 2\ln 2 + \frac{173}{60}\zeta(3) - \frac{11}{6}\zeta(3)\ln 2 - \frac{51}{80}\zeta^2(3) - \frac{55}{24}\zeta(5) + \frac{85}{24}\zeta(5)\ln 2. \end{aligned} \quad (167)$$

References

1. Kreimer, D. (2000) *Knots and Feynman Diagrams*, Cambridge University Press.
2. Boos, H. E. and Korepin, V. E. (2001) *Quantum Spin Chains and Riemann Zeta Function with Odd Arguments*, hep-th/0104008.
3. Heisenberg, W. (1928) *Zeitschrift für Physik*, **49**, 619.
4. Bethe, H. (1931) *Zeitschrift für Physik*, **76**, 205.
5. Hulthén, L. (1939) *Ark. Mat. Astron. Fysik*, **A 26**, 1.
6. Korepin, V. E., Izergin, A. G. and Bogoliubov, N. M. (1993) *Quantum inverse scattering method and correlation functions*, Cambridge Univ. Press, Cambridge, 1993.
7. Jimbo, M., Miki, K., Miwa, T. and Nakayashiki, A. (1992) Correlation Functions of the XXZ model for $\Delta < -1$, *Phys. Lett.* **A 166**; hep-th/9205055.
8. Korepin, V. E., Izergin, A. G., Essler, F. and Uglov, D. B. (1994) *Phys. Lett.* **A 190**, 182–184.
9. Takahashi, M. (1977) Thermodynamical Bethe Ansatz and Condensed Matter, *J. Phys.* **C 10**, 1289; cond-mat/9708087.
10. Dittrich, J. and Inozemtsev, V. I. (1997) On the second-neighbor correlator in 1D XXX quantum anti-ferromagnetic spin chain, cond-mat/9706263.
11. Berruto, F., Grignani, G., Semenoff, G. W. and Sodano, P. (1999) On the Correspondence Between the Strongly Coupled 2-Flavor Lattice Schwinger Model and the Heisenberg Antiferromagnetic Chain, hep-th/9901142; preprint DFUPG-190-98, UBC/GS-6-98.
12. de Gier, J. and Korepin, V. E. (2001) Six-Vertex Model with Domain wall boundary conditions. Variable inhomogeneities. math-ph/0101036.
13. Razumov, A. V. and Stroganov, Yu. G. (2000) Spin chains and combinatorics, cond-mat/0012141; (2001) Spin chains and combinatorics: twisted boundary conditions, cond-mat/0102247.
14. Titchmarch, E. C. (1987) *The Theory of the Riemann Zeta-Function*, Clarendon Press.
15. Riemann, B. (1860) Über die Anzahl der Primzahlen unter einer gegebenen Grösse, Monat. der Koenigl. Preuss. Akad. der Wissen. zu Berlin aus dem Jahre 1859, 671–680; also *Gesammelte mat. Werke und wissensch*, Nachlass, 2 Aufl. 1892, 145–155.
16. Tanguy Rivoal (2000) Irrationalite d'au moins un des neuf nombres $\zeta(5), \zeta(7), \ldots, \zeta(21)$, *C. R. Acad. Sci. Paris Sér. I Math.* **331**, no. 4, 267–270 and math.NT/0104221.
17. Don Zagier (1994) Values of Zeta Functions and their Applications, *First European Congress of Mathematics, Vol. II (Paris, 1992)*, Prog. Math., Birkhauser, Basel–Boston, p. 497.

18. Borwein, J. M., Bradley, D. M., Broadhurst, D. J. and Lisonek, P. (1999) Special values of multiple polylogarithms, `math.CA/9910045`.
19. Knauf, A. (1999) *Reviews in Mathematical Physics*, **11**, 1027–1060.
20. Kitanine, N., Maillet, J. M. and Terras, V. (1999) Correlation functions of the XXZ Heisenberg spin-1/2 chain in a magnetic field, `math-ph/9907019`.
21. Forrester, P. J. and Odlyzko, A. M. (1996) Phys. Rev. E **54**, p. 4493.
22. Tracy, C. A. and Widom, H. (1994) Fredholm determinants, differential equations and matrix models, *Commun. Math. Phys.* **163**, 33–72.
23. Its, A. R., Izergin, A. G., Korepin, V. E. and Slavnov, N. A. (1990) *Int. Jour. Mod. Phys. B* **4** p. 1003.
24. Odlyzko, A. M. (1989) The 10^{20} zero of the Riemann zeta function and 70 million of its neighbors, AT&T Bell Lab preprint.
25. Montgomery, H. L. (1973) *Proc. Symp. Pure Math.* **24**, 181–193.

QUANTIZATION OF THERMODYNAMICS AND THE BARDEEN–COOPER–SCHRIFFER–BOGOLYUBOV EQUATION

V. P. MASLOV (islc@dol.ru)
Physics Department
Moscow State University
Moscow
Russia

Abstract. New quantization methods for quantum problems of many-bodies are presented. These methods are used for quantization of thermodynamics of Quantum Statistics. In particular, this leads to an explanation of the superconductivity and superfluidity for nonzero temperatures.

1. Ultrasecondary quantization

The ultrasecondary quantization for problems of quantum mechanics and statistical physics was introduced in the papers [1–3]. Let us recall the notation and the main facts. Suppose that $\mathcal{F}$ is the boson Fock space [8], $\widehat{b}^+(x,s)$ is the creation operator for particles with number s and $\widehat{b}^-(x,s)$ is the annihilation operator for particles with number s in the space $\mathcal{F}$ [8], and $\widehat{B}^+(x,x')$ is the creation operator for a pair of particles and $\widehat{B}^-(x,x')$ is the annihilation operator for a pair of particles. These operators satisfy the commutation relations

$$\begin{aligned}
&[\widehat{b}^-(x,s),\widehat{b}^+(x',s')] = \delta_{ss'}\delta(x-x'), \qquad [\widehat{b}^\pm(x,s),\widehat{b}^\pm(x',s')] = 0,\\
&[\widehat{B}^-(x_1,x_2)\widehat{B}^+(x_1',x_2')] = \delta(x_1-x_1')\delta(x_2-x_2'),\\
&[\widehat{B}^\pm(x_1,x_2),\widehat{B}^\pm(x_1',x_2')] = 0,\\
&[\widehat{b}^\pm(x,s),\widehat{B}^\pm(x_1',x_2')] = [\widehat{b}^\pm(x,s),\widehat{b}^\mp(x_1',x_2')] = 0.
\end{aligned} \tag{1}$$

Further, Φ_0 is the vacuum vector in the space $\mathcal{F}$. This vector possesses the following properties:

$$\widehat{b}^-(x,s)\Phi_0 = 0, \qquad \widehat{B}^-(x_1,x_2)\Phi_0 = 0. \tag{2}$$

The variable x lies on the three-dimensional torus $L\times L\times L$, which we denote by $\mathbf{T}^3$. The variable s is discrete, $s = 0,1,\dots$; s is called the *number* or the *statistical spin.*

V. A. Malyshev and A. M. Vershik (eds.),
Asymptotic Combinatorics with Application to Mathematical Physics, 209–220.

Any vector Φ in the space $\mathcal{F}$ can be represented uniquely as

$$\Phi = \sum_{k=0}^{\infty}\sum_{M=0}^{\infty}\frac{1}{k!M!}\sum_{s_1=0}^{\infty}\cdots\sum_{s_k=0}^{\infty}\int\cdots\int dx_1\ldots dx_k dy_1\ldots dy_{2M}$$
$$\cdot\Phi_{k,M}(x_1,s_1;\ldots;x_k,s_k;y_1,y_2;\ldots;y_{2M-1},y_{2M})$$
$$\cdot\widehat{b}^+(x_1,s_1)\cdot\ldots\cdot\widehat{b}^+(x_k,s_k)\widehat{B}^+(y_1,y_2)\cdot\ldots\cdot\widehat{B}^+(y_{2M-1},y_{2M})\Phi_0, \quad (3)$$

where the function $\Phi_{k,M}(x_1,s_1;\ldots;x_k,s_k;y_1,y_2;\ldots;y_{2M-1},y_{2M})$ is symmetric with respect to permutations of the pairs of variables (x_j,s_j) and (x_i,s_i) and symmetric with respect to permutations of the pairs of variables (y_{2j-1},y_{2j}) and (y_{2i-1},y_{2i}).

In the boson case we introduce the subspace $\mathcal{F}_{k,M}^{\mathrm{Symm}}$ consisting of vectors Φ such that $\Phi_{k',M'}=0$ for $(k',M')\neq(k,M)$ and $\Phi_{k,M}$ is a symmetric function of the variables $x_1,x_2,\ldots,x_k,y_1,y_2,\ldots,y_{2M}$.

Similarly, in the fermion case we introduce the subspace $\mathcal{F}_{k,M}^{\mathrm{Asymm}}$ consisting of vectors Φ such that $\Phi_{k',M'}=0$ for $(k',M')\neq(k,M)$ and $\Phi_{k,M}$ is an antisymmetric function of the variables $x_1,x_2,\ldots,x_k$, $y_1,y_2,\ldots,y_{2M}$.

The orthogonal projection operator in the space $\mathcal{F}$ on the subspace $\mathcal{F}_{k,M}^{\mathrm{Symm}}$ has the form [1–3]:

$$\widehat{\Pi}_{k,M}^{\mathrm{Symm}} = \frac{1}{k!M!}\sum_{s_1=0}^{\infty}\cdots\sum_{s_k=0}^{\infty}\int\cdots\int dx_1\ldots dx_k dy_1\ldots dy_{2M}$$
$$\cdot\,\widehat{b}^+(x_1,s_1)\cdot\ldots\cdot\widehat{b}^+(x_k,s_k)\widehat{B}^+(y_1,y_2)\cdot\ldots\cdot\widehat{B}^+(y_{2M-1},y_{2M})$$
$$\cdot\ \operatorname*{Symm}_{x_1\ldots x_k y_1\ldots y_{2M}}\Big(\widehat{b}^-(x_1,s_1)\cdot\ldots\cdot\widehat{b}^-(x_k,s_k)\widehat{B}^-(y_1,y_2)\cdot\ldots\cdot\widehat{B}^-(y_{2M-1},y_{2M})\Big)$$
$$\cdot\exp\Big(-\sum_{s=0}^{\infty}\int dx\,\widehat{b}^+(x,s)\widehat{b}^-(x,s)-\iint dy\,dy'\,\widehat{B}^+(y,y')\widehat{B}^-(y,y')\Big), \quad (4)$$

where $\mathrm{Symm}_{x_1\ldots x_k y_1\ldots y_{2M}}$ is the symmetrization operator with respect to the variables $x_1,\ldots,x_k$, $y_1,\ldots,y_{2M}$ and the operators $\widehat{b}^+(x,s)$, $\widehat{b}^-(x,s)$ and $\widehat{B}^+(y,y')$, $\widehat{B}^-(y,y')$ are Wick ordered [8].

The orthogonal projection operator in the space $\mathcal{F}$ on the subspace $\mathcal{F}_{k,M}^{\mathrm{Asymm}}$ has the form [1–3]:

$$\widehat{\Pi}_{k,M}^{\mathrm{Asymm}} = \frac{1}{k!M!}\sum_{s_1=0}^{\infty}\cdots\sum_{s_k=0}^{\infty}\int\cdots\int dx_1\ldots dx_k dy_1\ldots dy_{2M}$$
$$\cdot\,\widehat{b}^+(x_1,s_1)\cdot\ldots\cdot\widehat{b}^+(x_k,s_k)\widehat{B}^+(y_1,y_2)\cdot\ldots\cdot\widehat{B}^+(y_{2M-1},y_{2M})$$
$$\cdot\ \operatorname*{Asymm}_{x_1\ldots x_k y_1\ldots y_{2M}}\Big(\widehat{b}^-(x_1,s_1)\cdot\ldots\cdot\widehat{b}^-(x_k,s_k)\widehat{B}^-(y_1,y_2)\cdot\ldots\cdot\widehat{B}^-(y_{2M-1},y_{2M})\Big)$$
$$\cdot\exp\Big(-\sum_{s=0}^{\infty}\int dx\,\widehat{b}^+(x,s)\widehat{b}^-(x,s)-\iint dy\,dy'\,\widehat{B}^+(y,y')\widehat{B}^-(y,y')\Big), \quad (5)$$

where $\mathrm{Asymm}_{x_1\dots x_k y_1\dots y_{2M}}$ is the antisymmetrization operator with respect to the variables $x_1,\dots,x_k,y_1,\dots,y_{2M}$. ¿From now on, unless otherwise specified, the operators $\widehat{b}^+(x,s)$, $\widehat{b}^-(x,s)$ and $\widehat{B}^+(y,y')$, $\widehat{B}^-(y,y')$ are Wick ordered.

In what follows, we study a system of N identical particles on the torus $\mathbf{T}$. We assume that the Hamiltonian of N bosons or fermions has the form

$$\widehat{H}_N = -\frac{\hbar^2}{2m}\sum_{j=1}^{N}\Delta_j + \sum_{j=1}^{N}\sum_{l=j+1}^{N} V(x_j - x_l). \tag{6}$$

By [1–3], in the boson case to this operator there corresponds an ultrasecondary quantized Hamiltonian of the form

$$\begin{aligned}\overline{\widehat{H}}_B = \sum_{k=0}^{\infty}\sum_{M=0}^{\infty}\frac{1}{k!M!}\sum_{s_1=0}^{\infty}\cdots\sum_{s_k=0}^{\infty}\int\cdots\int dx_1\dots dx_k dy_1\dots dy_{2M}\\ \cdot\widehat{b}^+(x_1,s_1)\cdot\ldots\cdot\widehat{b}^+(x_k,s_k)\widehat{B}^+(y_1,y_2)\cdot\ldots\cdot\widehat{B}^+(y_{2M-1},y_{2M})\widehat{H}_{k+2M}\\ \cdot\ \mathop{\mathrm{Symm}}_{x_1\dots x_k y_1\dots y_{2M}}\Big(\widehat{b}^-(x_1,s_1)\cdot\ldots\cdot\widehat{b}^-(x_k,s_k)\widehat{B}^-(y_1,y_2)\cdot\ldots\cdot\widehat{B}^-(y_{2M-1},y_{2M})\Big)\\ \cdot\exp\Big(-\sum_{s=0}^{\infty}\int dx\,\widehat{b}^+(x,s)\widehat{b}^-(x,s) - \iint dy\,dy'\widehat{B}^+(y,y')\widehat{B}^-(y,y')\Big).\end{aligned} \tag{7}$$

In the fermion case, the corresponding operator $\overline{\widehat{H}}_F$ is expressed by a similar formula where Symm is replaced by Asymm.

By analogy with (6) and (7), the ultrasecondary quantized operator $\overline{\widehat{A}}$ is associated with any N-particle operator [1–3]

$$\overline{\widehat{H}}_N\Big(\overset{2}{x}_1,\dots,\overset{2}{x}_N;-i\overset{1}{\frac{\partial}{\partial x_1}},\dots,-i\overset{1}{\frac{\partial}{\partial x_N}}\Big).$$

For example, in the boson case, with the unity operator we associate an ultrasecondary quantized unity operator of the form

$$\begin{aligned}\overline{\widehat{E}}_B = \sum_{k=0}^{\infty}\sum_{M=0}^{\infty}\frac{1}{k!M!}\sum_{s_1=0}^{\infty}\cdots\sum_{s_k=0}^{\infty}\int\cdots\int dx_1\dots dx_k dy_1\dots dy_{2M}\\ \cdot\widehat{b}^+(x_1,s_1)\cdot\ldots\cdot\widehat{b}^+(x_k,s_k)\widehat{B}^+(y_1,y_2)\cdot\ldots\cdot\widehat{B}^+(y_{2M-1},y_{2M})\\ \cdot\ \mathop{\mathrm{Symm}}_{x_1\dots x_k y_1\dots y_{2M}}\Big(\widehat{b}^-(x_1,s_1)\cdot\ldots\cdot\widehat{b}^-(x_k,s_k)\widehat{B}^-(y_1,y_2)\cdot\ldots\cdot\widehat{B}^-(y_{2M-1},y_{2M})\Big)\\ \cdot\exp\Big(-\sum_{s=0}^{\infty}\int dx\widehat{b}^+(x,s)\widehat{b}^-(x,s) - \iint dydy'\widehat{B}^+(y,y')\widehat{B}^-(y,y')\Big),\end{aligned} \tag{8}$$

which is the sum of the projection operators (4).

Similarly, in the fermion case the ultrasecondary quantized unity operator is

$$\overline{\widehat{E}}_F = \sum_{k=0}^{\infty}\sum_{M=0}^{\infty}\widehat{\Pi}_{k,M}^{\mathrm{Asymm}},$$

and Symm in formula (8) is replaced by Asymm.

Let us consider the eigenvalue problem

$$\overline{\overline{\widehat{H}}}_{B,F}\Phi = \lambda \overline{\overline{\widehat{E}}}_B\Phi, \qquad \overline{\overline{\widehat{E}}}\Phi \neq 0 \tag{9}$$

in the boson and fermion cases The following assertion holds.

Theorem 1. *On the subspaces $\mathcal{F}^{\mathrm{Symm}}_{k,M}$ and $\mathcal{F}^{\mathrm{Asymm}}_{k,M}$ of the space $\mathcal{F}$, the operators $\overline{\overline{\widehat{H}}}_B$ and $\overline{\overline{\widehat{H}}}_F$ coincide, respectively, with the operator $\widehat{H}_{k+2M}$.*

The proof is given in [1–3].

Corollary 2. *The eigenvalues λ of problem* (9) *in the boson and fermion cases coincide with the corresponding eigenvalues of the operators $\widehat{H}_N$* (6).

If the commutators of the operators $\widehat{b}^-(x,s)$ and $\widehat{b}^+(x,s)$, as well as of $\widehat{B}^-(x,y)$ and $\widehat{B}^+(x,y)$, are small like $1/N$, then, by [5–7], the asymptotics of the solutions of problem (9) is determined by the extremum points of the symbol corresponding to problem (9). In the boson case the symbol has the form

$$\begin{aligned}
&\mathcal{H}_B[b^*(\cdot),b(\cdot),B^*(\cdot),B(\cdot)] \\
&= \Big\{ \sum_{k,M=0}^{\infty} \frac{1}{k!M!} \sum_{s_1=0}^{\infty}\cdots\sum_{s_k=0}^{\infty} \int\cdots\int dx_1\ldots dx_k dy_1\ldots dy_{2M} \\
&\qquad \cdot b^*(x_1,s_1)\cdot\ldots\cdot b^*(x_k,s_k)B^*(y_1,y_2)\cdot\ldots\cdot B^*(y_{2M-1},y_{2M})H_{k+2M} \\
&\qquad \cdot \operatorname*{Symm}_{x_1\ldots x_k y_1\ldots y_{2M}}\Big(b(x_1,s_1)\cdot\ldots\cdot b(x_k,s_k)B(y_1,y_2)\cdot\ldots\cdot B(y_{2M-1},y_{2M})\Big)\Big\} \\
&\cdot\Big\{ \sum_{k',M'=0}^{\infty} \frac{1}{k'!M'!} \sum_{s'_1=0}^{\infty}\cdots\sum_{s'_{k'}=0}^{\infty} \int\cdots\int dx'_1\ldots dx'_{k'} dy'_1\ldots dy'_{2M'} \\
&\qquad \cdot b^*(x'_1,s'_1)\cdot\ldots\cdot b^*(x'_{k'},s'_{k'})B^*(y'_1,y'_2)\cdot\ldots\cdot B^*(y'_{2M'-1},y'_{2M'}) \\
&\qquad \cdot \operatorname*{Symm}_{x'_1\ldots x'_{k'},y'_1\ldots y'_{2M'}}\Big(b(x'_1,s'_1)\cdot\ldots\cdot b(x'_{k'},s'_{k'})B(y'_1,y'_2)\cdot\ldots\cdot B(y'_{2M'-1},y'_{2M'})\Big)\Big\}.
\end{aligned} \tag{10}$$

In the fermion case, the symbol is expressed in a similar way, only Symm in formula (10) is replaced by Asymm.

In the boson case, we have the following identity for the symbol (10):

$$\mathcal{H}_B[b^*(\cdot),b(\cdot),B^*(\cdot),B(\cdot)] = \frac{\mathrm{Sp}(\widehat{\rho}_B\widehat{H})}{\mathrm{Sp}(\widehat{\rho}_B)}, \tag{11}$$

where $\widehat{H}$ and $\widehat{\rho}_B$ are secondary quantized operators:

$$\begin{aligned}
\widehat{H} &= \int dx\,\widehat{\psi}^+(x)\Big(-\frac{\hbar^2}{2m}\Delta\Big)\widehat{\psi}^-(x) \\
&\quad +\frac{1}{2}\iint dx\,dy\,V(x,y)\widehat{\psi}^+(y)\widehat{\psi}^+(x)\widehat{\psi}^-(y)\widehat{\psi}^-(x).
\end{aligned} \tag{12}$$

Here $\widehat{\rho}_B$ depends on the functions $b(x,s)$ and $B(y,y')$ as follows:

$$\begin{aligned}\widehat{\rho}_B &= \sum_{k=0}^{\infty}\sum_{M=0}^{\infty}\frac{1}{k!M!(k+2M)!}\Big(\sum_{s=0}^{\infty}\iint dxdx' b(x,s)b^*(x',s)\widehat{\psi}^+(x)\widehat{\psi}^-(x')\Big)^k \\ &\quad\cdot\Big(\iint dy_1dy_2 B(y_1,y_2)\widehat{\psi}^+(y_1)\widehat{\psi}^+(y_2)\Big)^M \\ &\quad\cdot\Big(\iint dy_1'dy_2' B(y_1',y_2')\widehat{\psi}^-(y_1')\widehat{\psi}^-(y_2')\Big)^M \\ &\quad\cdot\exp\Big(-\int dz\widehat{\psi}^+(z)\widehat{\psi}^-(z)\Big),\end{aligned}\tag{13}$$

where $\widehat{\psi}^+(x)$ and $\widehat{\psi}^-(x)$ are the Bose creation and annihilation Wick-ordered operators [8].

In the fermion case, we have the similar identity

$$\mathcal{H}_F[b^*(\cdot),b(\cdot),B^*(\cdot),B(\cdot)] = \frac{\mathrm{Sp}(\widehat{\rho}_F\widehat{H})}{\mathrm{Sp}(\widehat{\rho}_F)},\tag{12$'$}$$

where $\widehat{H}$ and $\widehat{\rho}_F$ are the following secondary quantized operators:

$$\begin{aligned}\widehat{H} &= \int dx\,\widehat{\psi}^+(x)\Big(-\frac{\hbar^2}{2m}\Delta\Big)\widehat{\psi}^-(x) \\ &\quad+\frac{1}{2}\iint dx\,dy\,V(x,y)\widehat{\psi}^+(x)\widehat{\psi}^+(y)\widehat{\psi}^-(y)\widehat{\psi}^-(x)\end{aligned}\tag{13$'$}$$

and

$$\begin{aligned}\widehat{\rho}_F &= \sum_{k=0}^{\infty}\sum_{M=0}^{\infty}\frac{1}{k!M!(k+2M)!}\Big(\iint dy_1dy_2 B(y_1,y_2)\widehat{\psi}^+(y_1)\widehat{\psi}^+(y_2)\Big)^M \\ &\quad\cdot\sum_{s_1=0}^{\infty}\cdots\sum_{s_k=0}^{\infty}\int\ldots\int dx_1dx_1'\ldots dx_kdx_k' \\ &\quad\cdot b(x_1,s_1)b^*(x_1',s_1)\cdot\ldots\cdot b(x_k,s_k)b^*(x_k',s_k) \\ &\quad\cdot\widehat{\psi}^+(x_1)\cdot\ldots\cdot\widehat{\psi}^+(x_k)\widehat{P}_0\widehat{\psi}^-(x_k')\cdot\ldots\cdot\widehat{\psi}^-(x_1') \\ &\quad\cdot\Big(\iint dy_1'dy_2' B(y_1',y_2')\widehat{\psi}^-(y_2')\widehat{\psi}^-(y_1')\Big)^M.\end{aligned}\tag{14}$$

Here $\widehat{\psi}^+(x)$ and $\widehat{\psi}^-(x)$ are the Fermi creation and annihilation operators and $\widehat{P}_0$ is the projection operator on the vacuum vector of the fermion Fock space.

In general, for an arbitrary secondary quantized operator $\widehat{A}$, the symbol of the corresponding ultrasecondary quantized operator $\overline{\widehat{A}}$ is expressed [1–3] by the formula

$$A_{B,F}[b^*(\cdot),b(\cdot),B^*(\cdot),B(\cdot)] = \frac{\mathrm{Sp}(\widehat{\rho}_{B,F}\widehat{A})}{\mathrm{Sp}(\widehat{\rho}_{B,F})}.$$

In the space $\mathcal{F}$ we introduce [1–3] the ultrasecondary quantized operators of the number of particles:

$$\hat{\hat{N}}_B = \sum_{k=0}^{\infty}\sum_{M=0}^{\infty}(k+2M)\hat{\Pi}_{k,M}^{\text{Symm}}, \qquad \hat{\hat{N}}_F = \sum_{k=0}^{\infty}\sum_{M=0}^{\infty}(k+2M)\hat{\Pi}_{k,M}^{\text{Asymm}}. \tag{15}$$

Respectively, in the boson case, the symbol of the operator $\hat{\hat{N}}_B$ has the form

$$\begin{aligned} N_B = \Bigg\{ &\sum_{k=0}^{\infty}\sum_{M=0}^{\infty}\frac{k+2M}{k!M!}\sum_{s_1=0}^{\infty}\cdots\sum_{s_k=0}^{\infty}\int\cdots\int dx_1\dots dx_{k+2M} \\ &\cdot b^*(x_1,s_1)\cdot\ldots\cdot b^*(x_k,s_k)B^*(x_{k+1},x_{k+2})\cdot\ldots\cdot B^*(x_{k+2M-1},x_{k+2M}) \\ &\cdot \operatorname*{Symm}_{x_1\dots x_{k+2M}}\Big(b(x_1,s_1)\cdot\ldots\cdot b(x_k,s_k)B(x_{k+1},x_{k+2})\cdot\ldots\cdot B(x_{k+2M-1},x_{k+2M})\Big)\Bigg\} \\ &\cdot\Bigg\{\sum_{k'=0}^{\infty}\sum_{M'=0}^{\infty}\frac{1}{k'!M'!}\sum_{s'_1=0}^{\infty}\cdots\sum_{s'_{k'}=0}^{\infty}\int\cdots\int dz_1\dots dz_{k'+2M'} \\ &\cdot b^*(z_1,s'_1)\cdot\ldots\cdot b^*(z_{k'},s'_{k'})B^*(z_{k'+1},z_{k'+2})\cdot\ldots\cdot B^*(x_{k'+2M'-1},x_{k'+2M'}) \\ &\cdot \operatorname*{Symm}_{z_1\dots z_{k'+2M'}}\Big(b(z_1,s'_1)\cdot\ldots\cdot B(z_{k'+2M'-1},z_{k'+2M'})\Big)\Bigg\}^{-1}. \end{aligned} \tag{16}$$

In the corresponding formula in the fermion case, Symm is replaced by Asymm.

2. Ultrasecondary quantization of intrinsic energy, entropy, and free energy

With each vector $\Phi \in \mathcal{F}$ we associate entropy equal to

$$S = -\operatorname{Sp}(\hat{R}(\Phi(\Phi)\ln\hat{R}(\Phi)),$$

where $\hat{R}(\Phi)$ is the secondary quantized operator:

$$\begin{aligned} \hat{R}(\Phi) = \Bigg\{ &\sum_{k=0}^{\infty}\sum_{M=0}^{\infty}\frac{1}{k!M!(k+2M)!} \\ &\cdot\sum_{s_1=0}^{\infty}\cdots\sum_{s_k=0}^{\infty}\int\cdots\int dx_1dx'_1\dots dx_kdx'_kdy_1dy'_1\dots dy_{2M}dy'_{2M} \\ &\cdot\hat{\psi}^+(y_1)\dots\hat{\psi}^+(y_{2M})\hat{\psi}^+(x_1)\dots\hat{\psi}^+(x_k)\cdot\hat{\mathcal{P}}_0\hat{\psi}^-(x'_k)\dots\hat{\psi}^-(x'_1)\hat{\psi}^-(y'_{2M})\dots\hat{\psi}^-(y'_1) \\ &\cdot\Big(\Phi,\overset{2}{\hat{b}^+}(x_1,s_1)\cdot\ldots\cdot\overset{2}{\hat{b}^+}(x_k,s_k)\,\overset{2}{\hat{B}^+}(y_1,y_2)\cdot\ldots\cdot\overset{2}{\hat{B}^+}(y_{2M-1},y_{2M}) \\ &\cdot\exp\Big(-\sum_{s=0}^{\infty}\int dx\,\overset{2}{\hat{b}^+}(x,s)\,\overset{1}{\hat{b}^-}(x,s)-\iint dy\,dy'\,\overset{2}{\hat{B}^+}(y,y')\,\overset{1}{\hat{B}^-}(y,y')\Big) \\ &\cdot\overset{1}{\hat{B}^-}(y'_{2M-1},y'_{2M})\cdot\ldots\cdot\overset{1}{\hat{B}^-}(y'_1,y'_2)\,\overset{1}{\hat{b}^-}(x'_k,s_k)\cdot\ldots\cdot\overset{1}{\hat{b}^-}(x'_1,s_1)\Phi\Big)\Bigg\}\Big/(\Phi,\hat{\hat{E}}_{B,F}\Phi). \end{aligned}$$

Here, depending on the case under study, $\widehat{\psi}^{\pm}(x)$ are either boson or fermion Fock operators and $\widehat{\mathcal{P}}_0$ is the orthogonal projection operator on the vacuum vector of the Fock space.

In addition, on the space $\mathcal{F}$ we introduce the functionals

$$\mathcal{E} = \frac{(\Phi, \overline{\overline{\widehat{H}}}_{B,F}\Phi)}{(\Phi, \overline{\overline{\widehat{E}}}_{B,F}\Phi)}. \tag{17}$$

These functionals are called *intrinsic energy*.

Definition 3. The *eigenvalues* of intrinsic energy are defined to be the extremum values of the functional $\mathcal{E}$ under the additional conditions

$$S = \text{const}, \qquad \frac{(\Phi, \overline{\overline{\widehat{N}}}_{B,F}\Phi)}{(\Phi, \overline{\overline{\widehat{E}}}_{B,F}\Phi)} = N. \tag{18}$$

The extremum equation for the functionals (17) under the conditions (18) takes the form

$$\big(\overline{\overline{\widehat{H}}} - \mu\overline{\overline{\widehat{N}}} + \theta\overline{\overline{\widehat{L}}}(\Phi)\big)\Phi = \lambda\overline{\overline{\widehat{E}}}\Phi, \tag{19}$$

where θ and μ are the Lagrange multipliers (the temperature and the chemical potential, respectively) and $\overline{\overline{\widehat{L}}}_{B,F}(\Phi)$ is the ultrasecondary quantized operator corresponding to $\ln \widehat{R}(\Phi)$.

The values of λ constitute the spectrum of free energy.

Definition 4. The *eigenvalues* of free energy of an N-particle system at temperature θ are defined to be the values of λ for which Eq. (19) has solutions satisfying the conditions $\overline{\overline{\widehat{N}}}_{B,F}\Phi = N\overline{\overline{\widehat{E}}}\Phi$, $\overline{\overline{\widehat{E}}}_{B,F}\Phi \neq 0$.

To find the asymptotic expansion of solutions of Eq. (19) in the case in which the commutators are small like $1/N$ as $N \to \infty$, in the functionals $\mathcal{E}$, N, and S, we set the operators $\widehat{b}^{+}(x,s)$ and $\widehat{b}^{-}(x,s)$ to be the c-numbers $b^*(x,s)$ and $b(x,s)$, respectively, and the operators $\widehat{B}^{+}(x,y)$ and $\widehat{B}^{-}(x,y)$ to be $B^*(x,y)$ and $B(x,y)$. From $\mathcal{E}$ and N, we find the symbols $\mathcal{H}_B$ (10) and N_B (16) or $\mathcal{H}_F$ and N_F in the corresponding boson and fermion cases. From S we obtain the entropy symbols

$$S_{B,F} = -\operatorname{Sp}\left(\frac{\widehat{\rho}_{B,F}}{\operatorname{Sp}\widehat{\rho}_{B,F}}\ln\left(\frac{\widehat{\rho}_{B,F}}{\operatorname{Sp}\widehat{\rho}_{B,F}}\right)\right). \tag{20}$$

Under the additional conditions $S_F = S$ and $N_F = N$, the symbol $\mathcal{H}_F$ attains its minimum at $B(x,y)$ and $b(x,s)$ such that

$$\frac{\delta\mathcal{H}_F}{\delta b^*(x,s)} - \mu\frac{\delta N_F}{\delta b^*(x,s)} - \theta\frac{\delta S_F}{\delta b^*(x,s)} = 0,$$

$$\frac{\delta\mathcal{H}_F}{\delta B^*(x,y)} - \mu\frac{\delta N_F}{\delta B^*(x,y)} - \theta\frac{\delta S_F}{\delta B^*(x,y)} = 0.$$

By a change of variables, these equations can be reduced to the Bardeen–Cooper–Schriffer–Bogolyubov equations [1–3, 9] of conductivity theory.

From the equation for the spectrum of free energy (19) we pass to the following Cauchy problem:

$$i\overline{\widehat{H}}\frac{\partial}{\partial t}\Phi(t) = \Big(\overline{\widehat{E}} - \mu\overline{\widehat{N}} + \theta\overline{\widehat{L}}(\Phi(t))\Big)\Phi(t). \tag{21}$$

Suppose that at the initial moment of time the vector $\Phi(t) \in \mathcal{F}$ satisfies the relation

$$\begin{aligned}\overline{\widehat{E}}\Phi(0) = \sum_{k=0}^{\infty}\sum_{M=0}^{\infty}\frac{1}{\sqrt{k!M!}}\sum_{s_1=0}^{\infty}\cdots\sum_{s_k=0}^{\infty}\int\cdots\int dx_1\dots dx_k dy_1\dots dy_{2M}\\ \cdot\Phi_{k,M}(x_1,s_1;\dots;x_k,s_k;y_1,y_2;\dots;y_{2M-1},y_{2M})\\ \cdot\widehat{b}^+(x_1,s_1)\cdot\ldots\cdot\widehat{b}^+(x_k,s_k)\widehat{B}^+(y_1,y_2)\cdot\ldots\cdot\widehat{B}^+(y_{2M-1},y_{2M})\Phi_0,\end{aligned} \tag{22}$$

where the functions $\Phi_{k,M}$ can be represented as

$$\begin{aligned}&\Phi_{k,M}(x_1,s_1;\dots;x_k,s_k;y_1,y_2;\dots;y_{2M-1},y_{2M})\\ &\quad=\sum_{\alpha=0}^{\infty}\sqrt{a_{k,M,\alpha}}\Psi_{k,M,\alpha}(x_1,\dots,x_k,y_1,\dots,y_{2M})C_{k,M,\alpha}(s_1,\dots,s_N)\end{aligned} \tag{23}$$

and $\Psi_{k,M,\alpha}$, $C_{k,M,\alpha}$ are symmetric or antisymmetric (in the boson or fermion cases, respectively) functions such that

$$\begin{aligned}\int\cdots\int dx_1\dots dx_k dy_1\dots dy_{2M}\Psi^*_{k,M,\alpha}(x_1,\dots,x_k,y_1,\dots,y_{2M})\\ \cdot\Psi_{k,M,\alpha'}(x_1,\dots,x_k,y_1,\dots,y_{2M}) = \delta_{\alpha\alpha'},\end{aligned}$$

$$\sum_{s_1=0}^{\infty}\cdots\sum_{s_k=0}^{\infty}C^*_{k,M,\alpha}(s_1,\dots,s_k)C_{k,M,\alpha'}(s_1,\dots,s_k) = \delta_{\alpha\alpha'}.$$

Then at the time moment t, to solve problem (21) with the initial condition (22), (23) the operator $\overline{\widehat{E}}\Phi(t)$ has a form similar to (22), only $C_{k,M,\alpha}(s_1,\dots,s_N)$ is replaced by

$$C_{k,M,\alpha}(s_1,\dots,s_k)\exp\bigg(-it\theta\ln\bigg(\frac{a_{k,M,\alpha}}{\sum_{k',M',\beta}a_{k,M,\alpha}}\bigg)\bigg)$$

and $\Psi_{k,M,\alpha}(x_1,\dots,x_k,y_1,\dots,y_{2M})$ is replaced by

$$e^{-it\widehat{H}_{k+2M}}\Psi_{k,M,\alpha}.$$

3. Ultrasecondary quantization and quantization of thermodynamics in the case of a semidensity matrix

In statistical physics the mixed states of an N-particle system on the torus $\mathbf{T}^3$ is described [9] by the density matrix $\rho(x_1,\dots,x_N;y_1,\dots,y_N;t) \equiv \rho(x,y;t)$. This function satisfies the equation

$$i\frac{\partial}{\partial t}\rho(x,y;t) = \big(H_N(x) - H_N(y)\big)\rho(x,y;t) \tag{24}$$

and the conditions: $\widehat{\rho}^+ = \widehat{\rho}$ and $\widehat{\rho}$ is a nonnegative defined operator in $L_2(\mathbf{T}^{3N})$ whose kernel is $\rho(x,y;t)$.

Next, we consider the semidensity matrix

$$d(x_1,\dots,x_N;y_1,\dots,y_N).$$

The density matrix is expressed via $d(x,y)$ as follows:

$$\begin{aligned}
&\rho(x_1,\dots,x_N;y_1,\dots,y_N)\\
&\quad= \int\dots\int dz_1\dots dz_n d^*(x_1,\dots,x_N;z_1,\dots,z_N)d(y_1,\dots,y_N;z_1,\dots,z_N)\\
&\quad= \rho^*(y_1,\dots,y_N;x_1,\dots,x_N).
\end{aligned} \tag{25}$$

In the boson case, the functions $\rho(x,y)$ and $d(x,y)$ are symmetric with respect to the variables $x_1,\dots,x_N$ and the variables $y_1,\dots,y_N$. In the fermion case, these functions are antisymmetric.

Suppose that the semidensity matrix satisfies the equation

$$\begin{aligned}
&\Big\{-\frac{\hbar^2}{2m}\sum_{j=1}^{N}(\Delta x_j - \Delta y_j) + \sum_{j=1}^{N}\sum_{l=j+1}^{N}\big(V(x_j-x_l) - V(y_j-y_l)\big)\Big\}\\
&\qquad\cdot d(x_1,\dots,x_N;y_1,\dots,y_N) = \lambda d(x_1,\dots,x_N;y_1,\dots,y_N).
\end{aligned} \tag{26}$$

Then the density matrix $\rho(x,y)$ of the form (25) is a stationary solution of Eq. (24).

Let us study the ultrasecondary quantization of the equation for the density matrix (26).

Similarly to Section 2, we introduce the boson Fock space $\mathcal{L}$. Here $\widehat{b}^+(x,y)$, $\widehat{b}^-(x,y)$ are the creation and annihilation operators in the space $\mathcal{L}$ and D_0 is the vacuum vector. All the vectors in $\mathcal{L}$ can be represented uniquely in the form

$$\begin{aligned}
D &= \sum_{k=0}^{\infty}\frac{1}{\sqrt{k!}}\int\dots\int dx_1dy_1\dots dx_kdy_kd_k(x_1,\dots,x_k;y_1,\dots,y_k)\\
&\qquad\cdot\widehat{b}^+(x_1,y_1)\cdot\ldots\cdot\widehat{b}^+(x_k,y_k)D_0.
\end{aligned}$$

The semidensity matrix in the representation of the ultrasecondary quantization satisfies the equation

$$\overline{\overline{\widehat{W}}}D = \lambda\overline{Y}D, \tag{27}$$

where $\overline{\overline{\widehat{W}}}$ is the ultrasecondary quantized Hamiltonian of the equation for the semi-density matrix and $\overline{Y}$ is an operator in the space $\mathcal{L}$, the ultrasecondary quantized unity operator.

In the boson case we have

$$\overline{\overline{\widehat{W}}} = \sum_{k=0}^{\infty} \frac{1}{k!} \int \cdots \int dx_1 \ldots dx_k dy_1 \ldots dy_k$$
$$\cdot \widehat{b}^+(x_1, y_1) \cdot \ldots \cdot \widehat{b}^+(x_k, y_k) \big(\widehat{H}_k(x) - \widehat{H}_k(y)\big)$$
$$\cdot \operatorname*{Symm}_{x_1 \ldots x_k} \big(\widehat{b}^-(x_1, y_1) \cdot \ldots \cdot \widehat{b}^-(x_k, y_k)\big) \exp\Big(- \iint dx\, dy\, \widehat{b}^+(x, y) \widehat{b}^-(x, y)\Big),$$

as well as

$$\overline{\widehat{Y}} = \sum_{k=0}^{\infty} \frac{1}{k!} \int \cdots \int dx_1 \ldots dx_k dy_1 \ldots dy_k \widehat{b}^+(x_1, y_1) \cdot \ldots \cdot \widehat{b}^+(x_k, y_k)$$
$$\cdot \operatorname*{Symm}_{x_1 \ldots x_k} \big(\widehat{b}^-(x_1, y_1) \cdot \ldots \cdot \widehat{b}^-(x_k, y_k)\big) \exp\Big(- \iint dx\, dy\, \widehat{b}^+(x, y) \widehat{b}^-(x, y)\Big),$$

where, form now on, the operators $\widehat{b}^+(x, y)$ and $\widehat{b}^-(x, y)$ are Wick ordered. In the fermion case the operators $\overline{\overline{\widehat{W}}}$ and $\overline{\widehat{Y}}$ are expressed by similar formulas where Symm is replaced by Asymm.

In the limit, when the commutators $\widehat{b}^-(x, y)$ and $\widehat{b}^+(x, y)$ are small like $1/N$ as $N \to \infty$, according to [5–7] we set the operators $\widehat{b}^+(x, y)$ and $\widehat{b}^-(x, y)$ to be c-numbers. Then we obtain the following symbol corresponding to Eq. (27) in the boson case:

$$X[b^*(\cdot), b(\cdot)] = \Big\{ \sum_{k=0}^{\infty} \frac{1}{k!} \int \cdots \int dx_1 \ldots dx_k dy_1 \ldots dy_k b^*(x_1, y_1) \cdot \ldots$$
$$\cdot b^*(x_k, y_k) \big(\widehat{H}_k(x) - \widehat{H}_k(y)\big) \operatorname*{Symm}_{x_1 \ldots x_k} \big(b(x_1, y_1) \cdot \ldots \cdot b(x_k, y_k)\big) \Big\}$$
$$\cdot \Big\{ \sum_{k'=0}^{\infty} \frac{1}{k'!} \int \cdots \int dx'_1 \ldots dx'_k dy'_1 \ldots dy'_{k'} b^*(x'_1, y'_1) \cdot \ldots$$
$$\cdot b^*(x'_{k'}, y'_{k'}) \operatorname*{Symm}_{x'_1 \ldots x'_k} \big(b(x'_1, y'_1) \cdot \ldots \cdot b(x'_{k'}, y'_{k'})\big) \Big\}^{-1}. \tag{28}$$

In the fermion case the symbol $X[b^*(\cdot), b(\cdot)]$ is expressed by a similar formula where Symm is replaced by Asymm.

If the number of particles in the problem is fixed, then to Eq. (27) we add the condition

$$\overline{\mathcal{N}} D = N D. \tag{29}$$

Here $\overline{\mathcal{N}}$ is an operator in the space $\mathcal{L}$, which in the boson case has the form

$$\overline{\widehat{\mathcal{N}}} = \sum_{k=0}^{\infty} \frac{1}{k!} k \int \dots \int dx_1 \dots dx_N dy_1 \dots dy_k \widehat{b}^+(x_1, y_1) \cdot \ldots \cdot \widehat{b}^+(x_k, y_k)$$
$$\cdot \operatorname*{Symm}_{x_1 \dots x_k} \left(\widehat{b}^-(x_1, y_1) \cdot \ldots \cdot \widehat{b}^-(x_k, y_k)\right) \exp\left(- \iint dx\, dy\, \widehat{b}^+(x, y) \widehat{b}^-(x, y)\right).$$

We set $\widehat{b}^+(x,y)$ and $\widehat{b}^+(x,y)$ to be c-numbers. Then to Eq. (29) there corresponds the symbol

$$\mathcal{N}[b^*(\cdot), b(\cdot)] = \left\{ \sum_{k=0}^{\infty} \frac{1}{k!} k \int \dots \int dx_1 \dots dx_k dy_1 \dots dy_k \right.$$
$$\left. \cdot b^*(x_1, y_1) \cdot \ldots \cdot b^*(x_k, y_k) \operatorname*{Symm}_{x_1 \dots x_k} \left(b(x_1, y_1) \cdot \ldots \cdot b(x_k, y_k)\right) \right\}$$
$$\cdot \left\{ \sum_{k'=0}^{\infty} \frac{1}{k'!} \int \dots \int dx'_1 \dots dx'_{k'} dy'_1 \dots dy'_{k'} b^*(x'_1, y'_1) \cdot \ldots \cdot b^*(x'_{k'}, y'_{k'}) \right.$$
$$\left. \cdot \operatorname*{Symm}_{x'_1 \dots x'_{k'}} \left(b(x'_1, y'_1) \cdot \ldots \cdot b(x'_{k'}, y'_{k'})\right) \right\}^{-1}. \tag{30}$$

In the fermion case the operator $\overline{\widehat{\mathcal{N}}}$ and the symbol $\mathcal{N}$ are expressed similarly, only Symm is replaced by Asymm.

Let us consider the extremum of $X[b^*(\cdot), b(\cdot)]$ under the condition $\mathcal{N}[b^*(\cdot), b(\cdot)] = N$. The extremum equation has the form

$$\frac{\delta X}{\delta b^*(x,y)} - \mu \frac{\delta \mathcal{N}}{\delta b^*(x,y)} = 0, \qquad \frac{\delta X}{\delta b(x,y)} - \mu \frac{\delta \mathcal{N}}{\delta b(x,y)} = 0. \tag{31}$$

Taking into account the form of the functionals (28) and (30), we can rewrite Eqs. (31) as

$$[\widehat{L}, \widehat{A}] = 0, \tag{32}$$

where $\widehat{A}$ is an operator in $L_2(\mathbf{T}^3)$ with the kernel

$$\begin{aligned} A(x,y) \;&=\; -\frac{\hbar^2}{2m} \Delta_x \delta(x-y) + \delta(x-y) \int dz \\ &\quad \cdot V(x-z) L(z,z) \pm V(x-y) L(y,x) - \mu \delta(x-y). \end{aligned}$$

The kernel of the operator $\widehat{L}$ has the form

$$\begin{aligned} L(x,y) \;&=\; \sum_{k=1}^{\infty} \frac{(\pm 1)^{k-1}}{N^k} \int \dots \int dz_1 \dots dz_{2k-1}\, b(x_1, z_1) b^*(z_2, z_1) \\ &\quad \cdot b(z_3, z_2) b^*(z_4, z_3) \cdot \ldots \cdot b(z_{2k-1}, z_{2k}) b^*(y, z_{2k-1}). \end{aligned}$$

Possible solutions of Eq. (32) have the form $\widehat{L} = f(\widehat{A})$. In particular, for

$$f(\xi) = \frac{1}{e^{\xi/\theta} \mp 1} \pm \frac{1}{2}.$$

we obtain the temperature solutions.

Everywhere below the upper sign corresponds to the boson case and the lower sign to the fermion case.

The thermodynamics quantization for the density matrix is carried out similarly to Section 2.

The entropy is determined as follows:

$$S = -\operatorname{Sp}(\widehat{r}(D) \ln \widehat{r}(D)),$$

where $\widehat{r}(D)$ is a secondary quantized operator, depending on $D \in \mathcal{L}$, of the form

$$\begin{aligned}\widehat{r}(D) = \Bigg\{ \sum_{k=0}^{\infty} \frac{1}{(k!)^2} \int \dots \int dx_1 dy_1 dz_1 \dots dx_k dy_k dz_k \\ \cdot \widehat{\psi}^+(x_1) \cdot \ldots \cdot \widehat{\psi}^+(x_k) \widehat{\mathcal{P}}_0 \widehat{\psi}^-(y_k) \cdot \ldots \cdot \widehat{\psi}^-(y_1) \\ \cdot \Bigg(D, \overset{2}{\widehat{b}^+} (x_1, z_1) \dots \overset{2}{\widehat{b}^+} (x_k, z_k) \exp \Bigg(- \iint d\xi d\eta\, \overset{2}{\widehat{b}^+} (\xi, \eta)\, \overset{1}{\widehat{b}^-} (\xi, \eta) \Bigg) \\ \cdot \overset{1}{\widehat{b}^-} (y_k, z_k) \dots \overset{1}{\widehat{b}^-} (y_1, z_1) D \Bigg) \Bigg\} \Big/ (D, \widehat{\overline{Y}} D). \qquad (33)\end{aligned}$$

Similarly to (17) and (18), the functionals of intrinsic energy $\mathcal{E}$ and of the number of particles are introduced as follows:

$$\mathcal{E} = \operatorname{Sp}(\widehat{H}\widehat{r}), \qquad N = \operatorname{Sp}(\widehat{H}\widehat{r}).$$

Correspondingly, in the case of the density matrix the eigenvalues of intrinsic energy are defined to be the extremum values of the functional $\mathcal{E}$ under the additional conditions $N = \text{const}$, $S = \text{const}$. In this case the extremum equation is similar to Eq. (19) and the eigenvalues of this equations are called the eigenvalues of free energy in the case of ultrasecondary quantization of the density matrix.

References

1. V. P. Maslov, Funktsional. Anal. i Prilozhen., 2000, v.34, N4, 35–48.
2. V. P. Maslov, Teoret. Mat. Fiz., 2000, v.125, N2, 297–314.
3. V. P. Maslov, Teoret. Mat. Fiz., 2001, v.129, N3, 464–490.
4. V. P. Maslov, Complex WKB Method in Nonlinear Equations, Nauka, Moscow, 1977.
5. V. P. Maslov, Perturbation Theory and Asymptotic Methods, Nauka, Moscow, 1978.
6. V. P. Maslov, O. Yu. Shvedov, Complex Germ Method in Many-Particle Problems and Problems of Quantum Field Theory, URSS, Moscow, 2000.
7. F. A. Berezin, Secondary Quantization Method, Nauka, Moscow, 1987.
8. E. M. Lifshits, L. P. Pitaevskii, Statistical Physics, Nauka, Moscow, 1979.

APPROXIMATE DISTRIBUTION OF HITTING PROBABILITIES FOR A REGULAR SURFACE WITH COMPACT SUPPORT IN 2D

D. S. GREBENKOV (denis.grebenkov@polytechnique.fr)
Laboratoire de Physique de la Matière Condensée
Ecole Polytechnique
91128 Palaiseau Cedex France

Department of Statistical Physics
Saint Petersburg State University
ul. Ulyanovskaya 1
Petrodvorets, 198904
Saint Petersburg, Russia

Abstract.
Generalizing the well-known relations on characteristic functions on a plane to the case of a one-dimensional regular surface (curve) with compact support, we establish implicit equations for these functions. After solving the combinatorial problems, we introduce an approximation allowing to reduce these equations to a set of linear equations for a finite number of unknown functions. Imposing natural conditions, we obtain a closed system of linear equations which can be solved for a given surface. Its solutions can be used to approximate the distribution of hitting probabilities for a regular surface with compact support.
In order to verify the accuracy of the approximation, numerical analysis is being made for a chosen surface.

Introduction

Search for the distribution of hitting probabilities is an old and a well-known problem. Consider random walk on d-dimensional lattice (in continuous case consider Brownian motion). Then fix a surface of interest $\mathcal{S}$. Suppose that any random walk starts from a given point z which does not lie on $\mathcal{S}$. The problem is to calculate the distribution $P_z(x)$ of probabilities of first contact with points x of the surface $\mathcal{S}$. In other words, we are looking for the probability that random walks from $z \notin \mathcal{S}$ to $x \in S$ do not touch other points $y \in \mathcal{S}\backslash\{x\}$. Of course, the distribution $P_z(x)$ depends on z and $\mathcal{S}$.

V. A. Malyshev and A. M. Vershik (eds.),
Asymptotic Combinatorics with Application to Mathematical Physics, 221–242.

This problem has been solved for some particular surfaces. For example, the case of a planar surface in 2D (an ordinary straight line) is described in any book on probability theory (see [1], [2]). Its generalization for d-dimensional hyperplane is also simple (Section 2). Note that exact solutions have been found only for some particular surfaces but not in the general case. In the general the asymptotic behavior is widely studied, [2].

Problems of the hitting probabilities do not only have a purely mathematical interest. They are important for a wide class of physical problems, in particular, for the problems of Laplacian transfer across an interface, for instance, diffusion through a membrane, electrod problems, heterogeneous catalysis, etc. (see [3], [4], [5] for details). Indeed, if we are interested in diffusion through a semi-permeable membrane (points of this membrane can absorb or reflect touching particle with certain probabilities), we can write the total probability of absorption by a chosen point of the membrane as a sum of probabilities to be absorbed after 0, 1, 2, etc. reflections (rigorous formalism is described in [5], [6]). Here we face the task to calculate the distribution of hitting probabilities. Note that using this distribution solely for a planar membrane, we have recently obtained some important results about general characteristics of the Laplacian transfer across an interface, [7]. To solve these problems one needs to know the distribution of hitting probabilities for a general surface. As exact solution does not exist, one can try to find an approximation. Here we propose a method to approximate the distribution of hitting probabilities for a rather general case in 2D.

In the first section we introduce definitions and conditions which are required in what follows. In the second section we briefly describe a well-known case of the hitting probabilities on a horizontal axis. Main results are contained in the third section. Section 4 is devoted to some numerical results. In the last section we make conclusions and discuss possible generalizations.

1. Definitions

Consider a square lattice on a plane. Let us define a *regular surface*[1] *with compact support* $\mathcal{S} = \{(x, S(x))\}$ by a function $S(x)$ with integer x subject to the following conditions:

1. *Bijection*: The function $S(x)$ is a bijection between the set of integer numbers (absiccae x) and the set of surface points;

2. *Regularity*: For any x, $|S(x+1) - S(x)| \leq 1$;

3. *Compactness*: $\exists M : S(x) = 0$ for $|x| \geq M$, i.e. the non-plane part of the surface has a finite size. In other words, function $S(x)$ has a compact support. Moreover, we suppose that the surface is centered: $S(\pm(M-1)) \neq 0$.

Let us briefly discuss this definition. The second condition allows to simplify all calculations and formulae, but it does not seem to be essential (see Section 5). Note that this assumption can be viewed as *a regularity condition* for the surface in continuous case: $S'(x) \leq 1$.

[1] Even for two-dimensional case we prefer to use the word "surface" instead of "curve" or something else.

On the contrary, the third condition is important. It tells us that the surface in question is a finite "perturbation" of a planar surface (line). In other words, this surface is composed of two parts: a complex but compact part in the center with two plane "tails". Moreover, it is important that both tails lie on the same height (which is chosen as 0). This feature will allow us to obtain an approximate distribution of hitting probabilities by using the same ideas as for a planar surface (see Section 2).

We call all the points $\{(x,y)\ :\ y=n\}$ the n^{th} *level.* Denote

$$N = \max\{S(x)\}+1, \qquad N^* = -\min\{S(x)\}+1,$$

i.e. the surface lies between $(-N^*)$th and Nth levels.

All points $\mathcal{M} = \{(x,y)\ :\ \forall x\quad y < S(x)\}$ are called *internal.* All points $\mathcal{E} = \{(x,y)\ :\ \forall x\quad y > S(x)\}$ are called *external.* The external points near the surface, $\{(x, S(x)+1)\}$ are called *near-boundary points.* The functions defined on these points, are called *near-boundary functions* (see below). Often we will use the words "surface", "near-boundary functions", etc. thinking only about the non-trivial part, i.e. for $|x| < M$.

The external points with $y=0$ are called *ground points.* The functions defined on these points, are called *ground functions.* Let $J = \{k\ :\ (k,0)\in\mathcal{E}\}$ the set of abscissae of ground points. Let also $J_0 = \{k\in(-M,M)\ :\ (k,0)\in S\}$ the set of abscissae of boundary points on zeroth level (only non-plane part!).

Now, we introduce the hitting probabilities $P_{k,n}(x)$, i.e. the probability of the first contact with the surface at point $(x,S(x))$ if started from (k,n). Their characteristic functions $\phi_{m,n}(\theta)$ are

$$\phi_{k,n}(\theta) = \sum_{x=-\infty}^{\infty} P_{k,n}(x)e^{ix\theta}. \tag{1}$$

The inverse Fourier transform allows to obtain $P_{k,n}(x)$,

$$P_{k,n}(x) = \int_{-\pi}^{\pi} \frac{d\theta}{2\pi} e^{-ix\theta}\phi_{k,n}(\theta). \tag{2}$$

2. Planar surface

At the beginning, we consider the trivial and well-known case of a planar surface (horizontal axis): $S(x)=0$. This case is useful to remind the technique of manipulation with characteristic functions.

Suppose that $n>0$. The probability $P_{k,n}(x)$ satisfies a simple identity

$$P_{k,n}(x) = \frac{1}{4}\left[P_{k+1,n}(x)+P_{k-1,n}(x)+P_{k,n+1}(x)+P_{k,n-1}(x)\right], \tag{3}$$

which can be also written for characteristic functions,

$$\phi_{k,n}(\theta) = \frac{1}{4}\left[\phi_{k+1,n}(\theta) + \phi_{k-1,n}(\theta) + \phi_{k,n+1}(\theta) + \phi_{k,n-1}(\theta)\right]. \tag{4}$$

Due to translational invariance along the horizontal axis,

$$\phi_{k,n}(\theta) = e^{ik\theta}\phi_{0,n}(\theta). \tag{5}$$

Using the obvious condition $P_{k,0}(x) = \delta_{k,x}$, we obtain

$$\phi_{k,0}(\theta) = e^{ik\theta}. \tag{6}$$

The last trick is the following. If the starting point is placed in the n-th level, the random walk must cross the $(n-1)$-th level at some point $(m, n-1)$ to reach zeroth level. The probability to pass from (k,n) to $(m, n-1)$ without touching other points in the $(n-1)$-th level is exactly $P_{k,1}(m)$. Therefore we can write

$$P_{k,n}(x) = \sum_m P_{m,n-1}(x)P_{k,1}(m).$$

In terms of characteristic functions this convolution is just a product of the two corresponding characteristic functions,

$$\phi_{k,n}(\theta) = \phi_{k,n-1}(\theta)\phi_{k,1}(\theta).$$

This means simply that

$$\phi_{k,n}(\theta) = [\phi_{k,1}(\theta)]^n. \tag{7}$$

Substitution of expressions (5), (6) and (7) into relation (4) for $n = 1$ and $k = 0$ leads to

$$\phi_{0,1}(\theta) = \frac{1}{4}\left(e^{-i\theta}\phi_{0,1}(\theta) + e^{i\theta}\phi_{0,1}(\theta) + 1 + [\phi_{0,1}(\theta)]^2\right), \qquad \text{or}$$

$$\phi_{0,1}^2 - (4 - 2\cos\theta)\phi_{0,1} + 1 = 0. \tag{8}$$

This quadratic equation has two solutions, and we should choose the one for which $\phi_{0,1}(\theta) \le 1$ (property of characteristic function). It is denoted $\varphi(\theta)$,

$$\varphi(\theta) = 2 - \cos\theta - \sqrt{(2-\cos\theta)^2 - 1}. \tag{9}$$

Using again expressions (5) and (7), we obtain for the planar surface

$$\phi_{k,n}(\theta) = e^{ik\theta}\varphi^n(\theta). \tag{10}$$

Inverting this relation according to (2), we obtain the distribution of hitting probabilities for the planar surface,

$$P_{k,n}^{planar}(x) = \int_{-\pi}^{\pi} \frac{d\theta}{2\pi} e^{i(k-x)\theta}\varphi^n(\theta) = H_{k-x}^n. \tag{11}$$

This well-known result will be used for a general case (some properties of coefficients H_k^n are described in Appendix 1). The formulae (9) and (10) can be generalized for d-dimentional hypercubic lattice,

$$\varphi(\theta_1 \dots \theta_{d-1}) = d - \sum_{i=1}^{d-1} \cos(\theta_i) - \sqrt{\left(d - \sum_{i=1}^{d-1} \cos(\theta_i)\right)^2 - 1},$$

$$\phi_{\mathbf{x},n}(\theta_1 \dots \theta_{d-1}) = \exp\left[i \sum_{i=1}^{d-1} x_i \theta_i\right] \varphi^n(\theta_1 \dots \theta_{d-1}).$$

3. Regular surface with compact support

We will consider the characteristic functions $\phi_{k,n}(\theta)$ as a vector

$$\Phi_n(\theta) = \begin{pmatrix} \phi_{-L,n}(\theta) \\ \phi_{-L+1,n}(\theta) \\ \dots \\ \phi_{L,n}(\theta) \end{pmatrix}$$

of $(2L+1)$ components where parameter L is supposed large, and it will tend to infinity at the end of calculation.

For the planar surface we had relation (4) which can be written in matrix form

$$A\Phi_n = \Phi_{n-1} + \Phi_{n+1}, \tag{12}$$

where the matrix A is

$$A_{i,i} = 4, \qquad A_{i,i+1} = A_{i+1,i} = -1, \qquad A_{2L+1,1} = A_{1,2L+1} = -1.$$

The last equalities are artificial: we added them to obtain a cyclic structure of A. But at the limit $L \to \infty$ this little modification vanishes. The eigenvalues of A are

$$\lambda_m = 4 - 2\cos(\theta_m), \qquad \text{with } \theta_m = \frac{2\pi m}{2L+1},$$

and the eigenvectors are given as

$$V_m = \begin{pmatrix} e^{-iL\theta_m} \\ e^{-i(L-1)\theta_m} \\ \dots \\ e^{iL\theta_m} \end{pmatrix}.$$

Now we will generalize the relation (12) to the case of a regular surface with compact support by introducing vector I_n,

$$A\Phi_n = \Phi_{n-1} - \Phi_{n+1} + I_n \tag{13}$$

(this relation can be regarded as the definition of I_n).

The crucial point of this method is the possibility to treat the walks above the N-th level as previously, in Section 2. It means that

$$\Phi_{N+1} = \varphi \Phi_N, \tag{14}$$

where function φ is defined by (9). This property allows us to close the system of vector equations (13),

$$A\Phi_N = \varphi \Phi_N + \Phi_{N-1} + I_N.$$

The analogous relation in the lower half plane is simply

$$\Phi_{-N^*} = 0, \tag{15}$$

because it is impossible to penetrate through the surface. More generally, according to the definition of hitting probabilities, we should maintain

$$\phi_{m,n}(\theta) = \begin{cases} e^{im\theta}, & \text{if } (m,n) \in \mathcal{S}, \\ 0, & \text{if } (m,n) \in \mathcal{M}. \end{cases} \tag{16}$$

Under this *boundary condition* we are going to find $\phi_{m,n}(\theta)$ on external points $(m,n) \in \mathcal{E}$.

For any $-N^* < n < N$ let us introduce

$$c_n(\theta, \theta_m) = (V_m, \Phi_n), \qquad b_n(\theta, \theta_m) = (V_m, I_n). \tag{17}$$

We can rewrite (13), (14) and (15) in terms of c_n and b_n,

$$\lambda_m c_n = c_{n-1} + c_{n+1} + b_n, \qquad c_{N+1} = \varphi c_N, \qquad c_{-N^*} = 0. \tag{18}$$

If we can express c_n in terms of λ_m, $\varphi(\theta)$, c_0 and $\{b_k\}$, we find Φ_n as a decomposition in the base of eigenvectors V_m,

$$\Phi_n = \frac{1}{2L+1} \sum_m c_n(\theta, \theta_m) V_m \tag{19}$$

(factor $(2L+1)^{-1}$ is due to normalization (V_m, V_m)).

The main idea is to step down from Nth and $(-N^*)$th levels to zeroth level. We will consider the upper and lower half planes separately because the relations (14) and (15) are different. Note the essential complication of the general case with respect to the planar surface. For the planar surface we had $b_n = 0$ for any n, and the system of equations (18) was closed. It was sufficient to solve these recurrence relations by substitution $c_n = c_0 c^n$ in (18), and we obtained the final form of Φ_n. On the contrary, for the general surface the coefficients $b_n \neq 0$, and they depend on the near-boundary functions $\phi_{m,n}$ (see below). Consequently, the decomposition (19) itself becomes a system of implicit equations for $\phi_{m,n}$. For the moment, the problem is complex. It will be solved by several steps. First, we obviate the combinatorial problems, i.e. we solve the reccurence relations (18). Second, we propose an approximation to solve the equations for $\phi_{m,n}$.

3.1. SOLUTION OF RECCURENCE RELATIONS

A direct verification shows that

$$c_n = \beta_{N-n} c_N - \sum_{k=1}^{N-n-1} \alpha_k b_{n+k}, \tag{20}$$

$$c_{-n} = \alpha_{N^*-n} c_{-N^*+1} - \sum_{k=1}^{N^*-n-1} \alpha_k b_{-n-k} \tag{21}$$

is a general solution of (18) (we omitted the index m which does not change the structure of the solution), where

$$\alpha_0 = 0, \quad \alpha_1 = 1, \qquad \alpha_{n+2} = \lambda_m \alpha_{n+1} - \alpha_n, \tag{22}$$

$$\beta_n = \alpha_n(\lambda_m - \varphi) - \alpha_{n-1}, \qquad \beta_0 = 1. \tag{23}$$

Formulae (20) and (21) are valid for any $n \geq 0$, in particular, for $n = 0$, and we can express c_N and c_{-N^*+1} in terms of c_0 and $\{b_k\}$,

$$c_N = \frac{1}{\beta_N}\left(c_0 + \sum_{k=1}^{N-1} \alpha_k b_k\right), \qquad c_{-N^*+1} = \frac{1}{\alpha_{N^*}}\left(c_0 + \sum_{k=1}^{N^*-1} \alpha_k b_{-k}\right),$$

hence

$$c_n = \frac{\beta_{N-n}}{\beta_N}\left(c_0 + \sum_{k=1}^{N-1} \alpha_k b_k\right) - \sum_{k=1}^{N-n-1} \alpha_k b_{n+k}, \tag{24}$$

$$c_{-n} = \frac{\alpha_{N^*-n}}{\alpha_{N^*}}\left(c_0 + \sum_{k=1}^{N^*-1} \alpha_k b_{-k}\right) - \sum_{k=1}^{N^*-n-1} \alpha_k b_{-n-k}. \tag{25}$$

Let us introduce

$$f_n^{(N)}(\theta, \theta_m) = \frac{\beta_{N-n}}{\beta_N}, \qquad \tilde{f}_n^{(N^*)}(\theta_m) = \frac{\alpha_{N^*-n}}{\alpha_{N^*}}$$

(the right-hand side of $f_n^{(N)}$ depends on θ due to factor φ; the dependence on θ_m is due to λ_m containing in α_n).

Using the reccurence properties of α_k (see Appendix 2), we simplify relations (24) and (25),

$$c_n = f_n^{(N)}\left(c_0 + \sum_{k=1}^{n} \alpha_k b_k\right) + \alpha_n \sum_{k=n+1}^{N-1} f_k^{(N)} b_k, \tag{26}$$

$$c_{-n} = \tilde{f}_n^{(N^*)}\left(c_0 + \sum_{k=1}^{n} \alpha_k b_{-k}\right) + \alpha_n \sum_{k=n+1}^{N^*-1} \tilde{f}_k^{(N^*)} b_{-k}. \tag{27}$$

Note that the structure of these two solutions is similar. The only distinction is the function φ which appears in (14) for the upper half plane (whereas in the lower half plane we have (15), i.e. corresponding "function $\tilde{\varphi}$" is equal to 0).

So, we have solved the combinatorial problems.

3.2. APPROXIMATION

Let us consider again the definition (22) of coefficients α_n. We try to find the explicit solution for α_n in the form $\alpha_n = x^{n-1}$. Substituting this into (22), we obtain equation

$$x^2 - \lambda_m x + 1 = 0,$$

which has two well-known solutions: φ and φ^{-1} (compare this equation with (8)). As the expression (22) is linear, we find a general solution as linear combination of φ^{n-1} and φ^{1-n} such that $\alpha_0 = 0$. We obtain

$$\alpha_n = \frac{1-\varphi^{2n}}{1-\varphi^2}\varphi^{1-n}, \tag{28}$$

or as a geometrical sequence

$$\alpha_n = \sum_{k=0}^{n-1} \varphi^{2k+1-n}. \tag{29}$$

Rewriting the definition (23) of β_n as

$$\beta_n(\theta, \theta_m) = \alpha_{n+1}(\theta_m) - \varphi(\theta)\alpha_n(\theta_m),$$

after simplifications we obtain

$$f_n^{(N)}(\theta, \theta_m) = \frac{\varphi^n(\theta_m)[1-\varphi(\theta)\varphi(\theta_m)] - \varphi^{2N-n+1}(\theta_m)[\varphi(\theta_m)-\varphi(\theta)]}{[1-\varphi(\theta)\varphi(\theta_m)] - \varphi^{2N+1}(\theta_m)[\varphi(\theta_m)-\varphi(\theta)]}.$$

We know that $\varphi(\theta_m) \ll 1$ if θ_m is not in vicinity of 0 (see Section 4, Fig. 1), and in this case we can neglect φ^{2N-n+1} and φ^{2N+1}, thus[2]

$$f_n^{(N)}(\theta, \theta_m) \approx \varphi^n(\theta_m) \tag{30}$$

(for θ_m in vicinity of 0 there is a little difference between $f_n^{(N)}(\theta, \theta_m)$ and $\varphi^n(\theta_m)$ that depends, of course, on θ, see Section 4, Fig. 2). We are going to use this approximation in the following.

Using the formula (28), we obtain

$$\tilde{f}_n^{(N^*)}(\theta_m) = \frac{1-\varphi^{2(N^*-n)}(\theta_m)}{1-\varphi^{2N^*}(\theta_m)}\varphi^n(\theta_m).$$

[2] This approximation works better for larger N and smaller n. However, it will be used generally for any reasonable n and N.

Now it does not matter to keep terms $\varphi^{2(N^*-n)}$ and φ^{2N^*} in this expression, thus

$$\tilde{f}_n^{(N^*)}(\theta_m) \approx \varphi^n(\theta_m). \tag{31}$$

Note that we can establish all following results without approximation (31) for $\tilde{f}_n^{(N^*)}$ (the formulae will be just more complex). On the contrary, the approximation (30) for $f_n^{(N)}$ is essential because it allows to separate the dependencies on θ and θ_m.

3.3. COEFFICIENT c_0

Let us calculate the coefficient c_0.

As it was mentioned above, the plane "tails" of the surface lie on the zeroth level, thus

$$(\Phi_0)_k = e^{ik\theta}, \text{ if } |k| \geq M.$$

Using this explicit form, we are going to compute the contribution $c_0^{(0)}$ of plane "tails"

$$c_0^{(0)}(\theta, \theta_m) = (V_m, \Phi_0) = \sum_{k=-L}^{L} e^{-ik\theta_m}\phi_{k,0} = 2\sum_{k=M}^{L} \cos(\theta - \theta_m),$$

and, with the help of a trigonometrical identity, we obtain

$$c_0^{(0)}(\theta, \theta_m) = \frac{\sin(L+0.5)(\theta-\theta_m)}{\sin(\theta-\theta_m)/2} - \frac{\sin(M-0.5)(\theta-\theta_m)}{\sin(\theta-\theta_m)/2} \tag{32}$$

(in the case $\theta = \theta_m$ one should consider this relation in the limit sense when $\theta \to \theta_m$). Note the simple relation

$$\int_{-\pi/2}^{\pi/2} \frac{d\theta}{\pi} \cos(2k\theta) \frac{\sin(2M-1)\theta}{\sin\theta} = \chi_{(-M,M)}(k), \tag{33}$$

where for any integer k we define

$$\chi_A(k) = \begin{cases} 1, & \text{if } k \in A, \\ 0, & \text{if } k \notin A. \end{cases}$$

The contribution $c_0^{(1)}$ of intermediate part of the surface (with $|x| < M$) can contain some non-trivial functions $\phi_{m,0}(\theta)$, with $k \in J$ and $k \in J_0$

$$c_0^{(1)}(\theta, \theta_m) = \sum_{k \in J} \phi_{k,0}(\theta) e^{-ik\theta_m} + \sum_{k \in J_0} e^{ik(\theta-\theta_m)}. \tag{34}$$

3.4. LIMIT $L \to \infty$

Using the solutions (26) and (27), we can write the integral expression for $\phi_{k,n}$ by taking the limit $L \to \infty$ (here we write only the expression for $n \geq 0$; the opposite case will be easily obtained later)[3]

$$\phi_{k,n}(\theta) = f_n^{(N)}(\theta,\theta)e^{ik\theta} - \int_{-\pi}^{\pi} \frac{d\theta'}{2\pi} e^{ik\theta'} f_n^{(N)}(\theta,\theta') \frac{\sin(M-0.5)(\theta-\theta')}{\sin(\theta-\theta')/2}$$
$$+ \int_{-\pi}^{\pi} \frac{d\theta'}{2\pi} e^{ik\theta'} \left(f_n^{(N)}(\theta,\theta') \left[c_0^{(1)}(\theta,\theta') + \sum_{l=1}^{n} \alpha_l(\theta') b_l(\theta,\theta') \right] \right.$$
$$\left. + \alpha_n(\theta') \sum_{l=n+1}^{N-1} f_l^{(N)}(\theta,\theta') b_l(\theta,\theta') \right). \quad (35)$$

The first term which is exactly equal to $\varphi^n(\theta)e^{ik\theta}$ is the contribution of plane "tails". The second term corresponding to the perturbation on zeroth level due to $c_0^{(0)}$ is denoted

$$\phi_{k,n}^{(0)}(\theta) = \int_{-\pi}^{\pi} \frac{d\theta'}{2\pi} e^{ik\theta'} f_n^{(N)}(\theta,\theta') \frac{\sin(M-0.5)(\theta-\theta')}{\sin(\theta-\theta')/2}. \quad (36)$$

If we apply the approximation (30) to the formulae (35) and (36), we obtain

$$\phi_{k,n}(\theta) = \varphi^n(\theta)e^{ik\theta} - \phi_{k,n}^{(0)}(\theta) +$$
$$\int_{-\pi}^{\pi} \frac{d\theta'}{2\pi} e^{ik\theta'} \varphi^n(\theta') c_0^{(1)}(\theta,\theta') + \int_{-\pi}^{\pi} \frac{d\theta'}{2\pi} e^{ik\theta'} \sum_{l=1}^{N-1} \gamma_l^{(n)}(\theta') b_l(\theta,\theta'), \quad (37)$$

where we introduced the function

$$\gamma_l^{(n)}(\theta) = \begin{cases} \varphi^n(\theta)\alpha_l(\theta), & \text{if } l \leq n, \\ \varphi^l(\theta)\alpha_n(\theta), & \text{if } l > n, \\ 0, & \text{if } l \leq 0 \end{cases} \quad (38)$$

(the last convention will be used in the following).

The last term of (37) contains some unknown functions $\phi_{m,l}(\theta)$ through the coefficients $b_l(\theta,\theta')$. It is denoted as $T[\phi]$, and we are going to calculate it.

[3] Note that the first term of (32), $\sin(2L+1)(\theta-\theta_m)/\sin(\theta-\theta_m)$, tends to δ-function in the limit $L \to \infty$. It removes the integration in the first term.

3.5. COEFFICIENTS b_l

To get ahead with the expression (37), we should write explicitly the coefficients $b_l(\theta,\theta')$. It is not so easy for the general case. Indeed, for these purposes one can calculate the contributions of each point on l-th level. The problem is that there are many conditions, and they lead to complex formulae difficult to manipulate. We are going to present the other way.

What is the origin of the vector (I_n)? Let us recall the definition (13) where vectors (I_n) were introduced to generalize the expression (12). A brief reflection shows that

$$(I_n)_m = \begin{cases} 0, & \text{if } (m,n) \in \mathcal{E}, \\ 4\phi_{m,n} - \phi_{m-1,n} - \phi_{m+1,n} - \phi_{m,n-1} - \phi_{m,n+1}, & \text{if } (m,n) \notin \mathcal{E} \end{cases} .$$

In other words, the relation (12) is satisfied automatically for any external point, but it should be imposed for each surface and internal points.

Now it is the moment to recall the formula (16) which tells that functions $\phi_{m,n}(\theta)$ are equal to zero on the internal points. Therefore we can consider only the points near the surface S. A direct verification shows that

$$\begin{aligned}(I_{S(m)})_m = 4e^{im\theta} - \phi_{m,S(m)+1} &- \begin{cases} e^{i(m+1)\theta}, & \text{if } S(m+1) - S(m) = 0 \\ \phi_{m+1,S(m)}, & \text{if } S(m+1) - S(m) = -1 \\ 0, & \text{if } S(m+1) - S(m) = 1 \end{cases} \\ &- \begin{cases} e^{i(m-1)\theta}, & \text{if } S(m-1) - S(m) = 0 \\ \phi_{m-1,S(m)}, & \text{if } S(m-1) - S(m) = -1 \\ 0, & \text{if } S(m-1) - S(m) = 1, \end{cases}\end{aligned} \tag{39}$$

$$\begin{aligned}(I_{S(m)-1})_m &= -e^{im\theta} - e^{i(m+1)\theta}\delta_{S(m),S(m+1)+1} - e^{i(m-1)\theta}\delta_{S(m),S(m-1)+1}, \\ (I_n)_m &= 0, \quad \text{if } n \neq S(m) \text{ and } n \neq S(m) - 1\end{aligned}$$

(here we use Kronecker δ-symbols, $\delta_{ii} = 1$ and $\delta_{ij} = 0$ if $i \neq j$). Usually there are several nonzero components of (I_n) for each n, because each level contains several surface points. But there exists one exception – zeroth level, where there is infinity of surface points due to the plane "tails". Thus, the vector (I_{-1}) has exceptional structure[4]. It contains the usual terms due to the non-trivial part of the surface, and the contribution of plane "tails". Note that the last one is equal to $-c_0^{(0)}$ which was calculated in Section 3.3. Later we will use this result for the lower half plane.

3.6. APPROXIMATE DISTRIBUTION OF HITTING PROBABILITIES

According to the definition (17), we have

[4] There exists another vector with exceptional structure, (I_0). However, we construct our treatment by the way when this vector does not appear in expressions (see (26), (27)). In other words, we step down from N-th and $(-N^*)$-th levels to zeroth level in order to do not pass through the zeroth level.

$$T[\phi] = \int_{-\pi}^{\pi} \frac{d\theta'}{2\pi} e^{ik\theta'} \sum_{m=-M+1}^{M-1} e^{-im\theta'} \sum_{l=1}^{N-1} \gamma_l^{(n)}(\theta')(I_l)_m,$$

where we changed the order of summation over m and l. However, in the last sum there are only two terms corresponding to $(I_{S(m)})_m$ and $(I_{S(m)-1})_m$, if $S(m) \geq 1$ (otherwise, this sum is equal to 0). Using expressions (39), we obtain simply

$$\begin{aligned} T[\phi] = \sum_{m=-M+1}^{M-1} \int_{-\pi}^{\pi} \frac{d\theta'}{2\pi} e^{i(k-m)\theta'} \Bigg(& \gamma_{S(m)}^{(n)}(\theta') \Big[4e^{im\theta} - \phi_{m,S(m)+1} \\ & - e^{i(m+1)\theta} \delta_{S(m),S(m+1)} - \phi_{m+1,S(m)} \delta_{S(m),S(m+1)+1} \\ & - e^{i(m-1)\theta} \delta_{S(m),S(m-1)} - \phi_{m-1,S(m)} \delta_{S(m),S(m-1)+1} \Big] \\ - \gamma_{S(m)-1}^{(n)}(\theta') \Big[& e^{im\theta} + e^{i(m-1)\theta} \delta_{S(m),S(m-1)+1} + e^{i(m+1)\theta} \delta_{S(m),S(m+1)+1} \Big] \Bigg). \end{aligned}$$

(here we have used the last convention in the definition (38) of $\gamma_l^{(n)}$ to avoid any terms with $S(m) \leq 0$).

The last step is to transform this huge expression for characteristic functions into hitting probabilities using the formula (2). Note that all functions $e^{im\theta}$ after integration over θ with $e^{-ix\theta}$ give δ-symbols that remove the summation over m in corresponding terms (but we should write factor $\chi_{(-M,M)}(x)$),

$$\begin{aligned} T[P] = - \sum_{m=-M+1}^{M-1} & D_{m,S(m)}^{(k,n)} \Big(P_{m,S(m)+1}(x) \\ & + P_{m+1,S(m)}(x) \delta_{S(m),S(m+1)+1} + P_{m-1,S(m)}(x) \delta_{S(m),S(m-1)+1} \Big) \\ + \chi_{(-M,M)}(x) \Big(& 4D_{x,S(x)}^{(k,n)} - D_{x-1,S(x-1)}^{(k,n)} \delta_{S(x),S(x-1)} - D_{x+1,S(x+1)}^{(k,n)} \delta_{S(x),S(x+1)} \\ - D_{x,S(x)-1}^{(k,n)} & - D_{x+1,S(x+1)-1}^{(k,n)} \delta_{S(x),S(x+1)-1} - D_{x-1,S(x-1)-1}^{(k,n)} \delta_{S(x),S(x-1)-1} \Big), \end{aligned} \tag{40}$$

where

$$D_{m,l}^{(k,n)} = \int_{-\pi}^{\pi} \frac{d\theta'}{2\pi} e^{i(k-m)\theta'} \gamma_l^{(n)}(\theta'). \tag{41}$$

Note that these coefficients $D_{m,l}^{(k,n)}$ are universal, they do not depend on a given surface. It means that once calculated, these coefficients can be used for any hitting problem in 2D. They can be also expressed in terms of H_k^n,

$$D_{m,l}^{(k,n)} = \sum_{j=0}^{\min\{n,l\}-1} H_{k-m}^{2j+1+|l-n|}$$

(if n or l is equal to 0, the sum is also equal to 0). We just indicate several useful properties of these coefficients,

$$D_{m,l}^{(k,n)} = D_{m-k,l}^{(0,n)} = D_{0,l}^{(k-m,n)} = D_{0,l}^{(m-k,n)}, \qquad D_{m,l}^{(k,0)} = D_{m,0}^{(k,n)} = 0.$$

The first part of (40), containing $P_{m,S(m)+1}$, can be represented as

$$T_1[P] = -\sum_{m=-M}^{M} G_m^{(k,n)} P_{m,S(m)+1}(x)$$

with coefficients

$$G_m^{(k,n)} = D_{m,S(m)}^{(k,n)} + \delta_{S(m),S(m+1)-1} D_{m+1,S(m+1)}^{(k,n)} + \delta_{S(m),S(m-1)-1} D_{m-1,S(m-1)}^{(k,n)}.$$

Now we can see that the essential advantage of using (30) is the factorization of dependencies on θ and θ' in the last term of (37).

The second part of (40) can be simplified. Indeed, using the properties of δ-symbols, we have

$$T_2 = \chi_{(-M,M)}(x)\Big(4D_{x,S(x)}^{(k,n)} - D_{x-1,S(x)}^{(k,n)} - D_{x+1,S(x)}^{(k,n)} - D_{x,S(x)-1}^{(k,n)} + D_{x-1,S(x)}^{(k,n)}\delta_{S(x),S(x-1)+1} + D_{x+1,S(x)}^{(k,n)}\delta_{S(x),S(x+1)+1}\Big).$$

Using (41), we finally obtain

$$T_2 = \chi_{(-M,M)}(x)\begin{cases} \int\limits_{-\pi}^{\pi} \frac{d\theta}{2\pi} e^{i(k-x)\theta}\varphi^{S(x)+1}(\theta)\alpha_n(\theta), & \text{if } S(x) > n, \\ \int\limits_{-\pi}^{\pi} \frac{d\theta}{2\pi} e^{i(k-x)\theta}\varphi^{n}(\theta)\alpha_{S(x)+1}(\theta), & \text{if } S(x) \le n \end{cases} + D_{x-1,S(x)}^{(k,n)}\delta_{S(x),S(x-1)+1} + D_{x+1,S(x)}^{(k,n)}\delta_{S(x),S(x+1)+1}. \quad (42)$$

Let us get back to the formula (37). Using the inverse Fourier transform (2), we write

$$P_{k,n}(x) = H_{k-x}^n - P_{k,n}^{(0)}(x) + \sum_{m\in J_0} H_{k-m}^n \delta_{m,x} + \sum_{m\in J} H_{k-m}^n P_{m,0}(x) + T_1[P] + T_2, \quad (43)$$

where the coefficients H_k^n defined by (11) are exactly the hitting probabilities for the planar surface.

Let us calculate the second term of (43),

$$P^{(0)}_{k,n}(x) = \int\limits_0^{2\pi} \frac{d\theta}{2\pi} e^{-ix\theta} \int\limits_{-\pi}^{\pi} \frac{d\theta'}{2\pi} e^{ik\theta'} \varphi^n(\theta') \frac{\sin(M-0.5)(\theta-\theta')}{\sin(\theta-\theta')/2}.$$

Replacing in the first integral $\theta_1 = \theta - \theta'$, we factorize these integrals. The first factor is exactly H^n_{k-x}. The second one was evaluated explicitly, see (33), and it equals to $\chi_{[-M+1,M-1]}(x)$. Consequently, the first two terms in (43) can be grouped into $H^n_{k-x}\chi_{(-\infty,-M]\cap[M,+\infty)}(x)$. It means that the solution H^n_{k-x} of the plane case is valid only for the plane "tails", whereas on the non-trivial surface (for $|x| < M$) the main contribution is due to other terms. So, we have obtained an important result,

$$P_{k,n}(x) = \tilde{P}_{k,n}(x) + \sum_{m\in J} H^n_{k-m} P_{m,0}(x) - \sum_{m=-M}^{M} G^{(k,n)}_m P_{m,S(m)+1}(x), \tag{44}$$

where

$$\tilde{P}_{k,n}(x) = H^n_{k-x}\chi_{(-\infty,-M]\cap[M,+\infty)}(x) + T_2 + H^n_{k-x}\chi_{J_0}(x)$$

(the third term is due to the second sum in (34)). For the trivial case $n = 0$ we have simply: $T_2 = 0$, $J_0 = \emptyset$, $H^n_{k-x} = \delta_{k,x}$, therefore $\tilde{P}_{k,0}(x) = \delta_{k,x}\chi_{(-\infty,-M]\cap[M,+\infty)}(x)$. Later we will suppose that $n \neq 0$.

Using (42), we can combine first two terms to obtain

for $n > 0$:

$$\tilde{P}_{k,n}(x) = \begin{cases} \int\limits_{-\pi}^{\pi} \frac{d\theta}{2\pi} \cos(x-k)\theta \left(\varphi^n(\theta)\alpha_{S(x)+1}(\theta) \right), & S(x) \le n, \\ \int\limits_{-\pi}^{\pi} \frac{d\theta}{2\pi} \cos(x-k)\theta \left(\varphi^{S(x)+1}(\theta)\alpha_n(\theta) \right), & S(x) > n \end{cases}$$
$$+ D^{(k,n)}_{x-1,S(x)}\delta_{S(x),S(x-1)+1} + D^{(k,n)}_{x+1,S(x)}\delta_{S(x),S(x+1)+1} + H^n_{k-x}\chi_{J_0}(x). \tag{45}$$

Using the explicit formula (29) for coefficients α_n and the definition (11) of H^n_k, we can rewrite (45) as

for $n > 0$:

$$\tilde{P}_{k,n}(x) = \begin{cases} \sum\limits_{j=0}^{S(x)} H^{2j+n-S(x)}_{k-x}, & S(x) \le n, \\ \sum\limits_{j=1}^{n} H^{2j+S(x)-n}_{k-x}, & S(x) > n \end{cases}$$
$$+ D^{(k,n)}_{x-1,S(x)}\delta_{S(x),S(x-1)+1} + D^{(k,n)}_{x+1,S(x)}\delta_{S(x),S(x+1)+1} + H^n_{k-x}\chi_{J_0}(x). \tag{46}$$

Here we use the convention that $\sum_{j=a}^{b}$ is equal 0 if $b < a$, i.e. $\tilde{P}_{k,n}(x)$ is equal to 0 for $S(x) < 0$.

To obtain analogous results for the case $n < 0$, we can remark that initial formulae (26) and (27) are almost identical. Indeed, it is sufficient to replace n by $-n$, N by N^*, b_k by b_{-k}, $f_n^{(N)}$ by $\tilde{f}_n^{(N^*)}$ to obtain (27) from (26). According to the approximations (30) and (31), $f_n^{(N)} \approx \varphi^n \approx \tilde{f}_n^{(N^*)}$, therefore these functions are considered as identical. The summation from $l = 1$ to $N - 1$ (or to $N^* - 1$) in (37) disappears due to expressions (39) of $(I_n)_m$. So, we see that in order to obtain analogous results for $n < 0$, we should "reflect" all "ordinates" with respect to horizontal axis. It means that we have

for $n < 0$:

$$P_{k,n}(x) = \tilde{P}^*_{k,n}(x) + \sum_{m \in J} H^{-n}_{k-m} P_{m,0}(x) - \sum_{m=-M}^{M} G^{(k,n)*}_m P_{m,S(m)+1}(x), \qquad (47)$$

with

$$\tilde{P}^*_{k,n}(x) = \chi_{(-M,M)}(x) \begin{cases} \sum\limits_{j=0}^{-S(x)} H^{2j-n+S(x)}_{k-x}, & -S(x) \leq -n, \\ \sum\limits_{j=1}^{-n} H^{2j-S(x)+n}_{k-x}, & -S(x) > -n \end{cases}$$
$$+ D^{(k,-n)}_{x-1,-S(x)} \delta_{S(x),S(x-1)-1} + D^{(k,-n)}_{x+1,-S(x)} \delta_{S(x),S(x+1)-1} + H^{-n}_{k-x} \chi_{J_0}(x), \qquad (48)$$

$$G^{(k,n)*}_m = D^{(k,-n)}_{m,-S(m)} + \delta_{S(m),S(m+1)+1} D^{(k,-n)}_{m+1,-S(m+1)} + \delta_{S(m),S(m-1)+1} D^{(k,-n)}_{m-1,-S(m-1)}.$$

In (48) there appears function $\chi_{(-M,M)}(x)$, because for $n < 0$ there is no contribution $P^{(0)}_{k,n}(x)$ due to plane "tails" (see remark at the end of Section 3.5).

Note that we cannot write analogous expressions (46) and (48) uniquely by taking simply $|n|$ and $|S(x)|$. It is due to the fact that functions $P_{m,n}$ in the upper half plane ($n > 0$) have no influence on functions $P_{m,n}$ in the lower half plane ($n < 0$) (except through the ground functions), and vice versa. For example, in the sum of near-boundary functions (the last term in (44) and (47)) coefficients $G^{(k,n)}_m$ should be equal to 0 if $n > 0$ and $S(x) \leq 0$ or if $n < 0$ and $S(x) \geq 0$.

$\tilde{P}_{k,n}(x)$ can be considered as first approximation to $P_{k,n}(x)$. Note that a priori there is no reason to neglect the second and the third terms in (44). Normally, we should take these terms into account, thus the relation (44) is considered as a system of linear equations on the near-boundary and ground functions. To find these functions, we remark that equations (44) must be satisfied for any k, n and x, and we can choose appropriate values of k and n. To close the system for near-boundary functions, we take $\{(k,n) : k \in [-M, M],\ n = S(k) + 1\}$, i.e. for any

$k \in [-M, M]$

$$P_{k,S(k)+1}(x) = \tilde{P}_{k,S(k)+1}(x) + \sum_{m \in J} H_{k-m}^{(S(k)+1)} P_{m,0}(x) - \sum_{m=-M}^{M} G_m^{(k,S(k)+1)} P_{m,S(m)+1}(x). \quad (49)$$

To close the system for ground functions, we can choose different conditions. For example, if we consider surfaces with $S(x) > 0$ on $x \in (-M, M)$, there are no ground functions, thus there is no additional condition other than (49). For the general case (where $J \neq \emptyset$), we propose the following condition

$$P_{k,0}(x) = \frac{1}{4}\Big(P_{k+1,0}(x) + P_{k-1,0}(x) + P_{k,1}(x) + P_{k,-1}(x)\Big), \qquad k \in J. \qquad (50)$$

We started from this relation for all the external points (see (3)). Here we just demand that this relation reminds valid if we substitute our approximations for $P_{k,1}(x)$ and $P_{k,-1}(x)$.

4. Some numerical verifications

In this section we briefly present some numerical results to check the validity of the approximation.

First of all, in Fig. 1 we depict the function $\varphi(\theta)$ which plays a central role in this work.

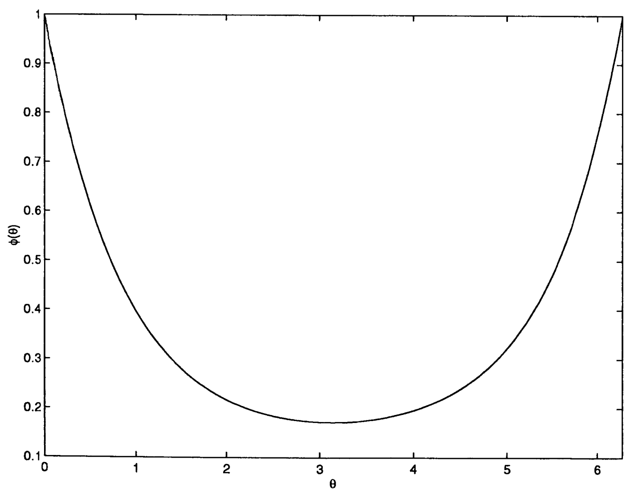


Figure 1. Function $\varphi(\theta)$. It changes in range from $\varphi(0) = 1$ to $\varphi(\pi) = 3 - \sqrt{8}$.

We can see that $\varphi(\theta) \ll 1$ if θ is not in vicinity of $2\pi m$ ($m \in Z$). This property was used in approximations (30) and (31). In order to understand the quality of the approximation (30) we depict the difference $f_n^{(N)}(\theta,\theta') - \varphi^n(\theta')$ for a given $N = 5$ and some n and θ (see Fig. 2).

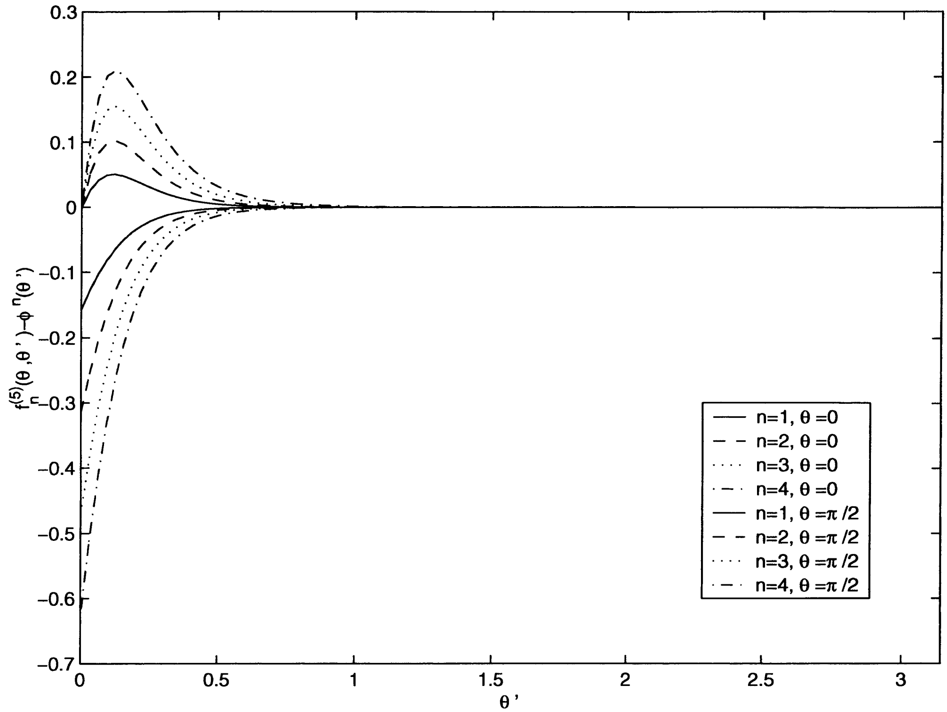


Figure 2. Difference $f_n^{(N)}(\theta,\theta') - \varphi^n(\theta')$ for $N = 5$. The four curves (for $n = 1,2,3,4$) in the upper half plane correspond to $\theta = 0$, others in the lower half plane – to $\theta = \pi/2$ (actually, the latest ones correspond to $\theta \geq \pi/8$ and almost do not depend on θ).

Here we can see that the approximation (30) is better for smaller n. It is not surprising because we have neglected the terms $\varphi^{2(N-n)}$. Also one can verify that for the same n the accuracy of (30) is higher for large N.

Let us consider again the planar surface, $S(x) = 0$. Without simulations, we easily obtain that

$$P_{k,n}(x) = \tilde{P}_{k,n}(x) = H^n_{k-x}.$$

So, in this trivial case our approximation gives the exact result (cf. (11)).

For a particular non-trivial surface the accuracy of the formula (44) can be obtained by comparing its values with numerical simulations of random walks. We have taken a simple surface represented in Fig. 3.

In this case there are no ground functions. On the contrary, there are 21 near-boundary functions which can be calculated with the help of (49). We present two

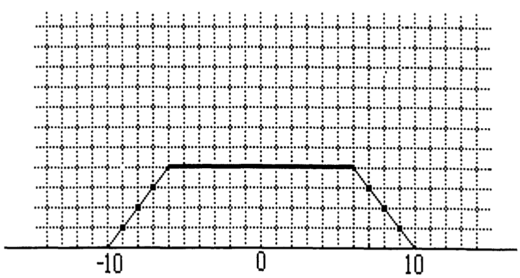


Figure 3. A simple surface with: $N = 5$, $N^* = 1$, $M = 10$.

distributions of hitting probabilities obtained numerically and through formula (44) (see Fig. 4).

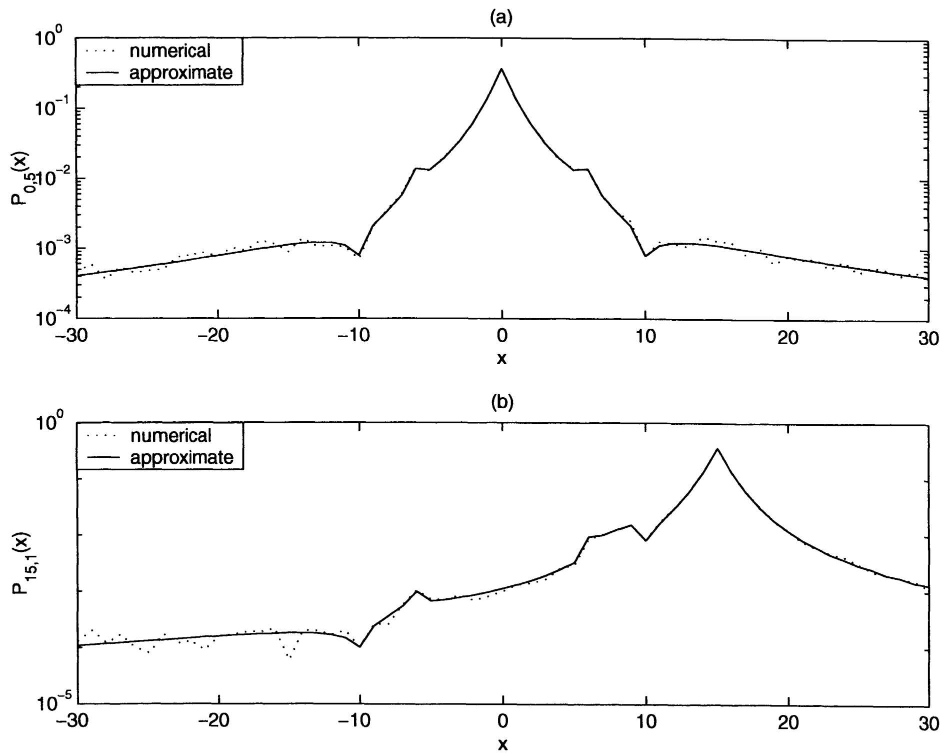


Figure 4. Distributions of hitting probabilities (in log-scale): the probability of the first contact with point $(x, S(x))$ of the surface if started from the point $(0, 5)$ (Fig. 4a) or $(15, 1)$ (Fig. 4b).

We can conclude that our approximation is quite good.

5. Conclusions and possible generalizations

Let us sum up what has been obtained. Using the same technique as for a planar surface, we expressed coefficients c_n in terms of c_0 and b_n. In these expressions (24), we approximated functions $f_n^{(N)}(\theta,\theta')$ and $\tilde{f}_n^{(N^*)}(\theta')$ by $\varphi^n(\theta')$. Such operation has several important features. First of all, it allows to separate the dependencies on θ and θ'. Thanks to this serapation, we finally obtain the *linear* equations (49) for near-boundary functions $P_{m,S(m)+1}(x)$ instead of *intergal* equations. Secondly, the approximation by $\varphi^n(\theta')$ unifies solutions for $n > 0$ and $n < 0$. Thirdly, it simplifies essentially our calculations. Note that expressions (30) and (31) are the only approximations in this treatment. If one leaves functions $f_n^{(N)}(\theta,\theta')$ and $\tilde{f}_n^{(N^*)}(\theta')$ in the calculation, one can obtain exact results. Unfortunately, such results have no practical sense: numerical solution of integral equations is usually more difficult than numerical simulation of random walks.

Having made the approximation, we express b_n in terms of near-boundary functions and combinations of exponential functions. Finally, we obtain a system of linear equations (49) for near-boundary functions $P_{m,S(m)+1}$ and ground functions $P_{m,0}$. It can be solved, and after that one can use the approximation (44) for any point (k,n).

Numerical analysis shows that this approximation works quite good.

The main conclusion is that we have found an approximate distribution of hitting probabilities for a rather general surface, pending certain conditions. In particular, one can make use of these results for a further study of the Laplacian transfer problems.

Now one needs to study the role of conditions which were imposed in the first section. As we said above, the compactness condition is the most important. It tells that

– the surface has a compact support, i.e. there is only *finite* "perturbation" of the planar surface;

– the plane "tails" have the same height (which is chosen as 0 of the vectical axis).

If we want to consider a surface with *infinite* support, we can obtain the same results but with an infinity of near-boundary functions. Thus the system (49) has infinitely many equations, and we cannot proceed any further. The same difficulty appears if the plane "tails" have different height: while we step down from the N-th level to the level of a lower "tail", we must pass through the level of a higher "tail". It means that there appears again an infinity of near-boundary functions. Only if we step down from the N-th level to the zeroth level (and from the $(-N^*)$-th level to the zeroth level), we can avoid the appearance of an infinity of near-boundary functions.

The regularity condition is used to simplify certain expressions. Nevertheless, it does not seem to be restrictive. Normally, to apply the technique of characteristic functions, we enumerate all sites (points) of the surface. To simplify the problem, one can make one of the following assumptions:

– either suppose that the surface obeys the regularity condition;

– or be interested in the total hitting probability $P^*_{k,n}(x)$ of points $(x, S(x))$, $(x, S(x)-1), \ldots, (x, S(x-1)+1)$ (if we authorize changes of $S(x)$ by more than one unit). In other words, we do not distinguish the points of the surface which have the same x-coordinate. Any of these assumptions allows to enumerate the points of the surface with their x coordinate using function $S(x)$. In the first assumption we consider the surfaces having only one point for each x; in the second assumption the surfaces can have some points with the same x, but we are interested in the total probability for each x. In order to generalize the method, one can introduce another *parametrization* of the surface. One possible generalization will be presented in our forthcoming paper.

6. Appendices

6.1. COEFFICIENTS H_k^n

As we have seen, coefficients H_k^n play a central role in all calculations of hitting probabilities in 2D. Here we briefly present some useful properties of H_k^n.

We can rewrite (11) in real form,

$$H_k^n = \int_0^\pi \frac{d\theta}{\pi} \cos(k\theta) \varphi^n(\theta). \tag{51}$$

First of all, let us write two inequalities for $\theta \in (0, \pi)$:

$$e^{-\theta} \le \varphi(\theta) \le e^{-\theta} \frac{1}{\cos(\theta/2)},$$

which can be used for estimations.

Now we are going to calculate the asymptotics of H_k^n for large k. Integrating the expression (51) by parts four times and using the values of the derivatives $\varphi^k(s)$ at the points 0 and π (see Table I), we obtain the asymptotic behaviour of H_k^n for large k (n is fixed),

$$H_k^n = \frac{n}{\pi k^2} - \frac{n(n^2 - 0.5)}{\pi k^4} + O(k^{-6}), \qquad k \gg 1. \tag{52}$$

TABLE I. The values of the derivatives $\varphi^{(k)}(s)$ at the points 0 and π.

θ	φ	φ'	φ''	φ'''
0	1	-1	1	$-1/2$
π	$3-\sqrt{8}$	0	$3\sqrt{2}/4 - 1$	0

This formula (52) works rather well for $k \ge 10$. At Table II we present values H_k^n for small k.

TABLE II. The values of coefficients H_k^n for small k (n in range from 1 to 5).

$n\backslash k$	0	1	2	3	4	5
1	0.3633	0.1366	0.0609	0.0319	0.0189	0.0124
2	0.1803	0.1221	0.0756	0.0477	0.0315	0.0219
3	0.1136	0.0958	0.0715	0.0517	0.0376	0.0278
4	0.0826	0.0759	0.0631	0.0501	0.0392	0.0307
5	0.0651	0.0620	0.0548	0.0464	0.0384	0.0315

The asymptotics of H_k^n for large n is

$$H_k^n = \frac{n}{\pi(n^2+k^2)} + O(n^{-3}),$$

i.e. we obtained the same behaviour as for the brownian motion. It is quite a reasonable result: if we look on the surface from a remoted point, there is no difference between continuous and discrete cases.

6.2. MANIPULATION WITH COEFFICIENTS α_n AND β_n

Here we present some properties of coefficients α_n and β_n. Also we prove the formula (26). Using only the definition (22), we find

$$\alpha_n = \alpha_k\alpha_{n-k+1} - \alpha_{k-1}\alpha_{n-k} \tag{53}$$

for any $k \le n$. Also using (23), we have

$$\beta_n = \alpha_{n+1} - \varphi\alpha_n.$$

Let us prove (26). According to (24), we have

$$c_n = \frac{\beta_{N-n}}{\beta_N}\left(c_0 + \sum_{k=1}^{n}\alpha_k b_k\right) + \frac{1}{\beta_N}\sum_{k=1}^{N-n-1} b_{n+k}(\beta_{N-n}\alpha_{n+k} - \beta_N\alpha_k).$$

Now we should simplify the difference in brackets in the last sum.

$$\beta_{N-n}\alpha_{n+k} - \beta_N\alpha_k = (\alpha_{N-n+1} - \varphi\alpha_{N-n})\alpha_{n+k} - (\alpha_{N+1} - \varphi\alpha_N)\alpha_k. \tag{54}$$

Consider the difference $\Delta = \alpha_{N-n+1}\alpha_{n+k} - \alpha_{N+1}\alpha_k$. Using the property (53), we can reduce the index $(n+k)$ in the first term and $(N+1)$ in the second term,

$$\Delta = \alpha_{N-n+1}(\alpha_k\alpha_{n+1} - \alpha_{k-1}\alpha_n) - (\alpha_{n+1}\alpha_{N-n+1} - \alpha_n\alpha_{N-n})\alpha_k =$$

$$= \alpha_n(\alpha_{N-n}\alpha_k - \alpha_{N-n+1}\alpha_{k-1}) = \alpha_n\alpha_{N-n-k+1}$$

(we used the property (53) in the last equality). Thus, we can represent (54) as

$$\beta_{N-n}\alpha_{n+k} - \beta_N\alpha_k = \alpha_n\alpha_{N-n-k+1} - \varphi\alpha_n\alpha_{N-n-k} = \alpha_n\beta_{N-n-k},$$

hence we find the formula (26),

$$c_n = f_n^{(N)} \left(c_0 + \sum_{k=1}^{n} \alpha_k b_k \right) + \alpha_n \sum_{k=n+1}^{N-1} f_k^{(N)} b_k.$$

References

1. Feller, W. (1971) *An Introduction to Probability Theory and Its Applications*, Second Edition.
2. Spitzer, F. (1964) *Principles of Random Walk.*
3. Sapoval, B. (1994) General Formulation of Laplacian Transfer Across Irregular Surfaces, *Phys. Rev. Lett.*, **73**, 3314.
4. Sapoval, B., Filoche, M., Karamanos, K. and Brizzi, R. (1999) Can one hear the shape of an electrode? I. Numerical study of the active zone in Laplacian tranfer, *Eur. Phys. J.* B, **9**, 739–753.
5. Filoche, M. and Sapoval B. (1999) Can one hear the shape of an electrode? II. Theoretical study of the Laplacian tranfer, *Eur. Phys. J.* B, **9**, 755–763.
6. Grebenkov D. S. (2000) Propriétés de l'opérateur d'auto-transport brownien, Rapport de stage à l'Ecole Polytechnique, France.
7. Grebenkov, D. S., Sapoval, B. and Filoche, M. (in press) Spectral Properties of the Brownian Self-Transport Operator.

Part THREE
Representation Theory

NOTES ON HOMOGENEOUS VECTOR BUNDLES OVER COMPLEX FLAG MANIFOLDS

SERGEI IGONIN * (igonin@mccme.ru)
University of Twente
The Netherlands

Abstract. Let P be a parabolic subgroup of a semisimple complex Lie group G defined by a subset $\Sigma \subset \Pi$ of simple roots of G, and let $\mathbf{E}_\varphi$ be a homogeneous vector bundle over the flag manifold $M = G/P$ corresponding to a linear representation φ of P. Using Bott's theorem, we obtain sufficient conditions on φ in terms of the combinatorial structure of $\Sigma \subset \Pi$ for cohomology groups $H^q(M, \mathcal{E}_\varphi)$ to be zero, where $\mathcal{E}_\varphi$ is the sheaf of holomorphic sections of $\mathbf{E}_\varphi$. In particular, we define two numbers $d(P)$, $\ell(P) \in \mathbb{N}$ such that for any φ obtained by natural operations from a representation $\tilde{\varphi}$ of dimension less than $d(P)$ one has $H^q(M, \mathcal{E}_\varphi) = 0$ for $0 < q < \ell(P)$. Applying this result to $H^1(M, \mathcal{E}_{\varphi\varphi^*})$, we see that the vector bundle $\mathbf{E}_\varphi$ is rigid.

Key words: Homogeneous vector bundle, complex flag manifold, cohomology of the sheaf of sections, Bott's theorem, root system, Dynkin diagram, rigid vector bundle.

Let $\mathbf{E}$ be a holomorphic vector bundle over a connected compact complex manifold M. Then there is a natural homomorphism of complex Lie groups $\mu\colon \operatorname{Aut}\mathbf{E} \to \operatorname{Bih} M$, where $\operatorname{Aut}\mathbf{E}$ is the automorphism group of the vector bundle and $\operatorname{Bih} M$ is the group of all biholomorphic transformations of M. The bundle $\mathbf{E}$ is said to be *homogeneous* if the action μ of $\operatorname{Aut}\mathbf{E}$ on M is transitive.

Assume that we have a homomorphism $\Phi\colon G \to \operatorname{Aut}\mathbf{E}$ such that the action $\mu\Phi$ of G on M is transitive. Then $\mathbf{E}$ is homogeneous; we say that $\mathbf{E}$ is *homogeneous with respect to* G. Let $o \in M$ and let $P = G_o$ be the stabilizer of o in G. Then P acts in a natural way on the fibre $E = E_o$, that is, we have a holomorphic linear representation $\varphi\colon P \to \mathrm{GL}(E)$. It is known that the bundle $\mathbf{E}$ is uniquely determined by the group G, the subgroup P, and the representation φ, which can all be arbitrary. We denote by $\mathbf{E}_\varphi$ the homogeneous vector bundle over $M = G/P$ defined by a representation φ of P.

Note also that corresponding to standard tensor operations on representations there are similar operations on vector bundles with a fixed base. In particular,

$$\mathbf{E}_{\varphi^*} = \mathbf{E}_\varphi^*, \quad \mathbf{E}_{\varphi_1+\varphi_2} = \mathbf{E}_{\varphi_1} \oplus \mathbf{E}_{\varphi_2}, \quad \mathbf{E}_{\varphi_1\varphi_2} = \mathbf{E}_{\varphi_1} \otimes \mathbf{E}_{\varphi_2}. \tag{1}$$

For a Lie group G denote by G° the identity component of G. We consider the case when M is a flag manifold, that is, when M is homogeneous and the stabilizers $(\operatorname{Bih} M)^\circ_x \subset (\operatorname{Bih} M)^\circ$, $x \in M$, are parabolic subgroups (see [1, 6]).

* Partially supported by the RFBR grants 01-01-00709, 00-01-00705.

V. A. Malyshev and A. M. Vershik (eds.),
Asymptotic Combinatorics with Application to Mathematical Physics, 245–254.

Then the group Bih M and all its transitive subgroups are semisimple Lie groups. Let $\mathbf{E} \to M$ be a homogeneous vector bundle and let $Q = \mu((\operatorname{Aut}\mathbf{E})^\circ)$. Then Q acts on M transitively and, therefore, is semisimple. By Levi's theorem, there exists a connected Lie subgroup G of Aut $\mathbf{E}$ such that $\mu\colon G \to Q$ is a local isomorphism. Thus $\mathbf{E}$ is homogeneous with respect to a connected semisimple Lie group G whose action on M is locally effective.

We choose a maximal torus T of G and denote by $\mathfrak{t}$ the corresponding Cartan subalgebra of the tangent algebra $\mathfrak{g}$ of G. Let $\Delta \subset \mathfrak{t}^*$ be the root system associated with T and let $\Delta_+ \subset \Delta$ be a subset of positive roots. We denote by Π the corresponding system of simple roots.

As usually, consider a non-degenerate G-invariant scalar product in $\mathfrak{g}$ inducing a non-degenerate scalar product $(\cdot,\cdot)$ in $\mathfrak{t}^*$ invariant under the Weyl group W. An element $\beta \in \mathfrak{t}^*$ is called a *weight* if

$$\frac{2(\beta,\alpha)}{(\alpha,\alpha)} \in \mathbb{Z} \quad \text{for all } \alpha \in \Delta. \tag{2}$$

A weight β is said to be *dominant* if $(\beta,\alpha) \geq 0$ for each $\alpha \in \Delta^+$.

Let $\mathfrak{g}_\alpha$ be the root subspace of $\mathfrak{g}$ corresponding to $\alpha \in \Delta$. If R is a Lie subgroup of G normalized by T, then its tangent subalgebra $\mathfrak{r} \subset \mathfrak{g}$ has the following form:

$$\mathfrak{r} = (\mathfrak{r} \cap \mathfrak{t}) \oplus \bigoplus_{\alpha \in \Delta(R)} \mathfrak{g}_\alpha,$$

where $\Delta(R) \subset \Delta$ is some subset called the *root system of* R. In particular, each system of positive roots Δ_+ determines two Borel (that is, maximal solvable) subgroups $B_\pm$ of G containing T with root systems $\Delta(B_\pm) = \pm\Delta_+$.

Let P be a parabolic subgroup of G, that is, a subgroup containing a Borel subgroup. We assume that P contains the subgroup B_- corresponding to the system of negative roots $\Delta_- = -\Delta_+$. It is known (see [1, 6]) that $\Delta(P) = \Delta_- \cup [\Sigma]$ in this case, where $[\Sigma]$ is the set of all roots of G that can be expressed as a linear combination of elements of a subset $\Sigma \subset \Pi$. We have the semidirect decomposition

$$P = H \ltimes N_-, \quad \mathfrak{p} = \mathfrak{h} \oplus \mathfrak{n}_-,$$

where H is a maximal reductive subgroup, N_- is the unipotent radical of P, and $\mathfrak{p}$, $\mathfrak{h}$, $\mathfrak{n}_-$ are the corresponding Lie algebras. Moreover,

$$\Delta(H) = [\Sigma],$$
$$\Delta(N_-) = \Delta_- \setminus [\Sigma],$$

and Σ coincides with the system of simple roots of the group H corresponding to its Borel subgroup $B_+ \cap H$. Thus the parabolic subgroup $P \subset G$ is uniquely determined by the subset $\Sigma \subset \Pi$ of simple roots.

Note that T is also a maximal torus of P. Each weight of a representation φ of P is a weight in the above sense. A representation φ is completely reducible if and only if it is trivial on N_-; in this case it is uniquely determined by the representation $\varphi|_H$ of H. A ***highest weight*** of φ is, by definition, a highest weight of $\varphi|_H$ understood in the sense of the ordering corresponding to the Borel subgroup $B_+ \cap H$.

One uses Bott's well-known theorem (see [1, 2]) to calculate the graded cohomology space $H^*(M, \mathcal{E}_\varphi)$, where $M = G/P$ is a flag manifold and $\mathcal{E}_\varphi$ is the sheaf of holomorphic sections of the homogeneous vector bundle $\mathbf{E}_\varphi \to M$ defined by a representation $\varphi\colon P \to \mathrm{GL}(E)$.

A weight $\lambda \in \mathfrak{t}^*$ is said to be ***singular*** if there exists a root $\alpha \in \Delta_+$ such that $(\lambda, \alpha) = 0$, and it is said to be ***regular*** otherwise. We set $\gamma = \frac{1}{2}\sum_{\alpha\in\Delta_+}\alpha$; then

$$(\gamma, \alpha) = \frac{1}{2}(\alpha, \alpha) \quad \text{for each } \alpha \in \Pi. \tag{3}$$

If λ is a regular weight, then there exists a unique $w \in W$ such that $w(\lambda) - \gamma$ is dominant. The *index* of the regular weight λ is the smallest integer s such that w expands into the product of s reflections s_α, $\alpha \in \Pi$. This is equal to the number of $\alpha \in \Delta_+$ such that $(\lambda, \alpha) < 0$. To each weight $\lambda \in \mathfrak{t}^*$ such that $\lambda + \gamma$ is regular we assign a dominant weight $I(\lambda) \in \mathfrak{t}^*$ by the rule $I(\lambda) = w(\lambda + \gamma) - \gamma$.

To formulate Bott's theorem, note that in each cohomology space $H^q(M, \mathcal{E})$ associated with a homogeneous vector bundle $\mathbf{E} \to M = G/P$ there exists a natural structure of a G-module.

Bott's theorem. *Let φ be an irreducible finite-dimensional representation of P with highest weight Λ. Then the graded cohomology space $H^*(M, \mathcal{E}_\varphi)$ is determined by Λ in the following way:*

- *if $\Lambda + \gamma$ is singular then $H^*(M, \mathcal{E}_\varphi) = 0$;*
- *if $\Lambda + \gamma$ is regular and its index is p, then $H^q(M, \mathcal{E}_\varphi) = 0$ for all $q \neq p$, while $H^p(M, \mathcal{E}_\varphi)$ is an irreducible G-module with highest weight $I(\Lambda)$.*

Using this theorem, we will obtain some sufficient conditions on representations φ of P for a cohomology group $H^q(M, \mathcal{E}_\varphi)$ to be zero.

Bott's theorem is not applicable directly when the representation is not completely reducible. But due to the following lemma to prove that cohomology is zero it is enough to consider completely reducible representations. For an arbitrary holomorphic representation $\varphi : P \to \mathrm{GL}(E)$ we construct a completely reducible representation φ^s as follows. There is a filtration (*Jordan–Hölder tower*)

$$0 = E_0 \subset E_1 \subset \ldots \subset E_m = E$$

of E by invariant subspaces of $\varphi(P)$ such that for all $i = 1, \ldots, m$ the induced representation $\bar{\varphi}_i$ in E_i/E_{i-1} is irreducible. We set

$$\varphi^s = \bar{\varphi}_1 + \cdots + \bar{\varphi}_m. \tag{4}$$

It is well-known that φ^s does not depend on a Jordan–Hölder tower of E.

Lemma 1. *If $H^q(M, \mathcal{E}_{\varphi^s}) = 0$ for some $q \geq 0$ then $H^q(M, \mathcal{E}_\varphi) = 0$.*

Proof. From $H^q(M, \mathcal{E}_{\varphi^s}) = 0$ and (4) we have

$$H^q(M, \mathcal{E}_{\bar{\varphi}_i}) = 0 \quad \text{for all } i = 1, \dots, m. \tag{5}$$

Denote by φ_i the restriction of φ onto the invariant subspace E_i and consider the short exact sequences of sheaves

$$0 \to \mathcal{E}_{\varphi_{i-1}} \to \mathcal{E}_{\varphi_i} \to \mathcal{E}_{\bar{\varphi}_i} \to 0, \ \ 1 \leq i \leq m,$$

corresponding to the exact sequences

$$0 \to E_{i-1} \to E_i \to E_i/E_{i-1} \to 0$$

of P-modules. Using the exact cohomology sequence, by induction on i from (5) we get $H^q(M, \mathcal{E}_{\varphi_i}) = 0$ for all $i = 1, \dots, m$. □

Each $\xi \in \mathfrak{t}^*$ is a linear combination of simple roots $\alpha \in \Pi$. We denote by $c_\alpha(\xi)$ the corresponding coefficients, that is, $\xi = \sum_{\alpha \in \Pi} c_\alpha(\xi) \cdot \alpha$, and set

$$\mathrm{C}(\xi) = \{\alpha \in \Pi : c_\alpha(\xi) \neq 0\} \subset \Pi.$$

The value $\sqrt{(\xi, \xi)}$ is called the *length* of ξ and denoted by $|\xi|$. For δ^1, $\delta^2 \in \Delta_+$ we write $\delta^1 \leq \delta^2$ if $c_\alpha(\delta^1) \leq c_\alpha(\delta^2)$ for all $\alpha \in \Pi$. We need two lemmas on semisimple Lie algebras.

Lemma 2. *If δ^1, $\delta^2 \in \Delta_+$ and $\delta^1 \leq \delta^2$ then there exists a sequence of positive roots $\delta_0, \delta_1, \dots, \delta_m$ such that $\delta_0 = \delta^1$, $\delta_m = \delta^2$, and $\delta_i - \delta_{i-1} \in \mathrm{C}(\delta^2 - \delta^1)$ for all $i = 1, \dots, m$.*

Proof. By induction on $|\mathrm{C}(\delta^2 - \delta^1)|$, we must prove that if $\delta^1 \neq \delta^2$ then there is $\alpha \in \mathrm{C}(\delta^2 - \delta^1)$ such that

$$\delta^2 - \alpha \in \Delta_+ \quad \text{or} \quad \delta^1 + \alpha \in \Delta_+. \tag{6}$$

We have $(\delta^2 - \delta^1, \delta^2 - \delta^1) > 0$, hence $(\delta^2 - \delta^1, \alpha) > 0$ for some $\alpha \in \mathrm{C}(\delta^2 - \delta^1)$. Then $(\delta^2, \alpha) > 0$ or $(\delta^1, \alpha) < 0$, which implies (6). □

Lemma 3. *Let $\mathfrak{g}_1, \dots, \mathfrak{g}_n$ be simple Lie algebras. Denote by d_i the minimal dimension of a nontrivial representation of $\mathfrak{g}_i$. Let ψ be a representation of the semisimple algebra $\oplus_i \mathfrak{g}_i$. If $\dim \psi < \sum_i d_i$ then ψ is trivial on some $\mathfrak{g}_i$.*

Proof. Each irreducible component of ψ is of the form

$$\psi_{i_1} \otimes \psi_{i_2} \otimes \cdots \otimes \psi_{i_k}, \tag{7}$$

where ψ_{i_j} is a nontrivial irreducible representation of $\mathfrak{g}_{i_j}$. The dimension of (7) is not less than

$$d_{i_1} d_{i_2} \dots d_{i_k} \geq d_{i_1} + d_{i_2} + \cdots + d_{i_k}.$$

Therefore, if ψ is nontrivial on each $\mathfrak{g}_i$ then $\dim \psi \geq \sum_i d_i$. □

Let $\mathrm{A} \subset \Pi \setminus \Sigma$ and $\mathrm{B} \subset \Sigma$. In the sequel we consider representations φ of P with highest weights Λ satisfying

$$(\Lambda, \alpha) < 0, \ (\Lambda, \beta) = 0 \quad \text{for all } \ \alpha \in \mathrm{A}, \ \beta \in \mathrm{B}. \tag{8}$$

We define a class of positive roots δ of $\mathfrak{g}$ (*significant* (A, B)-*roots*) such that if $\Lambda + \gamma$ is regular then $(\Lambda + \gamma, \delta) < 0$. By Bott's theorem and Lemma 1, this implies that $H^q(M, \mathcal{E}_\varphi) = 0$ for $0 \le q < \ell(\mathrm{A}, \mathrm{B})$, where $\ell(\mathrm{A}, \mathrm{B})$ is the number of significant (A, B)-roots.

Let us introduce the required concepts. A positive root

$$\delta = \sum_{\alpha \in \mathrm{A}} c_\alpha(\delta) \cdot \alpha + \sum_{\beta \in \mathrm{B}} c_\beta(\delta) \cdot \beta, \ \ c_\alpha(\delta), \, c_\beta(\delta) \ge 0,$$

of $\mathfrak{g}$ with

$$c_{\alpha_0}(\delta) > 0 \quad \text{for some } \ \alpha_0 \in \mathrm{A} \tag{9}$$

will be called an (A, B)-*root*. Denote by $\Delta_{\mathrm{A,B}}$ the set of (A, B)-roots. We say that $\delta \in \Delta_{\mathrm{A,B}}$ is *significant* if there exists a root $\sigma_\delta \in \Delta_{\mathrm{A,B}}$ such that

$$\sigma_\delta \le \delta, \quad \sum_{\alpha \in \mathrm{A}} c_\alpha(\sigma_\delta) \cdot (\alpha, \alpha) \ge \sum_{\beta \in \mathrm{B}} c_\beta(\sigma_\delta) \cdot (\beta, \beta), \tag{10}$$

and $\mathrm{C}(\delta - \sigma_\delta)$ consists of simple roots of the same length not greater than $|\sigma_\delta|$.

Lemma 4. *If a weight* Λ *satisfies* (8) *and* $\Lambda + \gamma$ *is regular then for each significant* $\delta \in \Delta_{\mathrm{A,B}}$ *one has* $(\Lambda + \gamma, \delta) < 0$.

Proof. By (8) and (3), we have

$$(\Lambda + \gamma, \sigma_\delta) = \sum_{\alpha \in \mathrm{A}} \left(c_\alpha(\sigma_\delta) \cdot (\Lambda, \alpha) + \frac{1}{2} c_\alpha(\sigma_\delta) \cdot (\alpha, \alpha) \right) + \sum_{\beta \in \mathrm{B}} \frac{1}{2} c_\beta(\sigma_\delta) \cdot (\beta, \beta). \tag{11}$$

Note that

$$\frac{2(\Lambda, \alpha)}{(\alpha, \alpha)} \ne -1 \quad \text{for each } \ \alpha \in \Pi \tag{12}$$

because otherwise $(\Lambda + \gamma)(\alpha) = 0$ and $\Lambda + \gamma$ is singular. Combining (2), (8), and (12), we obtain

$$\frac{2(\Lambda, \alpha)}{(\alpha, \alpha)} \le -2 \quad \text{for each } \ \alpha \in \mathrm{A}. \tag{13}$$

From (11), (13), and (10) one gets

$$(\Lambda + \gamma, \sigma_\delta) \le \sum_{\alpha \in \mathrm{A}} -\frac{1}{2} c_\alpha(\sigma_\delta) \cdot (\alpha, \alpha) + \sum_{\beta \in \mathrm{B}} \frac{1}{2} c_\beta(\sigma_\delta) \cdot (\beta, \beta) \le 0. \tag{14}$$

By Lemma 2, there exist $\delta_0, \, \delta_1, \ldots, \, \delta_m \in \Delta_{\mathrm{A,B}}$ such that $\delta_0 = \sigma_\delta$, $\delta_m = \delta$, and

$$\delta_i - \delta_{i-1} \in \mathrm{C}(\delta - \sigma_\delta) \subset \mathrm{A} \cup \mathrm{B}. \tag{15}$$

Denote

$$c = \frac{1}{2}(\delta_i - \delta_{i-1}, \delta_i - \delta_{i-1}).$$

By the definition of significant roots, c does not depend on $i = 1, \dots, m$ and

$$2c \leq (\sigma_\delta, \sigma_\delta). \tag{16}$$

For any roots α, β if $|\alpha| \geq |\beta|$ then $\frac{(\alpha,\alpha)}{(\beta,\beta)} \in \mathbb{N}$. Applying this to $\delta_0 = \sigma_\delta$ and $\delta_i - \delta_{i-1}$, from (16) we get $\frac{1}{2c}(\delta_0, \delta_0) \in \mathbb{N}$ and, therefore, from (2) one obtains

$$\frac{1}{c}(\Lambda + \gamma, \delta_0) = \frac{2(\Lambda + \gamma, \delta_0)}{(\delta_0, \delta_0)} \cdot \frac{(\delta_0, \delta_0)}{2c} \in \mathbb{Z}. \tag{17}$$

By (2) and (3),

$$\frac{1}{c}(\Lambda + \gamma, \delta_i - \delta_{i-1}) = \frac{1}{c}(\Lambda, \delta_i - \delta_{i-1}) + 1 \in \mathbb{Z}. \tag{18}$$

Moreover, from (15) and (8) we have $\frac{1}{c}(\Lambda, \delta_i - \delta_{i-1}) \leq 0$, hence

$$\frac{1}{c}(\Lambda + \gamma, \delta_i - \delta_{i-1}) \leq 1. \tag{19}$$

By induction on i, combining (14), (17), (18), and (19), we get $\frac{1}{c}(\Lambda + \gamma, \delta_i) \in \mathbb{Z}$ and

$$(\Lambda + \gamma, \delta_i) < 0 \quad \text{for all } i = 0, 1, \dots, m \tag{20}$$

because otherwise $(\Lambda + \gamma, \delta_i) = 0$ for some positive root δ_i and $\Lambda + \gamma$ is singular. In particular, $(\Lambda + \gamma, \delta) < 0$. □

Example 5. We prove three sufficient conditions for $\delta \in \Delta_{\mathrm{A,B}}$ to be significant by giving corresponding σ_δ.

1. *If* $\mathrm{C}(\delta)$ *contains simple roots of the same length*, we can, by (9), take $\sigma_\delta \in \mathrm{C}(\delta) \cap \mathrm{A}$.

2. *If there is* $\alpha_0 \in \mathrm{C}(\delta) \cap \mathrm{A}$ *longer than the other roots from* $\mathrm{C}(\delta) \setminus \{\alpha_0\}$, then $\mathrm{C}(\delta) \setminus \{\alpha_0\}$ contains roots of the same length, $c_{\alpha_0}(\delta) = 1$, hence $\mathrm{C}(\delta - \alpha_0) = \mathrm{C}(\delta) \setminus \{\alpha_0\}$ and one can put $\sigma_\delta = \alpha_0$.

3. *If there exists* $\alpha_0 \in \mathrm{C}(\delta) \cap \mathrm{A}$ *shorter than the other roots from* $\mathrm{C}(\delta) \setminus \{\alpha_0\}$ *and* $c_{\alpha_0}(\delta) = -\frac{2(\beta,\alpha_0)}{(\alpha_0,\alpha_0)}$ *for some* $\beta \in \mathrm{C}(\delta) \setminus \{\alpha_0\}$, then $\mathrm{C}(\delta) \setminus \{\alpha_0\}$ consists of roots of length $|\beta| = |\beta - \frac{2(\beta,\alpha_0)}{(\alpha_0,\alpha_0)} \cdot \alpha_0|$ and we set

$$\sigma_\delta = \beta - \frac{2(\beta, \alpha_0)}{(\alpha_0, \alpha_0)} \cdot \alpha_0.$$

Let $\ell(\mathrm{A}, \mathrm{B})$ be the number of significant (A, B)-roots. This number can be computed in terms of some graphs, if we regard A and B as subgraphs of the Dynkin diagram D of $\mathfrak{g}$.

Example 6. Denote by U the union of the connected components C of $\mathrm{A} \cup \mathrm{B}$ satisfying $\mathrm{C} \cap \mathrm{A} = \emptyset$ and set $\mathrm{B}' = \mathrm{B} \setminus \mathrm{U}$. Clearly, (A, B)-roots coincide with $(\mathrm{A}, \mathrm{B}')$-roots and $\ell(\mathrm{A}, \mathrm{B}) = \ell(\mathrm{A}, \mathrm{B}')$ because for each (A, B)-root δ the subgraph $\mathrm{C}(\delta) \subset \mathrm{A} \cup \mathrm{B}$ is connected and, by (9), $\mathrm{C}(\delta) \cap \mathrm{A} \neq \emptyset$.

Suppose that all the edges of $\mathrm{A} \cup \mathrm{B}'$ are simple, that is, each connected component of $\mathrm{A} \cup \mathrm{B}'$ contains simple roots of the same length. Then, by Example 5, each $(\mathrm{A}, \mathrm{B}')$-root is significant, hence $\ell(\mathrm{A}, \mathrm{B})$ equals the number of $(\mathrm{A}, \mathrm{B}')$-roots. The latter is equal to the number of positive roots of $\mathfrak{g}_{\mathrm{A} \cup \mathrm{B}'}$ minus the number of positive roots of $\mathfrak{g}_{\mathrm{B}'}$, where $\mathfrak{g}_{\mathrm{A} \cup \mathrm{B}'}$ and $\mathfrak{g}_{\mathrm{B}'}$ are semisimple Lie algebras with Dynkin diagrams $\mathrm{A} \cup \mathrm{B}'$ and B' respectively.

In what follows, we will denote by the same letter a representation $\varphi \colon P \to \mathrm{GL}(E)$ and its differential $\varphi \colon \mathfrak{p} \to \mathfrak{gl}(E)$.

Theorem 7. *Consider an irreducible representation φ of P with highest weight $\Lambda \in \mathfrak{t}^*$. Let* A *and* B *be subsets of* $\Pi \setminus \Sigma$ *and* Σ *respectively such that* (8) *holds. Then we have* $H^q(M, \mathcal{E}_\varphi) = 0$ *for* $0 \leq q < \ell(\mathrm{A}, \mathrm{B})$.

Proof. If $\Lambda + \gamma$ is singular then, by Bott's theorem, $H^*(M, \mathcal{E}_\varphi) = 0$; suppose that $\Lambda + \gamma$ is regular. By Lemma 4, $(\Lambda + \gamma, \delta) < 0$ for each significant $\delta \in \Delta_{\mathrm{A},\mathrm{B}}$. Therefore, the index of $\Lambda + \gamma$ is not less than $\ell(\mathrm{A}, \mathrm{B})$. Applying Bott's theorem, one completes the proof. □

We say that a representation $\varphi \colon P \to \mathrm{GL}(E)$ ***is obtained by natural operations from a representation*** $\tilde{\varphi} \colon P \to \mathrm{GL}(\tilde{E})$ if there is a homomorphism $\pi \colon \mathrm{GL}(\tilde{E}) \to \mathrm{GL}(E)$ such that

$$\varphi = \pi \circ \tilde{\varphi}. \tag{21}$$

Example 8. A representation φ obtained from $\tilde{\varphi}$ by tensor operations (1), for instance $\varphi = \tilde{\varphi}\tilde{\varphi}^*$, satisfies (21).

Denote by H' and $\mathfrak{h}'$ the semisimple commutator subgroup of H and its tangent subalgebra respectively. Let us regard Σ as a subdiagram of the Dynkin diagram D of $\mathfrak{g}$. Then $\mathfrak{h}'$ is a semisimple Lie algebra with Dynkin diagram Σ. As usually, each connected component C of Σ determines a simple ideal $\mathfrak{h}_{\mathrm{C}}$ of $\mathfrak{h}'$. If a representation $\tilde{\varphi}$ of P is trivial on some $\mathfrak{h}_{\mathrm{C}}$ then, by (21), any φ obtained from $\tilde{\varphi}$ by natural operations is also trivial on $\mathfrak{h}_{\mathrm{C}}$. This observation, Lemmas 1, 3, and Theorem 7 allow to prove that if φ is obtained by natural operations from a representation of relatively small dimension then some cohomology groups $H^q(M, \mathcal{E}_\varphi)$ are zero.

Indeed, Lemma 3 guarantees that $\tilde{\varphi}$ and, therefore, φ are trivial on some ideals of $\mathfrak{h}'$ if the dimension of $\tilde{\varphi}$ is sufficiently small. The less is the dimension of $\tilde{\varphi}$ the more ideals φ is trivial on. By the construction (4) of the corresponding completely reducible representation φ^s, $\varphi(\mathfrak{h}_{\mathrm{C}}) = 0$ implies $\varphi^s(\mathfrak{h}_{\mathrm{C}}) = 0$. Let B be the union of those connected components C of Σ for which one has $\varphi^s(\mathfrak{h}_{\mathrm{C}}) = 0$.

Consider an arbitrary irreducible component φ' of φ^s with highest weight Λ. Clearly, $(\Lambda, \beta) = 0$ for each $\beta \in \mathrm{B}$. We set

$$\mathrm{A}(\varphi') = \{\alpha \in \Pi \setminus \Sigma : (\Lambda, \alpha) < 0\}.$$

If $\mathrm{A}(\varphi')$ is empty then Λ is dominant and, by Bott's theorem, $H^q(M, \mathcal{E}_{\varphi'}) = 0$ for all $q > 0$. Otherwise, according to Theorem 7, we get $H^q(M, \mathcal{E}_{\varphi'}) = 0$ for $0 \leq q < \ell(\mathrm{A}(\varphi'), \mathrm{B})$. Thus, by Lemma 1,

$$H^q(M, \mathcal{E}_{\varphi^s}) = H^q(M, \mathcal{E}_{\varphi}) = 0 \quad \text{for all } 0 < q < \min_{\varphi'} \ell(\mathrm{A}(\varphi'), \mathrm{B}), \tag{22}$$

where φ' runs through the irreducible components of φ^s with $\mathrm{A}(\varphi') \neq \emptyset$.

The following theorem is an example of such a result. To formulate it we need some notations. We say that a subset $\mathrm{B} \subset \Sigma$ is *adjacent* to a vertex $\alpha \in \Pi \setminus \Sigma$ of D if there is an edge of D connecting α and B. Equivalently, there is a unique simple root $\beta \in \mathrm{B}$ such that $(\alpha, \beta) \neq 0$. For a simple root $\alpha \in \Pi \setminus \Sigma$ consider the connected components $\mathrm{C}_1, \dots, \mathrm{C}_n$ of $\Sigma \subset \mathrm{D}$ adjacent to α and the corresponding simple ideals $\mathfrak{h}_{\mathrm{C}_1}, \dots, \mathfrak{h}_{\mathrm{C}_n}$ of $\mathfrak{h}'$. Denote by d_i the minimal dimension of a nontrivial representation of $\mathfrak{h}_{\mathrm{C}_i}$ and set

$$\begin{aligned} d(\alpha) &= d_1 + \cdots + d_n, \\ \ell(\alpha) &= \min_{i=1,\dots,n} \ell(\{\alpha\}, \mathrm{C}_i). \end{aligned}$$

If there are no nonempty connected components adjacent to α, we put $d(\alpha) = \ell(\alpha) = 1$. Set also

$$\begin{aligned} d(P) &= \min_{\alpha \in \Pi \setminus \Sigma} d(\alpha), \\ \ell(P) &= \min_{\alpha \in \Pi \setminus \Sigma} \ell(\alpha). \end{aligned}$$

Theorem 9. *Suppose that a representation φ of P is obtained by natural operations from a representation $\tilde{\varphi}$ of dimension less then $d(P)$. Then we have*

$$H^q(M, \mathcal{E}_{\varphi}) = 0 \quad \textit{for } 0 < q < \ell(P). \tag{23}$$

Proof. If $d(P) = 1$, the statement is trivial. If $d(P) > 1$ then from Lemma 3 it follows that for each $\alpha \in \Pi \setminus \Sigma$ there is a nonempty connected component C_α of Σ adjacent to α such that $\tilde{\varphi}$ and, by (21), φ are trivial on $\mathfrak{h}_{\mathrm{C}_\alpha}$. By the definition of $\ell(P)$, we have $\ell(P) \leq \ell(\{\alpha\}, \mathrm{C}_\alpha)$ and from (22) one obtains (23). $\square$

Remark 10. We have $d(P) > 1$ if and only if for each $\alpha \in \Pi \setminus \Sigma$ there exists a nonempty connected component $\mathrm{C} \subset \Sigma$ adjacent to α. According to the following lemma, this is also equivalent to $\ell(P) > 1$. In this case the result of Theorem 9 is nontrivial.

Lemma 11. *If* B *is a nonempty subset of Σ adjacent to $\alpha \in \Pi \setminus \Sigma$ then*

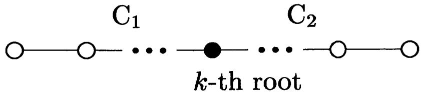
$$\ell(\{\alpha\}, \mathrm{B}) \geq 2.$$

Proof. The root α is clearly a significant (α, B)-root. By assumption, there is $\beta \in \mathrm{B}$ such that $(\alpha, \beta) \neq 0$. Consider the root δ of the form

$$\delta = a \cdot \alpha + b \cdot \beta, \quad a,\, b > 0,$$

such that $\delta + \alpha$ and $\delta + \beta$ are not roots of $\mathfrak{g}$. According to Example 5, δ is also significant. Thus we have at least two distinct significant roots. $\square$

Example 12. Consider the Grassmanian $\mathrm{Gr}_{n,k}$ of k-dimensional subspaces of $\mathbb{C}^n$. It is a flag manifold of the group $\mathrm{SL}_n(\mathbb{C})$. The stabilizer of a point $o \in \mathrm{Gr}_{n,k}$ is a parabolic subgroup with $\Sigma = \Pi \setminus \{\alpha\}$, where α is the k-th simple root in the Dynkin diagram



of $\mathrm{SL}_n(\mathbb{C})$. The subdiagram Σ consists of two components C_1 and C_2 adjacent to α, one of which is empty if $\mathrm{Gr}_{n,k} \simeq \mathbb{CP}^{n-1}$, that is, if $k = 1$ or $k = n-1$. Hence $d(P) = n$ if $1 < k < n-1$, and $d(P) = n-1$ if $k = 1$ or $k = n-1$. According to Example 6, we have $\ell(\{\alpha\}, \mathrm{C}_1) = k$, $\ell(\{\alpha\}, \mathrm{C}_2) = n-k$. Thus $\ell(P) = \min\{k, n-k\}$ if $1 < k < n-1$, and $\ell(P) = n-1$ if $k = 1$ or $k = n-1$. This computation gives the following result.

Theorem 13. *For any vector bundle* $\mathbf{E} \to \mathrm{Gr}_{n,k}$, $1 < k < n-1$, *obtained by tensor operations from a homogeneous vector bundle* $\tilde{\mathbf{E}} \to \mathrm{Gr}_{n,k}$ *of rank less than* n *one has* $H^q(\mathrm{Gr}_{n,k}, \mathcal{E}) = 0$ *for all* $0 < q < \min\{k, n-k\}$.

Proof. It is known (see [6]) that for $1 < k < n-1$ any transitive Lie subgroup of $\mathrm{Bih}\,\mathrm{Gr}_{n,k}$ coincides with $\mathrm{Bih}\,\mathrm{Gr}_{n,k} = \mathrm{PSL}_n(\mathbb{C})$. Therefore, each homogeneous vector bundle over $\mathrm{Gr}_{n,k}$ is homogeneous with respect to the simply connected group $\mathrm{SL}_n(\mathbb{C})$. Hence $\tilde{\mathbf{E}} = \mathbf{E}_{\tilde{\varphi}}$ for some representation $\tilde{\varphi}$ of $P \subset \mathrm{SL}_n(\mathbb{C})$. By (1), we have $\mathbf{E} = \mathbf{E}_\varphi$, where φ is obtained from $\tilde{\varphi}$ by tensor operations. According to Example 8, Theorem 9, and the above computation of $d(P)$ and $\ell(P)$, we complete the proof. $\square$

Cohomology groups of sheaves considered here appear, in particular, in deformation theory [8]. Theorems 7 and 9 allow to prove *rigidity*, that is, absence of nontrivial *deformations* [8], of some homogeneous vector bundles and *supermanifolds* [5, 7]. A result on rigidity of vector bundles is given by the following theorem, while homogeneous supermanifolds are studied in [4].

Theorem 14. *For any representation* φ *of* P *obtained by natural operations from a representaion* $\tilde{\varphi}$ *of dimension less than* $d(P)$ *the bundle* $\mathbf{E}_\varphi$ *is rigid.*

Proof. By (1), we have $\mathbf{E}_\varphi \otimes \mathbf{E}_\varphi^* = \mathbf{E}_{\varphi\varphi^*}$. According to Example 8, the representaion $\varphi\varphi^*$ is also obtained from $\tilde{\varphi}$ by natural operations. Then from Theorem 9 and Remark 10 one gets $H^1(M, \mathcal{E}_\varphi \otimes \mathcal{E}_\varphi^*) = 0$, which implies that $\mathbf{E}_\varphi$ is rigid (see [3, 8]). $\square$

Acknowledgements

I would like to express deep gratitude to Prof. A. L. Onishchik for his attention to this work and many useful remarks. It is a pleasure to thank the organizers of the

European Summer School 2001 on Asymptotic Combinatorics with Application to Mathematical Physics in St. Petersburg for the invitation and providing a stimulating environment. I am also grateful to the Erwin Schrödinger Institute in Vienna for hospitality in October–November 2000 where a part of this research was done.

References

1. Akhiezer, D. N. (1995) *Lie group actions in complex analysis*, Vieweg & Sohn, Braunschweig.
2. Bott, R. (1957) Homogeneous vector bundles, *Ann. of Math.* **66**, no. 2, 203–248.
3. Forster, O. and Knorr, K. (1974) Über die Deformationen von Vektorraumbündeln auf kompakten komplexen Räumen, *Math. Ann.* **209**, no. 4, 291–346.
4. Igonin, S. (in preparation) Homogeneous supermanifolds associated with Hermitian symmetric spaces.
5. Manin, Yu. I. *Gauge field theory and complex geometry*, Springer–Verlag, Berlin, 1988.
6. Onishchik, A. L. (1994) *Topology of transitive transformation groups*, Barth Verlag, Leipzig.
7. Onishchik, A. L. (1999) On the classification of complex analytic supermanifolds, *Lobachevskii J. Math.* **4** (1999), 47–70 (electronic).
8. Palamodov, V. P. (1990) Deformations of complex spaces, *Encycl. Math. Sci.* **10** (1990), 105–194.

REPRESENTATIONS THEORY AND DOUBLES OF YANGIANS OF CLASSICAL LIE SUPERALGEBRAS

VLADIMIR STUKOPIN (stukopin@math.rsu.ru)
Department of Mathematics
Don State Technical University
Rostov-on-Don,
Russia

Abstract. Some basic results of the theory of Yangians of Lie superalgebras are described. The Yangian of a classical Lie superalgebra is described as a result of quantization of the Lie superbialgebra of polynomial loops. Two systems of generators and defining relations are introduced, and their equivalence is proved. The PBW theorem for the Yangians of classical Lie superalgebras is formulated and proved. All irreducible finite-dimensional representations of the Yangians of Lie superalgebras of type $A(m,n)$ are described in terms of Drinfel'd polynomials. A notion of the double of a Yangian and a formula for the universal R-matrix for the double of a Yangian are discussed for the Yangian of a Lie superalgebra of type $A(m,n)$.

Introduction

The main goal of this article is to describe some basic results of the theory of Yangians of Lie superalgebras of classical type keeping in mind applications to integrable models of quantum field theory and statistical mechanics as a final aim. The notion of the Yangian of a simple Lie algebra g was introduced by V. Drinfel'd ([1]) as a deformation of the universal enveloping algebra $U(g[t])$ of the current algebra $g[t]$, and it was extended to the case of Lie superalgebras of classical type. The main tool of such applications in mathematical physics (and physics) is a quantum R-matrix. I recall the definition of an R-matrix ([1], [4], [6]). An R-matrix is a function $R(u)$ of a complex parameter u with values in $End(V \otimes V)$, where V is a finite-dimensional vector space (superspace) which satisfies the quantum Yang–Baxter equation (QYBE) (or the graded quantum Yang–Baxter equation (gQYBE) in the case of superspace):

$$R_{12}(u-v)R_{13}(u-w)R_{23}(v-w) = R_{23}(v-w)R_{13}(u-w)R_{12}(u-v),$$

where $R_{12} = R \otimes id \in End(V \otimes V \otimes V)$, etc.

Construction of an R-matrix is closely related to the representation theory of quantum groups (supergroups). In particular, every irreducible finite-dimensional representation V of the Yangian is associated with a solution $R_V(u)$ of QYBE

V. A. Malyshev and A. M. Vershik (eds.),
Asymptotic Combinatorics with Application to Mathematical Physics, 255–265.

which is a rational function of u. With the Yangian $Y(g)$ (where g is a simple Lie algebra) Drinfel'd associates an element R in some formal completion of $Y(g) \otimes Y(g)$ which is called the universal R-matrix (see [1]). More precisely, R is a formal series of a complex parameter u with coefficients in $Y(g) \otimes Y(g)$. The universal R-matrix is defined by the following condition: it intertwines the comultiplication and the opposite comultiplication: $\Delta'(a) = R\Delta(u)R^{-1}$. A quantum R-matrix R_V is the image of the universal R-matrix R under the action of a representation $\rho_V \otimes \rho_V : Y(g) \otimes Y(g) \to End(V \otimes V)$. The problem of description (or computation) of the universal R-matrix is very important since all formulas for quantum R-matrices are corollaries of the formula for the universal R-matrix. In many cases it is reasonable to work with the quantum double $DY(g)$ of the Yangian $Y(g)$ if we keep in mind applications to quantum field theory. A definition of the double for Hopf algebras was given by V. Drinfel'd, explicit formulas for the double of a Yangian were obtained by F. Smirnov ([11]) and S. Khoroshkin, V. Tolstoy ([10]). The notion of $DY(G)$ for a Lie superalgebra G is discussed below.

Yangians of Lie algebras are studied in many papers (see [4]). The theory of Yangians for Lie superalgebras (or super Yangians) is less developed (see [7], [8]). In this paper we make an attempt to develop the theory of Yangians for Lie superalgebras of classical type, and to generalize some results from [7].

This paper is organized as follows. In section 1 we recall basic definitions of the Lie superalgebras theory, and define a bisuperalgebra structure on superalgebras of polynomial loops. In section 2 we introduce the main object $Y(G)$, the Yangian of classical superalgebras, as a quantization of $U(G[t])$. We describe $Y(G)$ in terms of generators and defining relations. Starting from a system of zero-order generators, with relations and comultiplication law similar to those of the universal enveloping superalgebra, we introduce first-order generators with a more complicated comultiplication law. The remaining relations follow from the compatibility conditions for the algebra and coalgebra structures. From the finite system of generators and relations thus obtained we derive a new system of generators and relations (section 3) which can be used in the statement of the PBW theorem (see section 4). In section 5 we describe all finite-dimensional representations of $Y(A(m,n))$. In section 6 we discuss the structure of the double and the multiplicative formula of the universal R-matrix for the double of the super Yangian $Y(A(m,n))$. The proofs are either omitted, or given only as sketches.

1. Lie Superalgebras and Superbialgebras

Let us recall some basic definitions from the Lie superalgebras theory ([9]). A superalgebra $G = G_0 \oplus G_1$ is a Z_2-graded algebra with a product $[\cdot,\cdot]$, i.e., if $a \in G_\alpha$, $b \in G_\beta$, $\alpha, \beta \in Z_2 = \{0,1\}$, then $[a,b] \in G_{\alpha+\beta}$. A Lie superalgebra G is a superalgebra satisfying the following axioms:

$$[a,b] = -(-1)^{\alpha\beta}[b,a]; [a,[b,c]] = [[a,b],c] + (-1)^{\alpha\beta}[b,[a,c]], a \in G_\alpha, b \in G_\beta.$$

Let $A = (a_{ij})$ be a matrix of order r, and let τ be a subset of the set $I = \{1,...,r\}$ (as a rule, $\tau = \{k\}$, $1 \leq k \geq r$). Let us assume that the Cartan matrix A is a symmetrizable matrix, $B = DA$ is a symmetric matrix, and

$D = diag(d_1, ..., d_r)$, $d_k \neq 0$, is a diagonal matrix. Let $G = G(A, \tau)$ be a Lie superalgebra with generators x_i^+, x_i^-, h_i, $i \in I$, and the following defining relations:

$$[x_i^+, x_i^-] = \delta_{ij} h_i, \quad [h_i, h_j] = 0, \tag{1}$$

$$[h_i, x_i^\pm] = \pm b_{i,j} x_j^\pm, \quad (ad^{n_{ij}} x_i^\pm)(x_j^\pm) = 0, \quad i \neq j, \tag{2}$$

where $n_{i,j} = 1$, if $b_{i,i} = b_{i,j} = 0$; $n_{i,j} = 2$, if $b_{i,i} = 0$, $b_{i,j} \neq 0$; $n_{i,j} = 1 - 2b_{i,j}/b_{i,i}$, if $b_{i,i} \neq 0$; $ad\,a(b) = [a, b]$. Let $\deg(x) = \alpha$ if $x \in G_\alpha$, $\alpha \in Z_2$. Recall that a finite-dimensional Lie superalgebra $G = G_0 \oplus G_1$ is of classical type, if it is simple, and the representation of G_0 in G_1 is completely reducible.

Let $\Delta = \Delta_0 \cup \Delta_1$ be the root system of G, Δ_i be the root system of G_i, Δ^+, Δ_i^+ be the sets of positive roots of G, G_i respectively, $\Pi = \{\alpha_1, ..., \alpha_r\}$ be the system of simple roots. Recall that a simple root α_i is called white, if $i \in I \setminus \tau$, and grey or black, if $i \in \tau$, and $a_{ii} = 0$ or $a_{ii} = 2$ respectively.

Let G be a Lie superalgebra with symmetrizable Cartan matrix and nondegenerate Killing form. A superbialgebra G is a vector superspace with algebra and coalgebra structures which are compatible in the following sense: the cocommutator map $\varphi : G \to G \otimes G$ is a 1-cocycle (we assume that G acts on $G \otimes G$ by the adjoint representation). Let the cocycle φ be the coboundary of an element $r \in \bigwedge^2 g$. Consider the current superalgebra $G[t]$, and introduce the cocommutator φ by the formula

$$\varphi : a(u) \to [a(u) \otimes 1 + 1 \otimes a(v), r(u, v)],$$

where $r = \frac{t}{u-v}$, and t is a Casimir operator. The cocommutator φ defines the superbialgebra structure on $G[t]$. Notice that r satisfies the Yang–Baxter equation (or triangle equation):

$$yb(r) = [r^{12}, r^{13}] + [r^{12}, r^{23}] + [r^{13}, r^{23}] = 0,$$

$$r^{12} = \sum r_i' \otimes r"_i \otimes 1, r^{13} = \sum r_i' \otimes 1 \otimes r''i, r^{23} = \sum 1 \otimes r_i' \otimes r_i'' \text{ for } r = \sum r_i' \otimes r_i''.$$

2. Quantization of a Lie superbialgebra $G[t]$

The definition of quantization for Lie bialgebras from [1] can be extended to Lie superbialgebras. A quantization of a Lie superbialgebra $A = G[t]$ is a Hopf superalgebra $A_h = G[t]_h$ over the ring of formal power series $C[[h]]$ such that the following conditions hold:
1) $A_h/hA_h \cong U(G[t])$, where $U(G[t])$ is the universal enveloping superalgebra of the Lie superalgebra $G[t]$;
2) the Lie superalgebra A_h is isomorphic to $U(G[t])[[h]]$ as a vector space;
3) the correspondence principle is satisfied:

$$h^{-1}(\Delta(x) \ - \ \Delta^{op}(x)) mod\, h = \varphi(x) mod\, h,$$

where Δ is the comultiplication, Δ^{op} is the opposite comultiplication (i.e., if $\Delta(x) = \sum x_i' \otimes x"_i$, then $\Delta^{op}(x) = \sum (-1)^{p(x_i')p(x"_i)} x"_i \otimes x_i')$, the cobracket $\varphi : G[t] \to G[t] \bigwedge G[t]$ is a cocycle in $G[t]$.

It follows from 1), 2) that we can consider deformations of generators $h_i, x_i^{\pm}$, $h_i \cdot u$, $x_i^{\pm} \cdot u$, $i \in I$, as generators of A_h. Denote these deformations by $h_{i,0}$, $x_{i,0}^{\pm}$, $h_{i,1}$, $x_{i,1}^{\pm}$, $i \in I$. It follows from 1), 3) and the formula for the cocycle that we must introduce relations and comultiplication laws for $h_{i,0}, x_{i,0}^{\pm}$ as for $h_i, x_i^{\pm}$. Define a comultiplication law for $h_{i,1}$ by the formula

$$\Delta(h_{i,1}) = h_{i,i} \otimes 1 + 1 \otimes h_{i,i} + [h_{i,0} \otimes 1, \Omega], \quad i \in I, \tag{3}$$

where $\Omega = \sum_{i=1}^{r} (-1)^{\deg(x_i^-)} x_i^- \otimes x_i^+$ is a part of the Casimir operator.

It follows form 2), 4) that the comultiplication laws for $x_{i,1}^{\pm}$ are

$$\Delta(x_{i,1}^+) = x_{i,i}^+ \otimes 1 + (-1)^{\deg(x_i^-)} 1 \otimes x_{i,i}^+ + h[x_{i,0}^+ \otimes 1, \Omega], \quad i \in I;$$

$$\Delta(x_{i,1}^-) = x_{i,i}^- \otimes 1 + (-1)^{\deg(x_i^-)} 1 \otimes x_{i,i}^- + h[1 \otimes x_{i,0}^-, \Omega], \quad i \in I.$$

Other defining relations for A_h follow from the compatibility condition of algebra and coalgebra structures. Fixing $h = 1$, we obtain the Yangian of the Lie superalgebra $G : Y(G) = A_1$.

Theorem 1. *The Yangian $Y(G)$ is a Hopf superalgebra over C with generators $h_{i,0}, x_{i,0}^{\pm}, h_{i,1}, x_{i,1}^{\pm}, i \in I$, and defining relations*

$$[h_{i,0}, h_{j,0}] = [h_{i,0}, h_{j,1}] = [h_{i,1}, h_{j,1} = 0]; \tag{4}$$
$$[h_{i,0}, x_{j,0}^{\pm}] = \pm b_{ij} x_{j,0}^{\pm}, \quad [h_{i,1}, x_{j,0}^{\pm}] = \pm b_{ij} x_{j,1}^{\pm}; \tag{5}$$
$$[x_{i,0}^+, x_{j,0}^-] = \delta_{ij} h_{i,0}, \quad [x_{i,1}^+, x_{j,0}^-] = \delta_{ij}(h_{i,1} + \tfrac{1}{2} h_{i,0}^2); \tag{6}$$
$$[x_{i,1}^{\pm}, x_{j,0}^{\pm}] = [x_{i,0}^{\pm}, x_{j,1}^{\pm}] \pm (b_{ij}/2)(x_{i,0}^{\pm} x_{j,0}^{\pm} + x_{j,0}^{\pm} x_{i,0}^{\pm}); \tag{7}$$
$$(ad\, x_{i,0}^{\pm})^{n_{ij}}(x_{j,0}^{\pm}) = 0, \quad i \neq j; \tag{8}$$
$$[[h_{i,1}, x_{i,1}^+], x_{i,1}^-] + [x_{i,1}^+, [h_{i,1}, x_{i,1}^-]] = 0, \quad b_{ii} \neq 0; \tag{9}$$
$$[[h_{i,1}, x_{j,1}^+], x_{j,1}^-] + [x_{j,1}^+, [h_{i,1}, x_{j,1}^-]] = 0, \quad b_{ii} = 0, b_{ij} \neq 0. \tag{10}$$

The comultiplication law Δ is defined by the formulas

$$\Delta(h_{i,0}) = h_{i,0} \otimes 1 + 1 \otimes h_{i,0}; \tag{11}$$
$$\Delta(x_{i,0}^{\pm}) = x_{1,0}^{\pm} \otimes 1 + (-1)^{\deg(x_i^-)} 1 \otimes x_{i,0}^{\pm}; \tag{12}$$
$$\Delta(h_{i,1}) = h_{i,1} \otimes 1 + 1 \otimes h_{i,1} + h[h_{i,0} \otimes 1, \Omega_2]; \tag{13}$$
$$\Delta(x_{i,1}^+) = x_{i,1}^+ \otimes 1 + (-1)^{\deg(x_i^+)} 1 \otimes x_{i,1}^+ + h[x_{i,0}^+ \otimes 1, \Omega_2]; \tag{14}$$
$$\Delta(x_{i,1}^-) = x_{i,1}^- \otimes 1 + (-1)^{\deg(x_i^-)} 1 \otimes x_{i,1}^- - h[1 \otimes x_{i,0}^-, \Omega_2]. \tag{15}$$

3. A new system of generators

We shall assume for simplicity that $G = A(m,n)$. Let us introduce generators $\tilde{x}^{\pm}_{i,k}$, $\tilde{h}_{i,k}$ $(\in Y(G))$, $i \in I$, $k \in Z_+$, by the formulas

$$\begin{aligned} \tilde{x}^{\pm}_{i,k+1} &= \pm(a^{-1}_{i,i})[h_{i,1}, \tilde{x}^{\pm}_{i,k}], \qquad i \in I \setminus \tau (i \neq m+1); \\ \tilde{x}^{\pm}_{m+1,k+1} &= \mp[h_{m,1}, \tilde{x}^{\pm}_{m+1,k}]; \end{aligned} \tag{16}$$

$$\tilde{h}_{i,k} = [\tilde{x}^{+}_{i,k}, x^{-}_{i,0}], \qquad \tilde{x}^{\pm}_{i,0} = x^{\pm}_{i,0}, \qquad \tilde{h}_{i,0} = h_{i,0}. \tag{17}$$

Definition 2. Denote by $\bar{Y}(G)$ the superalgebra over C with generators $x^{\pm}_{i,k}$, $h_{i,k}$, $i \in \Gamma = I$, $k \in Z_+$, and defining relations

$$[h_{i,k}, h_{j,l}] = 0, \quad \delta_{i,j} h_{i,k+l} = [x^{+}_{i,k}, x^{-}_{j,l}]; \tag{18}$$

$$[h_{i,k+1}, x^{\pm}_{j,l}] = [h_{i,k}, x^{\pm}_{j,l+1}] + (b_{ij}/2)(h_{i,k} x^{\pm}_{j,l} + x^{\pm}_{j,l} h_{i,k}); \tag{19}$$

$$[h_{i,0}, x^{\pm}_{j,l}] = \pm b_{ij} x^{\pm}_{j,l}; \tag{20}$$

$$[x^{\pm}_{i,k+1}, x^{\pm}_{j,l}] = [x^{\pm}_{i,k}, x^{\pm}_{j,l+1}] + (b_{ij}/2)(x^{\pm}_{i,k} x^{\pm}_{j,l} + x^{\pm}_{j,l} x^{\pm}_{i,k}); \tag{21}$$

$$\textstyle\sum_{\sigma} [x^{\pm}_{i,k_{\sigma(1)}}, \ldots [x^{\pm}_{i,k_{\sigma(r)}}, x^{\pm}_{j,l}] \ldots] = 0, \quad i \neq j, \qquad r = n_{ij} \tag{22}$$

where the sum is taken over all permutations σ of $\{1, \ldots, r\}$.

Notice that the parity function takes the following values on the generators: $p(x^{\pm}_{j,k}) = 0$ for $k \in Z_+$, $j \in I \setminus \tau$; $p(h_{i,k}) = 0$ for $i \in I$, $k \in Z_+$; $p(x^{\pm}_{i,k}) = 1$ for $k \in Z_+$, $i \in \tau$.

Theorem 3. *The correspondence*

$$\tilde{x}^{\pm}_{i,k} (\in Y(G) \to x^{\pm}_{i,k} (\in \bar{Y}(G),$$
$$\tilde{h}_{i,k} (\in Y(G) \to h_{i,k} (\in \bar{Y}(G)$$

defines an isomorphism

$$Y(G) \to \bar{Y}(G).$$

First of all we shall give a sketch of the proof. We can assume that $\bar{Y}(G)$ is generated by generators $\tilde{x}^{\pm}_{i,k}$, $\tilde{h}_{i,k}$ which satisfy relations (4)–(10), (16), (17). Therefore, it suffices to prove that relations (18)–(22) follow from (4)–(10), (16), (17), and relations (4)–(10), (16), (17) follow from relations (18)–(22). The latter fact is almost evident.

Indeed, relations (4)–(10) are contained among relations (18)–(22). Relation (16) follows from (19) and the definition of $\tilde{h}_{i,1}$ (the second relation of (6)). Relation (17) follows from the second relation of (18). The following relations in terms of generating functions and logarithmic generators $\bar{h}_{i,k}$ play an important role in the proof.

Let

$$h_i(t) = \sum_{k \geq -1} h_{i,k} t^{-k-1}, \quad x^{\pm}_i(t) = \sum_{k \geq 0} x^{\pm}_{i,k} t^{-k-1}, h_{i,-1} = 1. \tag{23}$$

Define generators $\bar{h}_{i,k}$ by the formula

$$\bar{h}_i(t) = \sum_{k \geq 0} \bar{h}_{i,k} t^{-k-1} = \ln(h_i(t)). \tag{24}$$

Let A be a unital algebra with generators h_j, x_j, $j \in Z_+$, and the following defining relations:

$$[h_k, h_l] = 0, \tag{25}$$

$$[h_k, x_l] = [h_{k-1}, x_{l+1}] + \gamma(h_{k-1} x_l + x_l h_{k-1}), \tag{26}$$

where $h_{-1} = 1$, $\gamma \in R$. Let $h(t) = \sum_{k \geq -1} h_k t^{-k-1}$, $x(t) = \sum_{k \geq 0} x_k t^{-k-1}$, and define $\bar{h} \in A$ $(k \in Z_+)$ by the formula

$$\bar{h}(t) = \sum_{k \geq 0} \bar{h}_k t^{-k-1} = \ln(h(t)). \tag{27}$$

The following lemma can be proved by induction (see [12]).

Lemma 4. *Let* (25) *be satisfied for* $j \leq p$, $l \leq p$, *and let* (26) *be satisfied for* $k \leq p$, $l \in Z_+$. *Then the following relations hold for* $k \leq p$, $l \in Z_+$*:*

$$[\bar{h}_k, x_l] = 2\gamma \cdot x_{k+l} + 2 \cdot \sum_{\substack{0 \leq s \leq k-2 \\ k+s \in 2Z_+}} \frac{\gamma^{k+l-s}}{k+1} C_{k+1}^{s} x_{l+s}. \tag{28}$$

Corollary 5. *The following relations take place in* $Y(G)$*:*

$$[\bar{h}_{i,k}, x_{j,l}^{\pm}] = \pm(\alpha_i, \alpha_j) \cdot x_{j,k+l}^{\pm} \pm \sum_{\substack{0 \leq s \leq k-2 \\ 0 \leq s \in 2Z_+}} \frac{2^{s-k} \cdot (\alpha_i, \alpha_j)^{k+l-s}}{k+1} C_{k+1}^{s} x_{l+s}^{\pm}. \tag{29}$$

Let us sketch the proof of Theorem 3. The equivalence of defining relations for even generators is similar to the case of Yangians of simple Lie algebras (see [13]), and can be proved by induction. The proof of equivalence of defining relations in the case of odd generators is based on that for even generators, and is carried out in the same way. First of all we deduce (16)–(22) from (4)–(10), (16), (17) for small values of the second indices, preparing the induction base. Then we deduce relations (16)–(22) for $i = j$ (for even and odd generators), and finally we deduce (16)–(22) for $i \neq j$ for even and odd indices).

4. PBW theorem

We shall use the second realization of the super Yangian (see Definition 1 from [5]). We refer to the second index of the generators $x_{ik}^{\pm}, h_{rs}$ as the degree of the corresponding generator, and define the degree of a monomial in these generators as the sum of degrees of its factors. The degree of a polynomial in the generators $x_{ik}^{\pm}, h_{rs}$ is defined as the maximal degree of the monomials it is comprised of. Define the degree of a tensor product of monomials in $x_{ik}^{\pm}, h_{rs}$ as the sum of degrees of the tensor factors. The degree of a polynomial is defined as above. It follows from the formulas for the coproduct that

$$\deg(\Delta(x_{ik}^{\pm}) - x_{ik}^{\pm} \otimes 1 - 1 \otimes x_{ik}^{\pm}) < k;$$

$$\deg(\Delta(h_{ik}) - h_{ik} \otimes 1 - 1 \otimes h_{ik}) < k.$$

Denote by $Y(G)_k$ the space of elements of $Y(G)$ of degree at most k. We obtain a filtration $Y(G) = \sum_{k\geq 0} Y(G)_k$. We start constructing the PBW basis with constructing the root vectors for the super Yangian. Let $\alpha = \alpha_{i_1} + \alpha_{i_2} + \ldots + \alpha_{i_s}$ be the decomposition of a positive root into a sum of simple roots such that

$$x_{\alpha}^{\pm} = [x_{1_1}^{\pm}, [x_{i_2}^{\pm} \ldots [x_{i_{s-1}}^{\pm}, x_{i_s}^{\pm}] \ldots]]$$

is a non-zero root vector in the $\pm\alpha$ root space of G. Let $k = k_1 + \ldots + k_s$ be the decomposition of k in a sum of non-negative integers, and put

$$x_{\alpha,k}^{\pm} = [x_{i_1,k_1}^{\pm}, [x_{i_2,k_2}^{\pm} \ldots [x_{i_{s-1},k_{s-1}}^{\pm}, x_{i_s,k_s}^{\pm}] \ldots]],$$

$$x_{\pm\alpha,k} = x_{\alpha,k}^{\pm}, h_{\alpha,k} = [x_{\alpha,0}^{+}, x_{\alpha,k}^{-}].$$

Notice that if $k = k_1' + \ldots + k_s'$ is another decomposition of $k > 0$, and $x_{\alpha,k}'^{\pm}$, $h_{\alpha,k}'$ are the corresponding elements of $Y(G)$, then $x_{\alpha,k}^{\pm} - x_{\alpha,k}'^{\pm} \in Y(G)_{k-1}$, $h_{\alpha,k} - h_{\alpha,k}' \in Y(G)_{k-1}$. Choose any total ordering $\prec$ on the set $\{x_{\alpha,k}^{+}, x_{\beta,j}^{-}, h_{j,k} = h_{\alpha_j,k}$, and denote by $(B, \prec)$ the set of ordered monomials in $\{x_{\alpha,k}^{+}, x_{\beta,j}^{-}, h_{j,k}\}$.

Theorem 6. *$(B, \prec)$ is the PBW basis in $Y(G)$.*

Let us outline the proof of the theorem. First of all let us prove the completeness of $(B, \prec)$. Given a monomial $M \in (B, \prec)$, define its length in generators $x_{\alpha,k}^{+}, x_{\beta,j}^{-}, h_{j,k}$ as the number of factors of M. Notice that if we rearrange the factors of M, we shall obtain additional summands either of a smaller degree, or of the same degree but of a smaller length. By induction on the degree d of a polynomial in the above generators, and for a fixed d, by induction on the maximal length of its term, we obtain that $Y(G)$ is spanned by $(B, \prec)$. Now we have to prove that monomials from $(B, \prec)$ are linearly independent. Let $(B, \prec)_k = \{M \in B : \deg(M) \leq k\}$, and notice that the subspace $(B, \prec)_k$ depends only on k but not on $\prec$. This means that linear independence may be proved for only one total ordering. We choose this total ordering as follows. Let Δ_+ be the set of positive roots of G, and let $\beta(1) < \beta(2) < \cdots < \beta(n)$. We require that $x_{\alpha,k}^{+} \prec h_{j,k} \prec x_{\beta,j}^{-}$ for all $\alpha, \beta, j, k, l, m$; if $i < j$, then $x_{\beta(i),k}^{\pm} < x_{\beta(j),k}^{\pm}$, and $h_{i,k} < h_{j,l}$ for all k, l; if $k < l$, then $x_{\beta(i),k}^{\pm} < x_{\beta(i),l}^{\pm}$, and $h_{i,k} < h_{i,l}$ for $1 \leq i \leq n$, $1 \leq j \leq n$. Assume that the monomials are not linearly independent. Then there exist $M_1, \cdots, M_k \in B$, and numbers $c_1, \cdots, c_k \in C*$ such that

$$\sum_{j=1}^{k} c_j M_j = 0. \tag{30}$$

To show that the last equality leads to contradiction, we use an idea due S. Levendorskii ([12]). Namely, we consider the set of automorphisms $\tau_a, a \in C$, of the Hopf superalgebra $Y(G)$. In terms of the generators x^+_{ik} and h_{ik} the automorphism τ_a can be expressed as follows ([4, 12]):

$$\tau(h_{ik}) = \sum_{0 \le s \le k} C^k_s a^{k-s} h_{is}, \tau(x^{\pm}_{ik}) = \sum_{0 \le s \le k} C^k_s a^{k-s} x^{\pm}_{is}. \tag{31}$$

We also use a representation $\rho_0 : Y(G) \to End(V)$ satisfying the following property: $\rho_0(x^+_{\alpha,0})$, $\rho_0(x^-_{\beta,0})$, $\rho_0(h_{j0})$, $\alpha, \beta \in \Delta_+$, $1 \le j \le n$, are linearly independent. Such representation exists (see, for example, [9], [14]). The remaining part of the proof is similar to the proof for the Yangians of simple Lie algebras from [12] (and can be reduced to it).

5. Representations and Drinfel'd polynomials.

Finite-dimensional representations of simple Lie algebras were described by V. Drinfel'd in [3]. The Yangians of Lie superalgebras of type $A(m,n)$ were described in [7]. (We shall use basic definitions from [7]). Recall basic definitions of the representation theory of Yangians. Let $G = A(m,n)$, and let V be a finite-dimensional superspace. A representation of the Yangian $Y(G)$ of the Lie superalgebra G in V is a homomorphism of $Y(G)$ in $l(V)$. A representation V of $Y(G)$ is said to be a highest weight representation if it is generated by a vector v_0 such that

$$x^+_{ik} v_0 = 0,$$

$$h_{ik} v_0 = d_{ik} v_0$$

for all $i \in \{1, 2, \ldots, m+n+1\}$, $k \in Z_+, d_{ik} \in C$. $d = \{d_{ik}\}$ is called a highest weight. We shall denote by $V(d)$ the highest weight representation with highest weight $d = \{d_{ik}\}$.

Theorem 7. 1) *Every irreducible finite-dimensional representation of $Y(G)$ is a highest weight representation.*
2) *A representation $V(d)$ is finite-dimensional if and only if there exist polynomials P^d_i, $i \in \{1, 2, \ldots, m+n+1\}$, satisfying the following conditions:*
a) The leading coefficient of the polynomial P^d_i, $i \in \{1, 2, \ldots, m, m+2, \ldots, m+n+1\}$, equals to 1.
b)

$$\frac{P^d_i(u+1)}{P^d_i(u)} = 1 + \sum_{k=0}^{\infty} d_{ik} u^{-k-1},$$

in the sense that the right-hand side is the Laurent expansion of the left-hand side at $u = \infty$.
c) The correspondence $V(d) \to \{P^d_i(u)\}_{i \in I}$ defines a bijective map from the set of irreducible finite-dimensional representations of $Y(G)$ onto the set of polynomials $\{P^d_i(u)\}$ such that the leading coefficient of the polynomial $P^d_i(u)$ is equal to 1 for $i \in I \setminus \{m+1\}$.

A sketch of the proof. Let V be a finite-dimensional representation of the Yangian $Y(G)$, $G = A(m,n)$. The Hopf superalgebra G equals $G = G_0 \oplus G_1$, where $G_0 = sl(m+1) \oplus sl(n+1) \oplus C$, where C is a centre. Consider the restriction of the module structure on the algebra $Y(A_m) \otimes Y(A_n) \otimes C$. Notice that the algebra $Y(A_m) \otimes Y(A_n) \otimes C$ contains the algebra $U(A_m) \otimes U(A_n) \otimes C[x]$ (where $C[x]$ is an algebra of complex-valued polynomials) as a subalgebra. □

Lemma 8. *Let M be a finite-dimensional irreducible $Y(G)$-module. Consider the restriction of the representation M to $Y_1 = Y(A_m) \oplus Y(gl_n)$. Then the representation M is isomorphic to the tensor product of the representations M_1 and M_2 of the algebras $Y(A_m)$ and $Y(gl_n)$ respectively.*

Lemma 9. *Let V, U be irreducible finite-dimensional representations of $Y(G)$ with sets of Drinfel'd polynomials $\{P_i\}$, $\{\tilde{P}_i\}$. Let $v \in V$, $u \in U$ be their highest weight vectors. Then $v \otimes u$ is a highest weight vector in $V \otimes U$, and its set of Drinfel'd polynomials is $\{P_i\tilde{P}_i\}$.*

Using this lemmas we reduce Theorem 7 to a similar result for the Yangians of the Lie algebras A_m and gl_n which were formulated in [3] (see also [6]).

6. The double of a Yangian, and the universal R-matrix

Recall the definition of the double of a Hopf algebra (superalgebra). Let A^0 be the dual algebra A^* with the opposite comultiplication. Then there exists a unique quasitriangular Hopf algebra (superalgebra) $D(A)$ with universal R-matrix R such that the following conditions hold: 1) $D(A)$ contains A and A^0 as Hopf subalgebras (subsuperalgebras); 2) R is the image of the canonical element $A \otimes A^0$ under the inclusion in $D(A) \otimes D(A)$; 3) the linear map $A \otimes A \hookrightarrow D(A)$, $a \otimes b \mapsto ab$, is a bijection. Let us define the double of a super Yangian. Let $C(G)$ be the superalgebra generated by the elements $x^{\pm}_{ik}, h_{ik}$, $i \in I$, $k \in Z$, with relations (18)–(22).

As above, we refer to the second index of a generator as its degree. As in 5^0 above we obtain that the superalgebra $C(G)$ admits a $Z-$filtration:

$$\ldots \subset C_n \subset \ldots. \subset C_{-1} \subset C_0 \subset C_1 \subset \ldots \subset C(G).$$

Let $\bar{C}(G)$ be the corresponding formal completion of $C(G)$.

Theorem 10. *The superalgebra $\bar{C}(G)$ is isomorphic to $DY(G)$.*

To describe $DY(G)$, it is more convenient to use generating functions

$$x_i^{\pm +}(u) = \sum_{k \geq 0} x^{\pm}_{ik} u^{-k-1}, x_i^{\pm -}(u) = -\sum_{k<0} x^{\pm}_{ik} u^{-k-1},$$
$$h_i^{+}(u) = 1 + \sum_{k \geq 0} h_{ik} u^{-k-1}, h_i^{+}(u) = 1 - \sum_{k<0} h_{ik} u^{-k-1}.$$

In terms of these generating functions we can get comultiplication formulas and defining relations (using the Gauss decomposition and the dual realization of the super Yangian). At present, these formulas are proved only partially.

We can define the PBW basis for $DY(G)$ in a similar way as for $Y(G)$. Let $Y^{\pm}_{\mp}, Y^{\pm}_{0}$ be the subsuperalgebras generated by the fields $x^{\pm\mp}(u)$, $h^{\pm}(u)$ respectively.

Theorem 11. *Let $G = A(0,1)$. Then the universal R-matrix R of the super Yangian double $DY(G)$ can be presented in the factorizable form*

$$R = R_+ R_0 R_-, \tag{32}$$

where

$$R_+ \in Y^+_+ \otimes Y^-_-, R_- \in Y^+_- \otimes Y^-_+, R_0 \in Y^+_0 \otimes Y^-_0,$$

$$R_\pm = \sum_{\alpha\in\Delta_+} \sum_{k\geq 0} \exp(-x^{\pm}_{\alpha,k} \otimes x^{\mp}_{-\alpha,-k-1}). \tag{33}$$

The middle term R_0 has a more complicated structure. The computation of this term is not completed at present, and its explicit form is a conjecture.

Conjecture. *Let $\varphi^{\pm}_i(u) = \ln h^{\pm}_i(u)$, $A = (a_{ij})^n_{i,j=1}$ be a symmetrizable matrix for $A(m,n)$ $(m \neq n)$, $a_{ij}(q) = [a_{ij}]_q = \frac{q^{a_{ij}} - q^{-a_{ij}}}{q-q^{-1}}$, $A(q) = (a_{ij}(q))^n_{i,j=1}$. Let $C(q) = (c_{ij}(q))^n_{i,j=1}$ be a matrix proportional to $A(q)^{-1}$, and T be the shift operator, $Tf(v) = f(v-1)$. Then*

$$R_0 = \prod_{m\geq 0} exp \sum_{1\leq i,j\leq n} \mathrm{Res}_{u=v}(\varphi^+_i(u))' \otimes c_{ji}(T^{-1/2})\varphi^-_j(v + (n+1/2)h), \tag{34}$$

where h is a constant, and the operator $T^{-1/2}$ is substituted inside $c_{ji}(q)$ instead of q.

Acknowledgements

The author is grateful to the organizers of the NATO ASI–2001 in Saint-Petersburg where this work was discussed. This work was also supported by RFBR grant No. 00-01-00868.

References

1. Drinfeld, V. (1988) Quantum groups, *Proc. Int. Cong. Math.*, Berkley, **1**, 789–820.
2. Drinfeld, V. (1985) Hopf algebras and the quantum Yang–Baxter equation, *Soviet Math. Dokl.*, **32**, 254–258.
3. Drinfeld, V. (1988) A new realization of Yangians and quantized affine algebras, *Soviet Math. Dokl.*, **36**, 212–216.
4. Chari, V. and Pressley, A. (1995) *A guide to quantum groups*, Camb. Univ. Press, Cambridge.
5. Chari, V. and Pressley, A. (1990) Yangians and R-matrices, *L'Enseignment Mathematique*, **36**, 267–302.

6. Chari, V. and Pressley, A. (1991) Fundamental representations of Yangians and singularities of R-matrices, *J. Reine Angew. Math.*, **417**, 87–128.
7. Stukopin, V. (1994) On Yangians of Lie Superalgebras of Type $A(m,n)$. *Funct. Analysis and Its Appl.*, **28**, no. 3, 217–219.
8. Nazarov, M. (1991) Quantum Berezinian and the classical Capelly identity, *Lett. Math. Phys.*, **21**, 123–131.
9. Kac, V. (1977) A Sketch of Lie Superalgebra Theory, *Commun. Math. Phys.*, **53**, 31–64.
10. Khoroshkin, S. M. and Tolstoy, V. N. (1996) Yangian Double, *Lett. Math. Phys.*, **36**, 373–402.
11. Smirnov, F. (1992) Dynamical symmetries of massive integrable models, *J. Modern Phys.* A, **7**, suppl. 1B, 813–838.
12. Levendorskii, S. (1993) On PBW-Bases for Yangians, *Lett. Math. Phys.*, **27**, 37–42.
13. Levendorskii, S. (1993) On generators and defining relations of Yangians, *J. Geom. Phys.*, **12**, 1–11.
14. Frappat, L. and Sorba, P. (1996) Dictionary on Lie Superalgebras, `hep-th/9607161`.

IDEMPOTENT (ASYMPTOTIC) MATHEMATICS AND THE REPRESENTATION THEORY

G. L. LITVINOV (islc@dol.ru)
International Sophus Lie Centre
Moscow
Russia

V. P. MASLOV (islc@dol.ru)
Physics Department
Moscow State University
Moscow
Russia

G. B. SHPIZ* (islc@dol.ru)
International Sophus Lie Centre
Moscow
Russia

Abstract. A brief introduction to Idempotent Mathematics is presented. Idempotent Mathematics can be treated as a result of a dequantization of the traditional Mathematics as the Planck constant tends to zero taking pure imaginary values. In the framework of Idempotent Mathematics some basic concepts and results of the theory of group representations (including some unexpected theorems of the Engel type) are discussed.

1. Introduction

Idempotent Mathematics is based on replacing the usual arithmetic operations by a new set of basic operations (e.g., such as maximum or minimum), that is on replacing numerical fields by idempotent semirings and semifields. Typical (and the most common) examples are given by the so-called (max, +) algebra $\mathbf{R}_{\max}$ and (min, +) algebra $\mathbf{R}_{\min}$. Let $\mathbf{R}$ be the field of real numbers. Then $\mathbf{R}_{\max} = \mathbf{R} \cup \{-\infty\}$ with operations $x \oplus y = \max\{x, y\}$ and $x \odot y = x + y$. Similarly $\mathbf{R}_{\min} = \mathbf{R} \cup \{+\infty\}$ with the operations $\oplus = \min$, $\odot = +$. The new addition $\oplus$ is idempotent, i.e., $x \oplus x = x$ for all elements x. Idempotent Mathematics can be treated as a result of a dequantization of the traditional mathematics over numerical fields as the Planck constant $\hbar$ tends to zero taking pure imaginary values. Some problems that are

* This work was supported by the RFBR grant 99–01–00196 and the Dutch Organization for Scientific Research (N.W.O.).

V. A. Malyshev and A. M. Vershik (eds.),
Asymptotic Combinatorics with Application to Mathematical Physics, 267–278.

nonlinear in the traditional sense turn out to be linear over a suitable idempotent semiring (idempotent superposition principle [1]). For example, the Hamilton–Jacobi equation (which is an idempotent version of the Schrödinger equation) is linear over $\mathbf{R}_{\min}$ and $\mathbf{R}_{\max}$.

The basic paradigm is expressed in terms of an *idempotent correspondence principle* [2].This principle is similar to the well-known correspondence principle of N. Bohr in quantum theory (and closely related to it). Actually, there exists a heuristic correspondence between important, interesting and useful constructions and results of the traditional mathematics over fields and analogous constructions and results over idempotent semirings and semifields (i.e., semirings and semifields with idempotent addition).

A systematic and consistent application of the idempotent correspondence principle leads to a variety of results, often quite unexpected. As a result, in parallel with the traditional mathematics over rings, its "shadow", the Idempotent Mathematics, appears. This "shadow" stands approximately in the same relation to the traditional mathematics as classical physics to quantum theory. In many respects Idempotent Mathematics is simpler than the traditional one. However, the transition from traditional concepts and results to their idempotent analogs is often nontrivial.

There is an idempotent version of the theory of linear representations of groups. We shall present some basic concepts and results of the idempotent representation theory. In the framework of this theory the well-known Legendre transform can be treated as an $\mathbf{R}_{\max}$-version of the traditional Fourier transform (this observation is due to V. P. Maslov). We shall discuss some unexpected theorems of the Engel type.

In this paper we present the so-called algebraic approach to Idempotent Mathematics: basic notions and results are 'simulated' in pure algebraic terms. Historical surveys and the corresponding references can be found in [2–6].

2. Semirings, semifields, and idempotent dequantization

Consider a set S equipped with two algebraic operations: *addition* $\oplus$ and *multiplication* $\odot$. It is a *semiring* if the following conditions are satisfied:

- the addition $\oplus$ and the multiplication $\odot$ are associative;
- the addition $\oplus$ is commutative;
- the multiplication $\odot$ is distributive with respect to the addition $\oplus$: $x\odot(y\oplus z) = (x \odot y) \oplus (x \odot z)$ and $(x \oplus y) \odot z = (x \odot z) \oplus (y \odot z)$ for all $x, y, z \in S$.

A *unity* of a semiring S is an element $\mathbf{1} \in S$ such that $\mathbf{1} \odot x = x \odot \mathbf{1} = x$ for all $x \in S$. A *zero* of a semiring S is an element $\mathbf{0} \in S$ such that $\mathbf{0} \neq \mathbf{1}$ and $\mathbf{0} \oplus x = x$, $\mathbf{0} \odot x = x \odot \mathbf{0} = \mathbf{0}$ for all $x \in S$. A semiring S is called an *idempotent semiring* if $x \oplus x = x$ for all $x \in S$. A semiring S with neutral elements $\mathbf{0}$ and $\mathbf{1}$ is called a *semifield* if every nonzero element of S is invertible.

Let $\mathbf{R}$ be the field of real numbers and $\mathbf{R}_+$ the semiring of all nonnegative real numbers (with respect to the usual addition and multiplication). The change of variables $x \mapsto u = h \ln x$, $h > 0$, defines a map $\Phi_h \colon \mathbf{R}_+ \to S = \mathbf{R} \cup \{-\infty\}$.

Let the addition and multiplication operations be mapped from $\mathbf{R}$ to S by Φ_h, i.e., let $u \oplus_h v = h \ln(\exp(u/h) + \exp(v/h))$, $u \odot v = u + v$, $\mathbf{0} = -\infty = \Phi_h(0)$, $\mathbf{1} = 0 = \Phi_h(1)$. It can easily be checked that $u \oplus_h v \to \max\{u, v\}$ as $h \to 0$ and S forms a semiring with respect to addition $u \oplus v = \max\{u, v\}$ and multiplication $u \odot v = u + v$ with zero $\mathbf{0} = -\infty$ and unit $\mathbf{1} = 0$. Denote this semiring by $\mathbf{R}_{\max}$; it is *idempotent*, i.e., $u \oplus u = u$ for all its elements. The semiring $\mathbf{R}_{\max}$ is actually a semifield. The analogy with quantization is obvious; the parameter h plays the rôle of the Planck constant, so $\mathbf{R}_+$ (or $\mathbf{R}$) can be viewed as a "quantum object" and $\mathbf{R}_{\max}$ as the result of its "dequantization". A similar procedure gives the semiring $\mathbf{R}_{\min} = \mathbf{R} \cup \{+\infty\}$ with the operations $\oplus = \min$, $\odot = +$; in this case $\mathbf{0} = +\infty$, $\mathbf{1} = 0$. The semirings $\mathbf{R}_{\max}$ and $\mathbf{R}_{\min}$ are isomorphic. Connections with physics and imaginary values of the Planck constant are discussed below. The idempotent semiring $\mathbf{R} \cup \{-\infty\} \cup \{+\infty\}$ with the operations $\oplus = \max$, $\odot = \min$ can be obtained as a result of a "second dequantization" of $\mathbf{R}$ (or $\mathbf{R}_+$). Dozens of interesting examples of nonisomorphic idempotent semirings may be cited as well as a number of standard methods of deriving new semirings from these (see, e.g., [2–6] and below).

Idempotent dequantization is related to logarithmic transformations that were used in the classical papers of E. Schrödinger [7] and E. Hopf [8]. The subsequent progress of E. Hopf's ideas has culminated in the well-known vanishing viscosity method (the method of viscosity solutions), see, e.g., [9].

3. Idempotent Analysis

Let S be an arbitrary semiring with idempotent addition $\oplus$ (which is always assumed to be commutative), multiplication $\odot$, zero $\mathbf{0}$, and unit $\mathbf{1}$. The set S is supplied with the *standard partial order* $\preccurlyeq$: by definition, $a \preccurlyeq b$ if and only if $a \oplus b = b$. Thus all elements of S are positive: $\mathbf{0} \preccurlyeq a$ for all $a \in S$. Due to the existence of this order, Idempotent Analysis is closely related to lattice theory, the theory of vector lattices, and the theory of ordered spaces. Moreover, this partial order allows to model a number of basic notions and results of Idempotent Analysis at the purely algebraic level; in this paper we develop this line of reasoning systematically. Let us notice that the standard partial order can be defined for an arbitrary commutative semigroup with idempotent addition.

Calculus deals mainly with functions whose values are numbers. The idempotent analog of a numerical function is a map $X \to S$, where X is an arbitrary set and S is an idempotent semiring. Functions with values in S can be added, multiplied by each other, and multiplied by elements of S.

The idempotent analog of a linear functional space is a set of S-valued functions that is closed under addition of functions and multiplication of functions by elements of S, or an S-semimodule. Consider, e.g., the S-semimodule $\mathcal{B}(X, S)$ of functions $X \to S$ that are bounded in the sense of the standard order on S.

If $S = \mathbf{R}_{\max}$, then the idempotent analog of integration is defined by the formula

$$I(\varphi) = \int_X^{\oplus} \varphi(x)\, dx = \sup_{x \in X} \varphi(x), \tag{1}$$

where $\varphi \in \mathcal{B}(X,S)$. Indeed, a Riemann sum of the form $\sum\limits_i \varphi(x_i) \cdot \sigma_i$ corresponds to the expression $\bigoplus\limits_i \varphi(x_i) \odot \sigma_i = \max\limits_i\{\varphi(x_i) + \sigma_i\}$, which tends to the right-hand side of (1) as $\sigma_i \to 0$. Of course, this is a purely heuristic argument.

Formula (1) defines the idempotent integral not only for functions taking values in $\mathbf{R}_{\max}$, but also in the general case when any of bounded (from above) subsets of S has the least upper bound.

An idempotent measure on X is defined by $m_\psi(Y) = \sup\limits_{x\in Y} \psi(x)$, where $\psi \in \mathcal{B}(X,S)$. The integral with respect to this measure is defined by

$$I_\psi(\varphi) = \int_X^\oplus \varphi(x)\, dm_\psi = \int_X^\oplus \varphi(x) \odot \psi(x)\, dx = \sup_{x\in X} \big(\varphi(x) \odot \psi(x)\big). \tag{2}$$

Obviously, if $S = \mathbf{R}_{\min}$, then the standard order $\preccurlyeq$ is opposite to the conventional order $\leqslant$, so in this case equation (2) assumes the form

$$\int_X^\oplus \varphi(x)\, dm_\psi = \int_X^\oplus \varphi(x) \odot \psi(x)\, dx = \inf_{x\in X} \big(\varphi(x) \odot \psi(x)\big), \tag{3}$$

where inf is understood in the sense of the conventional order $\leqslant$.

The functionals $I(\varphi)$ and $I_\psi(\varphi)$ are linear over S; their values correspond to limits of Lebesgue (or Riemann) sums. The formula for $I_\psi(\varphi)$ defines the idempotent scalar product of the functions ψ and φ. Various idempotent functional spaces and an idempotent version of the theory of distributions can be constructed on the basis of the idempotent integration, see, e.g., [1, 3–6, 10].The analogy of idempotent and probability measures leads to spectacular parallels between optimization theory and probability theory. For example, the Chapman–Kolmogorov equation corresponds to the Bellman equation (see, e.g., the survey of Del Moral [11] and [5]). Many other idempotent analogs may be cited (in particular, for basic constructions and theorems of functional analysis [4]).

4. The superposition principle and linear problems

Basic equations of quantum theory are linear (the superposition principle). The Hamilton–Jacobi equation, the basic equation of classical mechanics, is nonlinear in the conventional sense. However it is linear over the semiring $\mathbf{R}_{\min}$. Also, different versions of the Bellman equation, the basic equation of optimization theory, are linear over suitable idempotent semirings (V. P. Maslov's idempotent superposition principle), see [1, 3, 6, 10]. For instance, the finite-dimensional stationary Bellman equation can be written in the form $X = H \odot X \oplus F$, where X, H, F are matrices with coefficients in an idempotent semiring S and the unknown matrix X is determined by H and F [12]. In particular, standard problems of dynamic programming and the well-known shortest path problem correspond to the cases $S = \mathbf{R}_{\max}$ and $S = \mathbf{R}_{\min}$, respectively. In [12], it was shown that main optimization algorithms for finite graphs correspond to standard methods for solving systems of linear equations of this type (i.e., over semirings). Specifically, Bellman's shortest path algorithm corresponds to a version of Jacobi's algorithm, Ford's algorithm corresponds to the Gauss–Seidel iterative scheme, etc.

Linearity of the Hamilton–Jacobi equation over $\mathbf{R}_{\min}$ (and $\mathbf{R}_{\max}$) is closely related to the (conventional) linearity of the Schrödinger equation. Consider a classical dynamical system specified by the Hamiltonian

$$H = H(p, x) = \sum_{i=1}^{N} \frac{p_i^2}{2m_i} + V(x), \tag{4}$$

where $x = (x_1, \dots, x_N)$ are generalized coordinates, $p = (p_1, \dots, p_N)$ are generalized momenta, m_i are generalized masses, and $V(x)$ is the potential. In this case the Lagrangian $L(x, \dot{x}, t)$ has the form

$$L(x, \dot{x}, t) = \sum_{i=1}^{N} m_i \frac{\dot{x}_i^2}{2} - V(x), \tag{5}$$

where $\dot{x} = (\dot{x}_1, \dots, \dot{x}_N)$, $\dot{x}_i = dx_i/dt$. The value function $S(x, t)$ of the action functional has the form

$$S(x, t) = \int_{t_0}^{t} L(x(t), \dot{x}(t), t)\, dt, \tag{6}$$

where the integration is performed along a trajectory of the system. The classical equations of motion are derived as the stationarity conditions for the action functional (the Hamilton principle, or the least action principle).

The action functional can be considered as a function taking the set of curves (trajectories) to the set of real numbers. Assume that its range lies in the semiring $\mathbf{R}_{\min}$. In this case the minimum of the action functional can be viewed as the idempotent integral of this function over the set of trajectories or the idempotent analog of the Feynman path integral. Thus the least action principle can be considered as the idempotent version of the well-known Feynman approach to quantum mechanics (which is presented, e.g., in [13]); here, one should remember that the exponential function involved in the Feynman integral is monotone on the real axis. The representation of a solution to the Schrödinger equation in terms of the Feynman integral corresponds to the Lax–Oleĭnik formula for a solution to the Hamilton–Jacobi equation.

Since $\partial S/\partial x_i = p_i$, $\partial S/\partial t = -H(p, x)$, the following Hamilton–Jacobi equation holds:

$$\frac{\partial S}{\partial t} + H\left(\frac{\partial S}{\partial x_i}, x_i\right) = 0. \tag{7}$$

Quantization (see, e.g., [13]) leads to the Schrödinger equation

$$-\frac{\hbar}{i}\frac{\partial \psi}{\partial t} = \widehat{H}\psi = H(\hat{p}_i, \hat{x}_i)\psi, \tag{8}$$

where $\psi = \psi(x, t)$ is the wave function, i.e., a time-dependent element of the Hilbert space $L^2(\mathbf{R}^N)$, and $\widehat{H}$ is the energy operator obtained by substitution of the momentum operators $\widehat{p}_i = \frac{\hbar}{i}\frac{\partial}{\partial x_i}$ and the coordinate operators $\widehat{x}_i\colon \psi \mapsto x_i\psi$ for the variables p_i and x_i in the Hamiltonian function, respectively. This equation is linear in the conventional sense (the quantum superposition principle). The standard procedure of limit transition from the Schrödinger equation to the Hamilton–Jacobi equation is to use the following ansatz for the wave function: $\psi(x,t) = a(x,t)e^{iS(x,t)/\hbar}$, and to keep only the leading order as $\hbar \to 0$ (the 'semiclassical' limit).

Instead of doing this, we switch to imaginary values of the Planck constant $\hbar$ by the substitution $h = i\hbar$, assuming $h > 0$. Thus the Schrödinger equation (8) turns to an analog of the heat equation:

$$h\frac{\partial u}{\partial t} = H\left(-h\frac{\partial}{\partial x_i}, \hat{x}_i\right) u, \tag{9}$$

where the real-valued function u corresponds to the wave function ψ. A similar idea (the switch to imaginary time) is used in the Euclidean quantum field theory (see, e.g., [14]); let us remember that time and energy are dual quantities.

Linearity of equation (8) implies linearity of equation (9). Thus if u_1 and u_2 are solutions of (9), then so is their linear combination

$$u = \lambda_1 u_1 + \lambda_2 u_2. \tag{10}$$

Let $S = -h\ln u$ or $u = e^{-S/h}$ as in Section 2 above. It can easily be checked that equation (9) thus turns to

$$\frac{\partial S}{\partial t} = V(x) + \sum_{i=1}^{N}\frac{1}{2m_i}\left(\frac{\partial S}{\partial x_i}\right)^2 - h\sum_{i=1}^{n}\frac{1}{2m_i}\frac{\partial^2 S}{\partial x_i^2}. \tag{11}$$

This equation is nonlinear in the conventional sense. However, if S_1 and S_2 are its solutions, then so is the function

$$S = \lambda_1 \odot S_1 \underset{h}{\oplus} \lambda_2 \odot S_2 \tag{12}$$

obtained from (10) by means of our substitution $S = -h\ln u$. Here the generalized multiplication $\odot$ coincides with the ordinary addition and the generalized addition $\oplus_h$ is the image of the conventional addition under the above change of variables. As $h \to 0$, we obtain the operations of the idempotent semiring $\mathbf{R}_{\min}$, i.e., $\oplus = \min$ and $\odot = +$, and equation (11) turns to the Hamilton–Jacobi equation (7), since the third term in the right-hand side of equation (11) vanishes.

Thus it is natural to consider the limit function $S = \lambda_1 \odot S_1 \oplus \lambda_2 \odot S_2$ as a solution of the Hamilton–Jacobi equation and to expect that this equation can be treated as linear over $\mathbf{R}_{\min}$. This argument (clearly, a heuristic one) can be extended to equations of a more general form. For a rigorous treatment of (semiring) linearity for these equations see [3, 6] and also [1]. Notice that if h is changed to $-h$, then the resulting Hamilton–Jacobi equation is linear over $\mathbf{R}_{\max}$.

The idempotent superposition principle indicates that there exist important problems that are linear over idempotent semirings.

5. Convolution and the Fourier–Legendre transform

Let G be a group. Then the space $\mathcal{B}(X, \mathbf{R}_{\max})$ of all bounded functions $G \to \mathbf{R}_{\max}$ (see above) is an idempotent semiring with respect to the following analog $\circledast$ of the usual convolution:

$$(\varphi(x) \circledast \psi)(g) == \int_G^{\oplus} \varphi(x) \odot \psi(x^{-1} \cdot g)\, dx = \sup_{x \in G}(\varphi(x) + \psi(x^{-1} \cdot g)). \tag{13}$$

Of course, it is possible to consider other "function spaces" (and other basic semirings instead of $\mathbf{R}_{\max}$). In [3] "group semirings" of this type are referred to as *convolution semirings.*

Let $G = \mathbf{R}^n$, where $\mathbf{R}^n$ is considered as a topological group with respect to the vector addition. The conventional Fourier–Laplace transform is defined as

$$\varphi(x) \mapsto \tilde{\varphi}(\xi) = \int_G e^{i\xi \cdot x} \varphi(x)\, dx, \tag{14}$$

where $e^{i\xi \cdot x}$ is a character of the group G, i.e., a solution of the following functional equation:

$$f(x + y) = f(x) f(y).$$

The idempotent analog of this equation is

$$f(x + y) = f(x) \odot f(y) = f(x) + f(y),$$

so "continuous idempotent characters" are linear functionals of the form $x \mapsto \xi \cdot x = \xi_1 x_1 + \cdots + \xi_n x_n$. As a result, the transform in (14) assumes the form

$$\varphi(x) \mapsto \tilde{\varphi}(\xi) = \int_G^{\oplus} \xi \cdot x \odot \varphi(x)\, dx = \sup_{x \in G}(\xi \cdot x + \varphi(x)). \tag{15}$$

The transform in (15) is nothing but the *Legendre transform* (up to some notation) [10]; transforms of this kind establish the correspondence between the Lagrangian and the Hamiltonian formulations of classical mechanics.

Of course, this construction can be generalized to different classes of groups and semirings. Transformations of this type convert the generalized convolution $\circledast$ to the pointwise (generalized) multiplication and possess analogs of some important properties of the usual Fourier transform. For the case of semirings of Pareto sets the corresponding version of the Fourier transform reduces the multicriterial optimization problem to a family of singlecriterial problems [15].

The examples discussed in this sections can be treated as fragments of an idempotent version of the representation theory. In particular, "idempotent" representations of groups can be examined as representations of the corresponding convolution semirings (i.e. idempotent group semirings) in semimodules. To present nontrivial examples from the idempotent version of the representation theory we need some preliminary material.

6. Idempotent semimodules and linear spaces

Recall that an idempotent semigroup is an arbitrary commutative (additive) semigroup with idempotent addition. It can be treated as an ordered set with the following partial order: $x \preccurlyeq y$ if and only if $x \oplus y = y$. It is easy to see that this order is well-defined and $x \oplus y = \sup\{x, y\}$. For an arbitrary subset X of an idempotent semigroup, we put $\oplus X = \sup(X)$ and $\wedge X = \inf(X)$ if the corresponding right-hand sides exist. An idempotent semigroup is called *b-complete* (or *boundedly complete*) if any of its subsets bounded from above (including the empty subset) has the least upper bound. In particular, any b-complete idempotent semigroup contains zero (denoted by $\mathbf{0}$), which coincides with $\oplus\varnothing$, where $\varnothing$ is the empty set. A homomorphism of b-complete idempotent semigroups is called a *b-homomorphism* if $g(\oplus X) = \oplus g(X)$ for any subset X bounded from above.

An idempotent semifield is called *b-complete* if it is b-complete as an idempotent semigroup. In any b-complete semifield, the generalized distributive laws

$$a \odot (\oplus X) = \oplus(a \odot X), \quad a \odot (\wedge X) = \wedge(a \odot X) \tag{16}$$

are valid; here a is an element of the semifield and X is a nonempty bounded subset. It is easy to see that $\mathbf{R}_{\max}$ is a b-complete semifield.

An *idempotent semimodule* over an idempotent semiring K is an idempotent semigroup V endowed with a multiplication $\odot$ by elements of K such that, for any $a, b \in K$ and $x, y \in V$, the usual laws

$$a \odot (b \odot x) = (a \odot b) \odot x, \tag{17}$$

$$(a \oplus b) \odot x = a \odot x \oplus b \odot x, \tag{18}$$

$$a \odot (x \oplus y) = a \odot x \oplus a \odot y, \tag{19}$$

$$\mathbf{0} \odot x = \mathbf{0} \tag{20}$$

are valid. An idempotent semimodule over an idempotent semifield is called an *idempotent space.* An idempotent b-complete space V over a b-complete semifield K is called an *idempotent b-space* if, for any nonempty bounded subset $Q \subset K$ and any $x \in V$, the relations

$$(\oplus Q) \odot x = \oplus(Q \odot x), \quad (\wedge Q) \odot x = \wedge(Q \odot x) \tag{21}$$

hold. A homomorphism $g : V \to W$ of b-spaces is called a *b-homomorphism*, or a *b-linear operator (mapping)*, if $g(\oplus X) = \oplus g(X)$ for any bounded subset $X \subset V$. More general definitions (for spaces which may not be b-complete) can be found in [4]. Homomorphisms taking values in K (treated as a semimodule over itself) are called *linear functionals.* A subset of an idempotent space is called a *subspace* if it is closed with respect to addition and multiplication by coefficients. A subspace in a b-space is called a *b-closed subspace* if it is closed with respect to summation over arbitrary bounded (in V) subsets. This subspace has a natural structure of b-space; it is also a b-subspace in V in the sense of [4].

For an arbitrary set X and an idempotent space V over a semifield K, we use $B(X, V)$ to denote the semimodule of all bounded mappings from X into V with pointwise operations. If V is an idempotent b-space, then $B(X,V)$ is a b-space. A mapping f from a topological space X into an ordered set V is called *upper semicontinuous* if, for any $b \in V$, the set $\{x \in X | f(x) \succcurlyeq b\}$ is closed in X, see [4]. In the case where V is the set of real numbers, this definition coincides with the usual definition of upper semicontinuity of a real function. The set of all bounded upper semicontinuous mappings from X to V is denoted by $USC(X,V)$. If V is an idempotent b-space, then $USC(X,V)$ is also a b-space with respect to the operations $f \oplus g = \sup\{f,g\}$ and $(k \odot f)(x) = k \odot f(x)$.

7. Archimedean spaces [16]

In what follows, unless otherwise specified, the symbol K stands for a b-complete idempotent semifield and all idempotent spaces are over K.

A subset M of idempotent b-space V is called *wo-closed* if $\wedge X \in M$ and $\oplus X \in M$ for any linearly ordered subset $X \subset M$ in V. A nondecreasing mapping $f : V \to W$ of b-spaces is called *wo*-continuous if $f(\oplus X) = \oplus f(X)$ and $f(\wedge X) = \wedge f(X)$ for any bounded linearly ordered subset $X \subset V$. Note that an arbitrary isomorphism of ordered sets is *wo*-continuous. It can be shown that the notions of *wo*-closedness and *wo*-continuity coincide with the closedness and continuity with respect to some $T1$ topology defined in an intrinsic way in terms of the order.

Proposition 1. *Suppose that V is an idempotent b-space and W is a wo-closed subsemigroup of V. Then $\oplus X \in W$ for any subset $X \subset W$ bounded in V. In particular, each wo-closed subspace is a b-closed subspace.*

An element x of an idempotent space V is called *Archimedean* if, for any $y \in V$, there exists a coefficient $\lambda \in K$ such that $\lambda \odot x \succcurlyeq y$. For an Archimedean element $x \in V$, the formula $x^*(y) = \wedge\{k \in K | k \odot x \succcurlyeq y\}$ defines a mapping $x^* : V \to K$. If V is an idempotent b-space, then x^* is a b-linear functional and $x^*(y) \odot x \succcurlyeq y$ for any $y \in V$[6]. We say that an Archimedean element $x \in V$ is *wo-continuous* if the functional x^* is *wo*-continuous, and that an idempotent b-space V is *Archimedean* if V contains a *wo*-continuous Archimedean element.

Proposition 2. *If X is a compact topological space, then USC(X,K) is an Archimedean space and the function* $\mathbf{e}$ *identically equal to* $\mathbf{1}$ *is a wo-continuous Archimedean element.*

Note that $\mathbf{e}^*(f) = \sup\{f(x) | x \in X\}$.

Theorem 3. *Any wo-closed subspace of an Archimedean space is an Archimedean space. Any linearly ordered (with respect to the inclusion) family of nonzero wo-closed subspaces of an Archimedean space V has a nonzero intersection.*

Let V be a b-space. A subset $W \subset V$ is called a $\wedge$-*subspace* if it is closed with respect to multiplication by scalars and taking greatest lower bounds of nonempty subsets. By this definition, any such W is a boundedly complete lattice with respect to the order inherited from V. Therefore, any $\wedge$-subspace $W \subset V$ can be treated as a semimodule with respect to the inherited multiplication by scalars and the operations $x \oplus_W y = \sup\{x, y\}$, where sup is over W. In what follows, all $\wedge$-subspaces are considered as semimodules with respect to these operations. The definitions immediately imply that any $\wedge$-subspace of a b-space is a b-space. It is easy to show that $USC(X, V)$ is a $\wedge$-subspace in $B(X, V)$ for any b-space V and any topological space X.

Proposition 4. *If V is an Archimedean b-space and $x \in V$ is a wo-continuous Archimedean element, then any $\wedge$-subspace W of V containing x is an Archimedean b-space.*

An arbitrary semiring K is called *algebraically closed* (or *radicable*, see, e.g. [5]) if for any element $x \in K$ and any positive integer number n there exists an element $y \in K$ such that $y^n = x$. It is easy to show that $\mathbf{R}_{\max}$ is a b-complete algebraically closed semifield.

Theorem 5. *An idempotent b-space V over an algebraically closed b-complete semifield K is Archimedean if and only if there exists a space of the form $USC(X, K)$, where X is a compact topological space, such that V is isomorphic to its $\wedge$-subspace containing constants.*

8. Representations of groups in Archimedean spaces

Suppose that V is an Archimedean idempotent b-space over an algebraically closed b-complete semifield (e.g., over $\mathbf{R}_{\max}$). By $\mathrm{End}(V)$ denote the set of all b-linear operators $V \to V$. This set is an idempotent semigroup with respect to the pointwise sum and it is a b-space over K with respect to the standard multiplication by coefficients from K. The usual multiplication (composition) of maps turns $\mathrm{End}(V)$ into an idempotent semiring (and a b-complete semialgebra over K).

Let G be an abstract group. A *linear representation* $\pi : G \to \mathrm{End}(V)$ of G in an Archimedean space V is a homomorphism from G to the group of all invertible elements in $\mathrm{End}(V)$ (with respect to the composition of operators). The representation π is (topologically) *irreducible* if the space V has no nontrivial *wo*-closed $\pi(G)$-invariant subspaces.

Theorem 6. *Every linear representation of a group G in an Archimedean idempotent space V has a nontrivial irreducible subrepresentation in a wo-closed subspace of V.*

Theorem 7. *Let π be a linear representation of a group G in an Archimedean idempotent space V and for a nonzero element $x \in V$ the orbit $\pi(G)x$ is bounded. Set $a = \oplus(\pi(G)x)$. Then $\pi(g)a = a$ for each $g \in G$.*

We shall say that a representation π of G in V has a (nonzero) *joint eigenvector* $a \in V$ if $\pi(g)a = \lambda(g)a$ for all $g \in G$, where $\lambda(g) \in K$.

Corollary 8. *Every linear representation of a finite group in an Archimedean idempotent space has a joint eigenvector with a unique eigenvalue* $\mathbf{1}$.

Corollary 9. *Every upper semicontinuous linear representation of a compact group in an Archimedean idempotent space has a joint eigenvector with a unique eigenvalue* **1**.

9. An Engel type theorem for representations of nilpotent groups

Let G be an abstract group. For elements $a, b \in G$ we set $[a,b] = a^{-1}b^{-1}ab$; for subsets X and Y in G we denote by $[X,Y]$ a subgroup in G generated by the set $\{[x,y] | x \in X, y \in Y\}$; we set $\Gamma_1(G) = G$, $\Gamma_i(G) = [G, \Gamma_{i-1}(G)]$, $i = 1,2,3,\ldots$ Recall that an abstract group G is *nilpotent* if and only if there exists a positive integer number n such that $\Gamma_n(G) = \{e\}$, where e is the neutral element (identity) of G.

Theorem 10. *Every linear representation of a nilpotent abstract group in an Archimedean idempotent space over an algebraically closed b-complete semifield (e.g., over* $\mathbf{R}_{\max}$*) has a joint eigenvector.*

Corollary 11. *Every collection of commuting invertible b-linear operators in an Archimedean idempotent space has a joint eigenvector.*

Corollary 12. *Every invertible b-linear operator in an arbitrary Archimedean idempotent space over an algebraically closed b-complete semifield has an eigenvector.*

Remark. There is no idempotent version of the well known Lie theorem for representations of abstract solvable groups in idempotent spaces. Moreover, there exists an irreducible linear representation of a solvable group in the idempotent space $V = \mathbf{R}_{\max} \times \mathbf{R}_{\max}$ over $\mathbf{R}_{\max}$.

References

1. Maslov, V. P. (1987) On a new superposition principle for optimization problems, *Russian Math. Surveys,* **42**, no. 3, 43–54.
2. Litvinov, G. L. and Maslov, V. P. (1995) Correspondence principle for idempotent calculus and some computer applications, Institut des Hautes Etudes Scientifiques, Bures-sur-Yvette, (IHES/M/95/33); also in: [5], 420–443.
3. Kolokoltsov, V. N. and Maslov, V. P. (1997) *Idempotent Analysis and Applications,* Kluwer Acad. Publ., Dordrecht.
4. Litvinov, G. L., Maslov, V. P. and Shpiz, G. B. (2001) Idempotent functionals analysis: an algebraic approach, *Mathematical Notes,* **69**, no. 5, 696–729; `math.FA/0009128`.
5. Gunawardena, J., ed. (1998) *Idempotency,* Publ. of the Newton Institute, Cambridge Univ. Press, Cambridge.
6. Maslov, V. P. and Samborskiĭ, S. N., eds. (1992) *Idempotent Analysis,* Providence, R.I. Amer. Math. Soc., Adv. Sov. Math., vol. 13.
7. Schrödinger, E. (1926) Quantization as an eigenvalue problem, *Annalen der Physik,* **364**, 361–376, in German.
8. Hopf, E. (1950) The partial differential equation $u_t + uu_x = \mu u_{xx}$, *Comm. Pure Appl. Math.,* **3**, 201–230.
9. Capuzzo Dolcetta, I. and Lions, P.-L., eds. (1997) Viscosity Solutions and Applications, in *Springer Lecture Notes in Math.,* Vol. 1660.

10. Maslov, V. P. (1987) *Méthodes opératorielles,* Mir, Moscou.
11. Del Moral, P. (1997) A survey of Maslov optimization theory, in V. N. Kolokoltsov and V. P. Maslov, (Eds.) *Idempotent Analysis and Applications,* Kluwer Acad. Publ., Dordrecht.
12. Carré, B. A. (1971) An algebra for network routing problems, *J. Inst. Math. Appl.*, **7**, 273–294.
13. Feynman, R. and Hibbs, A. (1965) *Quantum Mechanics and Path Integrals*, Mc Graw-Hill, New York.
14. Nelson, E. (1973) *Probability Theory and Euclidian Field Theory,* Constructive Quantum Field Theory, vol. 25, Springer, Berlin.
15. Samborskiĭ, S. N. and Tarashchan, A. A. (1992) The Fourier transform and semirings of Pareto sets, in V. P. Maslov and S. N. Samborskiĭ(Eds.), *Idempotent Analysis,* Providence, R.I. Amer. Math. Soc., Adv. Sov. Math., vol. 13, 139–150.
16. Shpiz, G. B. (2000) A theorem on eigenvector in idempotent spaces, *Doklady Mathematics,* **62**, no. 2 169–171.

A NEW APPROACH TO BEREZIN KERNELS AND CANONICAL REPRESENTATIONS

G. VAN DIJK (dijk@math.leidenuniv.nl)
Mathematical Institute
University of Leiden
P.O Box 9512
2300 RA Leiden
The Netherlands

Abstract. We propose, following ideas of S. C. Hille, another approach to the theory of Berezin kernels than the ususal one in the context of Hermitian symmetric spaces and spaces of Hermitian type. Our context is much more general and circles around so-called $\sigma\theta$-stable parabolic subgroups and intertwining operators. We present new examples of Berezin kernels and also highlight the new approach in the context of interpolation of representations between $L^2(SU(1,n,\mathbb{F})/S(U(1,\mathbb{F}) \times U(n,\mathbb{F})))$ and its compact analogue $L^2(SU(n+1,\mathbb{F})/S(U(1,\mathbb{F}) \times (U(n,\mathbb{F})))$ for $\mathbb{F} = \mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$, in the spirit of Neretin [27].

1. Introduction

Since the introduction of Berezin kernels [2–4] and canonical representations [36], both topics have been the object of research for several years, with a concentration in the past five years. Interesting ideas by Molchanov and Neretin, the first one from Berezin's school, have had great influence on this research. But also others have contributed to new insights. We mention in alphabetical order the names of Arazy, Engliš, Nomura, Olafsson, Ørsted, Pasquale, Peetre, Pevzner, Zhang. We refer to [7] for a rather complete list of references.

The decomposition of the canonical representations into irreducible unitary representations was given (without proof) by Berezin for the classical Hermitian symmetric spaces in [4]. Attempts to find a conceptual proof, valid for all Hermitian symmetric spaces, succeeded in 1994 by Upmeier and Unterberger [35]. The link with the theory of Jordan algebras which they exploited, has shown to be very fruitful. It supplied the ingredients for a generalization of canonical representations to spaces of Hermitian type (see [12], [21]). The decomposition in this case was almost simultaneously discovered by Van Dijk and Pevzner [12], Nomura and Zhang. Perhaps even more interesting was the observation by Repka [32], Ørsted and Zhang [29], and Van Dijk and Pevzner [12] that canonical representations occur as tensor products of holomorphic and anti-holomorphic representations,

V. A. Malyshev and A. M. Vershik (eds.),
Asymptotic Combinatorics with Application to Mathematical Physics, 279–305.

and of their analytic continuations. In the context of spaces of Hermitian type they can be seen as restrictions of holomorphic representations to a "symmetric" subgroup.

Generalization of the canonical representations to line bundles was a next step, set by Zhang [38], Van Dijk and Pevzner [12]. For a further step to vector bundles, see [11], [13].

Then comes the analytic continuation of the canonical representations τ_λ ($\lambda \in \mathbb{C}$). After a succesful attempt by Van Dijk and Hille for rank one spaces [8], the problem was settled in its generality in the context of Hermitian symmetric spaces by Neretin [26]. This is a very hard piece of work, which has lead to new observations. Taking, after analytic continuation, λ negative integer, all expressions with integrals turn out to disappear and we are left with a (discrete) sum, which can be related to a representation of a compact form of our Hermitian symmetric spaces X. This raises the question of a definition of canonical representations for compact Hermitian symmetric spaces, but shows primarily a nice feature of canonical representations: they show up as interpolation between two regular representations, one on $L^2(X)$ ($\lambda \to \infty$) and one on the L^2-space of a compact form of X ($\lambda \to -\infty$). All this is due to Neretin [27]. Extension of Neretin's results to spaces of Hermitian type would be interesting, see Section 10 for a first attempt.

We now come to attempts to generalize Berezin kernels to other spaces than the (Riemannian) Hermitian symmetric spaces and the spaces of Hermitian type. The first step was set here by Molchanov in [24], where he extended the notions of Berezin kernel and canonical representation to para-Hermitian symmetric spaces. Van Dijk and Molchanov gave a detailed account of the rank one para-Hermitian spaces $SL(n,\mathbb{R})/GL(n-1,\mathbb{R})$ [9], [10]. Clearly the new setup even deals with pseudo-Riemannian spaces. But more general spaces are allowed if we follow the, up to now most general, approach by S. C. Hille in his Leiden thesis [20]. We will give a review of this approach and add new spaces to those mentioned above, where Berezin kernels have a natural meaning. Among these spaces are in any case the compact Hermitian symmetric spaces. A new case is also considered in [25], where we investigate Berezin kernels on complex hyperbolic spaces (pseudo-Riemannian) and see what remains of the interpolation property. The result is amazing.

In Hille's approach not only the space is crucial, but in fact we have to consider a quadruple (G, σ, θ, P), where G is a semisimple Lie group (connected, with finite center), θ a Cartan involution, σ another involution commuting with θ and P a $\sigma\theta$-stable parabolic subgroup. The space we are considering for the associated Berezin kernel is then $H/H \cap P$, where H is the fixed point group of σ. Though we sometimes loose some important property for Berezin kernels, namely being positive-definite for $\lambda > \rho$ for some constant ρ, there remains enough interesting other properties in the general case, considered by Hille.

This paper is organized as follows. After introducing Hille's approach, we give several examples, old and new, where the theory can be applied. We finish with a section (Section 10) where one can see the new approach beautifully in action. This section also describes the amazing interpolation property, discovered by Neretin, for the hyperbolic spaces $X = SU(1,n,\mathbb{F})/S(U(1,\mathbb{F}) \times U(n,\mathbb{F}))$ where $\mathbb{F} = \mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$.

We thank S. C. Hille and V. F. Molchanov for numerous interesting and valuable discussions.

This paper is both a review and an extension of an important part of Hille's thesis, defended at Leiden University [20].

2. $\sigma\theta$-stable parabolic subgroups

Let $\mathfrak{g}$ be a noncompact semisimple real Lie algebra with complexification $\mathfrak{g}_c$. Let $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ be a Cartan decomposition and let θ denote the corresponding Cartan involution. We add another involution σ of $\mathfrak{g}$, commuting with θ, with decomposition $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{q}$ into $+1$ and -1 eigen spaces of σ.

Let G_c be a connected, simply connected Lie group with Lie algebra $\mathfrak{g}_c$ and G, K the analytic subgroups corresponding to $\mathfrak{g}$, $\mathfrak{k}$ respectively. Both G and K are closed in G_c.

The involution σ is extended linearly to $\mathfrak{g}_c$ and lifted to G_c. It leaves G invariant. The same holds for θ. Let H denote the closed subgroup of G consisting of the fixed points of σ. Notice that K is the one for θ. Clearly $\mathfrak{h}$ is the Lie algebra of H, but H may not be connected. It is reductive and has finitely many connected components.

Let P denote a $\sigma\theta$-stable parabolic subgroup of G with Langlands decomposition $P = MAN$. The Levi part $M_1 = MA = P \cap \theta(P)$ is σ- and θ-stable. It follows that both M and A are σ-stable. Moreover $\overline{N} := \theta(N) = \sigma(N)$. Let $\mathfrak{a}$ be the Lie algebra of A. Put $\mathfrak{a}_\mathfrak{q} = \mathfrak{a} \cap \mathfrak{q}$ and $\mathfrak{a}_\mathfrak{h} = \mathfrak{a} \cap \mathfrak{h}$. Then $\mathfrak{a} = \mathfrak{a}_\mathfrak{h} \oplus \mathfrak{a}_\mathfrak{q}$, an orthogonal sum with respect to the Killing form. Let $A_\mathfrak{h} = A \cap H$ and $A_\mathfrak{q} = \{a \in A : \sigma(a) = a^{-1}\}$. Then A has a direct product decomposition $A = A_\mathfrak{h}\, A_\mathfrak{q}$.

Set $M_\sigma = MA_\mathfrak{h}$. So $P = M_\sigma\, A_\mathfrak{q}\, N$, the σ-Langlands decomposition of P.

The following lemmae are due to Van den Ban [1]:

Lemma 1. *Let P be a $\sigma\theta$-stable parabolic subgroup with σ-Langlands decomposition $P = M_\sigma\, A_\mathfrak{q} N$. Then $M_1 = M_\sigma\, A_\mathfrak{q}$ is the centralizer of $A_\mathfrak{q}$ in G and $\mathfrak{a}_\mathfrak{q} = center(\mathfrak{m}_1) \cap \mathfrak{p} \cap \mathfrak{q}$.*

Let $\mathfrak{n}$ be the Lie algebra of N.

Lemma 2. *Set $\rho_P(X) = \frac{1}{2}\,\mathrm{trace}\, ad(X)\Big|_\mathfrak{n}$ $(X \in \mathfrak{a})$. If $P = MAN$ is $\sigma\theta$-stable, then $\rho_P = 0$ on $\mathfrak{a}_\mathfrak{h}$.*

3. Open H-orbits on G/P

Let P be, as before, a $\sigma\theta$-stable parabolic subgroup. Clearly, by dimension arguments, HP is open in G. The H-orbit of the origin of G/P is thus open in G/P and isomorphic to $H/H \cap M_1 = H/H \cap M_\sigma = H/H \cap M \cdot A_\mathfrak{h}$. So this orbit carries a H-invariant measure, since both H and $H \cap M_1$ are reductive groups with finitely many connected components. What about the other open H-orbits on G/P? According to a result by Helminck and Helminck [19] we have:

Proposition 3. *If HwP is open, then we can choose w such that wPw^{-1} is $\sigma\theta'$-stable with respect to some Cartan involution θ', commuting with σ. In particular HwP carries an H-invariant measure.*

We shall therefore restrict to open orbits of the form HP/P on G/P.

4. Induced representations

In this section, let P be an arbitrary parabolic subgroup of G with Langlands decomposition $P = MAN$. One has $G = KP$. Set $K_M = K \cap M = K \cap P$. Let (δ, V_δ) be a Hilbert representation of M. Denote by $\langle\,.\,,\,.\,\rangle_{V_\delta}$ the inner product on V_δ. We shall assume that $\delta\,|\,K_M$ is unitary and that δ has an infinitesimal character. The latter condition is necessary for the application of various results of Wallach. Let $\delta^*(m) = \delta(m^{-1})^*$, where $*$ denotes Hilbert space adjoint. Then (δ^*, V_δ) is also a representation of M with infinitesimal character. One has $\delta^* = \delta$ if δ is unitary.

Let $\nu \in \mathfrak{a}_c^*$ and denote by $\omega_{\delta,\nu}$ the representation of P on V_δ given by:

$$\omega_{\delta,\nu}(man) = a^{\rho_P+\nu}\,\delta(m) \quad (m \in M,\, a \in A,\, n \in N). \tag{1}$$

For simplicity of the presentation we shall assume for now on that $\dim V_\delta < \infty$.

The (Hilbert) induced representation $\mathrm{Ind}_P^G(\delta : \nu) = \pi_P(\delta : \nu)$ acts on the Hilbert space of functions satisfying

$$f(xp) = \omega_{\delta,\nu}(p^{-1})\,f(x) \quad (x \in G,\, p \in P) \tag{2}$$

and with inner product

$$\langle f_1,\, f_2\rangle := \int_K \langle f_1(k),\, f_2(k)\rangle_{V_\delta} dk. \tag{3}$$

The action is

$$\pi_P(\delta : \nu)(g) f(x) = f(g^{-1}x). \tag{4}$$

We denote the Hilbert space by $I_P(\delta : \nu)$ and by $I_P^\infty(\delta : \nu)$ the space of C^∞-functions in $I_P(\delta : \nu)$.

Define a sesqui-linear pairing $\langle\,.\,,\,.\,\rangle_\delta : I_P^\infty(\delta,\, \nu) \times I_P^\infty(\delta^*,\, -\overline{\nu}) \to \mathbb{C}$ by

$$\langle f_1,\, f_2\rangle_\delta := \int_{K/K\cap M} \langle f_1(k),\, f_2(k)\rangle_{V_\delta} dk = \int_{\overline{N}} \langle f_1(\overline{n}),\, f_2(\overline{n})\rangle_{V_\delta} d\overline{n}. \tag{5}$$

where $\overline{N} = \theta(N)$. It is G-invariant and non-degenerate.

5. Intertwining operators

Let P be as in Section 4 and let $\overline{P} = \theta(P) = MA\overline{N}$ be the opposite parabolic subgroup. We denote by $A_{\overline{P}\,|\,P}(\delta\ :\ \nu)$ the standard Knapp–Stein intertwining operator from $I_P^\infty(\delta\,,\,\nu)$ to $I_{\overline{P}}^\infty(\delta\,,\,\nu)$, defined formally by

$$A_{\overline{P}\,|\,P}(\delta\,,\,\nu)f(g) = \int_{\overline{N}} f(g\overline{n})\,d\overline{n}. \tag{6}$$

The available literature on these operators is vast. We shall use the approach of Wallach [37].

Let us denote by $\Sigma(P, A)$ the set of non-zero $\alpha \in \mathfrak{a}^*$ such that there is a non-zero $X \in \mathfrak{n}$ with $[H, X] = \alpha(H)X$ for all $H \in \mathfrak{a}$.

According to [37], Chapter 10, one has:

Theorem 4. (i) *There is a constant c_δ such that if* $\operatorname{Re}\langle\nu,\,\alpha\rangle \geq c_\delta$ *for all $\alpha \in \Sigma(P, A)$, then* (6) *converges absolutely and $A_{\overline{P}\,|\,P}(\delta\ :\ \nu)$ is a continuous non-zero G-intertwining operator from $I_P^\infty(\delta\ :\ \nu)$ to $I_{\overline{P}}^\infty(\delta,\,\nu)$.*

(ii) *If $f \in I_P^\infty(\delta,\,\nu)$, then $\nu \to A_{\overline{P}\,|\,P}(\delta,\,\nu)f$ is holomorphic on* $\operatorname{Re}\langle\nu, \alpha\rangle > c_\delta$ *$(\alpha \in \Sigma(P|A))$ and can be meromorphically extended to $\mathfrak{a}_c^*$.*

(iii) *$\langle A_{\overline{P}\,|\,P}(\delta,\,\nu)f_1,\,f_2\rangle_\delta = \langle f_1,\,A_{P\,|\,\overline{P}}(\delta^*,\,-\overline{\nu})f_2\rangle_\delta$ for $f_1 \in I_P^\infty(\delta,\,\nu)$, $f_2 \in I_{\overline{P}}^\infty(\delta^*,\,-\overline{\nu})$ as an identity of meromorphic functions on $\mathfrak{a}_c^*$.*

6. Generalized Berezin forms and Berezin distributions

The actual construction of the generalized Berezin forms is not complicated. They "live" on the space of smooth vectors of a parabolically induced representation, where the parabolic subgroup is $\sigma\theta$-stable. One easily writes down its formal expression (9) using the standard intertwining operator.

But here there are small complications. The intertwining operators form a meromorphic family and one considers parameters in a closed subset that has empty interior. So one has to show first that the operators are defined. This is done in Proposition 7. The invariance of the—now well defined—form originates from an intertwining property of an operator defined by means of the symmetric involution σ. It is *not* intertwining for G, but it is for $G^\sigma = H$.

Let G, σ, θ be as usual, $P = MAN$ a $\sigma\theta$-stable parabolic subgroup of G.

Let δ be as before. We will impose an additional condition on δ soon. We denote by σ_* the map $\sigma_*(f) = f \circ \sigma$ if f is a function on G.

Lemma 5. *σ_* is an isomorphism of the topological vector spaces $I_P^\infty(\delta\ :\ \lambda)$ and $I_{\overline{P}}^\infty(\sigma_*(\delta)\ :\ \sigma_*(\lambda))$. Moreover,*

$$\sigma_* \circ \pi_P\,(\delta\ :\ \lambda)(g) = \pi_{\overline{P}}\,(\sigma_*(\delta)\ :\ \sigma_*(\lambda))\,(\sigma(g)) \circ \sigma_*. \tag{7}$$

Proof. Evident. Notice that $\sigma_*(\rho_P) = -\rho_P = \rho_{\overline{P}}$. □

Lemma 6. *Let $f \in I_P^\infty(\delta : \lambda)$. Then*

$$\sigma_* \circ A_{\overline{P}|P}(\delta : \lambda) f = A_{P|\overline{P}}(\sigma_*(\delta) : \sigma_*(\lambda)) (\sigma_*(f)) \tag{8}$$

as an identity of meromorphic maps of $\lambda \in \mathfrak{a}_c^*$.

Proof. Recall the constant c_δ from Theorem 4. Let c be the maximum of c_δ and $c_{\sigma_*(\delta)}$. If $\lambda \in \mathfrak{a}^*$ is such that $\langle\lambda, \alpha\rangle \geq c$ for all $\alpha \in \Sigma(P, A)$, then $\langle\sigma_*(\lambda), \beta\rangle \geq c$ for all $\beta \in \Sigma(\overline{P}, A)$. Here we use that $\sigma_*(\Sigma(\overline{P}, A)) = \Sigma(P, A)$. Using $\sigma(d\overline{n}) = dn$ yields the result for these values of λ. Now apply analytic continuation. □

Clearly the set $\mathfrak{a}_\mathfrak{q}^+ = \{H : H \in \mathfrak{a}_\mathfrak{q}, \alpha(H) > 0 \; for \; all \, \alpha \in \Sigma(P, A)\}$ is non-empty open in $\mathfrak{a}_\mathfrak{q}$.

Proposition 7. *There exists a non-empty open affine cone Γ_δ in $\mathfrak{a}_\mathfrak{q}^* \subset \mathfrak{a}_c^*$ such that the integral $A_{\overline{P}|P}(\delta : \lambda)f$ is absolutely convergent for all $\lambda \in \Gamma_\delta + i\mathfrak{a}^*$, and $A_{\overline{P}|P}(\delta : \lambda)f$ is continuous and non-zero.*

Proof. There exists a $H_0 \in \mathfrak{a}_\mathfrak{q}$ such that $\alpha(H_0) \geq c_\delta$ for all $\alpha \in \Sigma(P, A)$. Applying the ususal identification by means of the Killing form between $\mathfrak{a}$ and $\mathfrak{a}^*$ (call it j), we set $\lambda_0 = j(H_0)$ and $\Gamma_\delta = \lambda_0 + j(\mathfrak{a}_q^+)$. Then $\operatorname{Re}\langle\lambda , \alpha\rangle \geq c_\delta$ for all $\lambda \in \Gamma_\delta + i\mathfrak{a}^*$ and we can apply Theorem 4 (i). □

For the construction of the generalized Berezin form we impose the following additional condition on δ:

- There is a non-zero symmetric linear endomorphism J of V_δ such that $J\,\delta(\sigma(m)) = \delta^*(m)\,J$ for all $m \in M$.

Such representations δ exist, e.g. one can take the trivial representation. Less trivial examples will appear soon.

J induces an operator on V_δ-valued functions which we also denote by J.

Notice that if $\lambda \in \mathfrak{a}_\mathfrak{q}^* + i\mathfrak{a}_\mathfrak{h}^*$, then $\sigma_*(\lambda) = -\overline{\lambda}$. Define for $\lambda \in \mathfrak{a}_\mathfrak{q}^* + i\mathfrak{a}_\mathfrak{h}^*$, regular for $A_{\overline{P}|P}(\delta : \lambda)$ (e.g. $\lambda \in \Gamma_\delta + i\mathfrak{a}^*$, see Prop. 7), and for $f, g \in I_P^\infty(\delta : \lambda)$:

$$B_{\delta,\lambda}^P(f, g) = \langle J \circ \sigma_* \circ A_{\overline{P}|P}(\delta : \lambda)f, \, g\rangle_{\delta^*}. \tag{9}$$

Proposition 8. *The form $B_{\delta,\lambda}^P$ has the following properties:*

(i) *It is a $\pi_P(\delta : \lambda)(H)$-invariant Hermitian form on $I_P^\infty(\delta : \lambda)$.*

(ii) *If $\lambda \in \mathfrak{a}_\mathfrak{q}^* + i\mathfrak{a}_\mathfrak{h}^*$ regular, then it is non-degenerate if and only if $A_{\overline{P}|P}(\delta : \lambda)$ is injective.*

Proof. (i) Using the relation $\langle J\sigma_*(f), g\rangle_{\delta^*} = \langle f, \sigma_*(J^* g)\rangle_{\delta^*}$, $J^* = J$, Theorem 4 (iii) and Lemma 6, and the fact that J and σ^* commute, one obtains

$$\begin{aligned} B_{\delta,\lambda}^P(f, g) &= \langle A_{\overline{P}|P}(\delta : \lambda)f, \, \sigma_*(Jg)\rangle_\delta = \langle f, \, A_{P|\overline{P}}(\delta^* : -\overline{\lambda})\sigma_*(Jg)\rangle_\delta \\ &= \overline{\langle\sigma_* J A_{\overline{P}|P}(\delta : \lambda)g, \, f\rangle_{\delta^*}} = \overline{B_{\delta,\lambda}^P(g, f)}. \end{aligned}$$

$B_{\delta,\lambda}^P$ is invaraint with respect to $\pi_P(\delta : \lambda)(x) \times \pi_P(\delta : \lambda)(\sigma(x))$, $x \in G$ as Lemma 5 shows. Therefore it is H-invariant.

(ii) is clear. □

Now we restrict f and g in (9) to functions with compact support mod P in HP. This space is naturally isomorphic with $\mathcal{D}(H/H \cap P, \operatorname{Im}\lambda, \delta)$, the space of C^∞-functions on H satisfying

$$f(hma) = \delta(m^{-1})\, a^{-i\operatorname{Im}\lambda} f(h) \tag{10}$$

for $h \in H$, $a \in A \cap H$, $m \in M \cap H$, $\operatorname{Supp} f$ compact modulo $H \cap P$. Given such a function, we assign to it the following function F on HP:

$$F(hp) = \omega_{\delta,\lambda}(p^{-1})\, f(h), \quad (h \in H, p \in P). \tag{11}$$

Applying Schwartz Kernel Theorem, the form $B^P_{\delta,\lambda}$ gives rise to a $\operatorname{Hom}(V_\delta, V_\delta^*)$-valued distribution Ψ on H satisfying:

- $\Psi^* = \Psi$,
- $\Psi(m'a'hma) = \delta^*(m^{-1})\, a^{-i\operatorname{Im}(\lambda)}\Psi(h)\, a'^{-i\operatorname{Im}(\lambda)}\, \delta(m'^{-1})$,

$(m, m' \in M \cap H,\ a, a' \in A \cap H,\ h \in H)$, and (by abuse of notation):

- $B^P_{\delta,\lambda}(f, g) = \langle g^* * f, \Psi\rangle$

for $f, g \in \mathcal{D}(H/H \cap P, \operatorname{Im}\lambda, \delta)$. [Here g^* stands for $g^*(x) = g(x^{-1})^*$, $x \in H$].

The distribution $\Psi = \Psi^P_{\delta,\lambda}$ on H is called a *Berezin distribution* and the associated sesqui-linear Hermitian form $B^P_{\delta,\lambda}$ a *generalized Berezin form.*

7. Explicit expression for $\Psi^P_{\delta,\lambda}$

The $\sigma\theta$-stable parabolic subgroup P gives rise to the Gauss decomposition:

$$G \supset \overline{N}\, M\, A\, N. \tag{12}$$

The right hand side is open and dense in G. Moreover the natural map $\overline{N} \times M \times A \times N \to \overline{N}\, M\, A\, N$ is a diffeomorphism.

For $g \in G$, $g \in \overline{N}\, M\, A\, N$, write $g = \overline{n}man$ and set $a = \underline{a}(g)$, $m = \underline{m}(g)$. Then both $\underline{a}$ and $\underline{m}$ are defined almost everywhere on G.

Proposition 9. *For $\lambda \in \Gamma_\delta + i\mathfrak{a}^*_\mathfrak{h}$, one has a.e. on H:*

$$\Psi^P_{\delta,\lambda}(h) = \underline{a}(h^{-1})^{\lambda-\rho}\, J \circ \delta(\underline{m}(h^{-1})) \quad (h \in H) \tag{13}$$

up to normalization of Haar measures. $\Psi^P_{\delta,\lambda}$ is a locally integrable function for those λ.

Proof. The proof is given in Hille's thesis. It is exactly the same as the proof of Lemma 10.13 in [21], which applies to the special situation of a non-compactly causal triple (G, σ, θ). □

Observe that

$$\Psi^P_{\delta,\lambda}(h^{-1})^* = \underline{a}(h)^{\overline{\lambda}+\rho}\, \delta^*(\underline{m}(h)) \circ J = \Psi^P_{\delta,\lambda}(h).$$

8. Expansion of Berezin distributions into extremal distributions of positive type

In this section we change the notation a little bit. The main ideas here are due to L. Schwartz [34].

Let $\underline{G}$ be a unimodular Lie group and $\underline{H}$ any closed unimodular subgroup. The invariant measure on $\underline{G}$ is denoted by dx, on $\underline{H}$ by dh and on $\underline{G}/\underline{H}$ by $d\overline{x}$. We normalize them in such a way that formally $dx = d\overline{x}\, dh$.

What will be said here corresponds, at least in the beginning of this section, to $\underline{G} = H$, $\underline{H} = H \cap P$, $\delta = 1$, and $\operatorname{Im} \lambda = 0$ in the context of the previous sections.

We will pay a few words to the (abstract) expansion of bi-$\underline{H}$-invariant distributions T on $\underline{G}$ into 'extremal' ones, and refer to [6] for more details. For concrete expansions, see e.g. Section 10 and [9], [10], [24].

First we assume T to be positive-definite.

Denote by $\Gamma_{\underline{G}}$ the cone of bi-$\underline{H}$-invariant positive-definite distributions and by $ext\,(\Gamma_{\underline{G}})$ the subset of distributions that correspond to its extremal rays.

Any positive-definite bi-$\underline{H}$-invariant distribution corresponds in a natural way to a unitary representation on a Hilbert subspace of $\mathcal{D}'(\underline{G}/\underline{H})$, the space of distributions on $\underline{G}/\underline{H}$, the $T \in ext\,(\Gamma_{\underline{G}})$ correspond to irreducible representations. We call the elements of $ext\,(\Gamma_{\underline{G}})$, extremal distributions.

Choose an (admissible) parametrization $s \to T_s$ $(s \in S)$ of $ext\,(\Gamma_{\underline{G}})$ (see [6] for explanation). Then one has:

Proposition 10. *For every $T \in \Gamma_{\underline{G}}$ there exists a (non-necessarily unique) positive Radon measure m on S such that*

$$\langle T, \varphi \rangle = \int_S \langle T_s, \varphi \rangle\, dm(s) \tag{14}$$

for all $\varphi \in \mathcal{D}(\underline{G})$.

Unicity always occurs if the pair $(\underline{G}, \underline{H})$ is a generalized Gelfand pair (see [6] for definition). A typical example is $(SL(n, \mathbb{R}), GL(n-1, \mathbb{R}))$ and, of course, any usual Gelfand pair $(\underline{G}, \underline{H})$ with $\underline{H}$ a compact subgroup.

Taking $T = \delta_{\underline{H}}$ with $\langle \delta_{\underline{H}}, \varphi \rangle = \int_{\underline{H}} \varphi(h) dh$ in Proposition 10, gives the Plancherel formula for $\underline{G}/\underline{H}$.

Our distribution $\Psi^P_{\delta,\lambda}$ $(\delta = 1)$ is not always positive-definite. What to do then? Often one can show that the associated Berezin form is bounded. See for example Theorem 12.3 in [25] and Theorem 8.4 in [9]. Then one can use the following proposition. For any $\varphi \in \mathcal{D}(\underline{G})$, we set $\widetilde{\varphi}(x) = \varphi(x^{-1})$ $(x \in \underline{G})$. Then $\widetilde{\varphi} \in \mathcal{D}(\underline{G})$.

Proposition 11. ([34], Proposition 38) *Let T be a Hermitian bi-$\underline{H}$-invariant distribution satisfying*

$$\langle T, \widetilde{\varphi} * \varphi \rangle \leq \langle L, \widetilde{\varphi} * \varphi \rangle \quad (\varphi \in \mathcal{D}(\underline{G})) \tag{15}$$

for some positive-definite bi-$\underline{H}$-invariant distribution L. Then one has

$$\langle T, \varphi \rangle = \int_S \langle T_s, \varphi \rangle\, dm(s) \quad (\varphi \in \mathcal{D}(\underline{G})) \tag{16}$$

for some, non-necessarily positive, measure m.

The *proof* is easy. Set $T = L + (T - L)$, then L and $T - L$ are positive-definite, bi-$\underline{H}$-invariant distributions on $\underline{G}$. Apply now Proposition 10. □

In particular, if T is bounded, i.e.

$$\langle T, \tilde{\varphi} * \varphi\rangle \leq C\|\varphi\|^2_{\underline{G}/\underline{H}} \quad (\varphi \in \mathcal{D}(\underline{G}))$$

for some constant $C > 0$, where

$$\|\varphi\|^2_{\underline{G}/\underline{H}} = \int_{\underline{G}/\underline{H}} |\varphi_0(\overline{x})|^2 d\overline{x}$$

with $\varphi_0(\overline{x}) = \int_{\underline{H}} \varphi(xh)\, dh$ $(x \in \underline{G};\ \overline{x} \in \underline{G}/\underline{H})$, then Proposition 11 applies with $L = C\delta_{\underline{H}}$.

The situation which corresponds to $\operatorname{Im}\lambda \neq 0$ and $\delta|M \cap H$ is unitary, can be treated in a similar way. Clearly we have therefore to generalize the results in [6] by admitting matrix-valued distributions which transform left and right under the action of $\underline{H}$ according to a finite-dimensional unitary representation of $\underline{H}$.

9. Examples

TABLE I. Irreducible causal symmetric pairs

$\mathfrak{g}$ compactly causal	$\mathfrak{g}^c$ non-compactly causal	$\mathfrak{h}$
$\mathfrak{su}(p,q) \oplus \mathfrak{su}(p,q)$	$\mathfrak{sl}(p+q;\mathbf{C})$	$\mathfrak{su}(p,q)$
$\mathfrak{so}^*(2n) \oplus \mathfrak{so}^*(2n)$	$\mathfrak{so}(2n;\mathbf{C})$	$\mathfrak{so}^*(2n)$
$\mathfrak{so}(2,n) \oplus \mathfrak{so}(2,n)$	$\mathfrak{so}(2+n;\mathbf{C})$	$\mathfrak{so}(2,n)$
$\mathfrak{sp}(n,\mathbf{R}) \oplus \mathfrak{sp}(\mathbf{n},\mathbf{R})$	$\mathfrak{sp}(n,\mathbf{C})$	$\mathfrak{sp}(n,\mathbf{R})$
$\mathfrak{e}_{6(-14)} \oplus \mathfrak{e}_{6(-14)}$	$\mathfrak{e}_6$	$\mathfrak{e}_{6(-14)}$
$\mathfrak{e}_{7(-25)} \oplus \mathfrak{e}_{7(-25)}$	$\mathfrak{e}_7$	$\mathfrak{e}_{7(-25)}$
$\mathfrak{su}(p,q)$	$\mathfrak{sl}(p+q;\mathbf{R})$	$\mathfrak{so}(p,q)$
$\mathfrak{su}(n,n)$	$\mathfrak{su}(n,n)$	$\mathfrak{sl}(n;\mathbf{C}) \oplus \mathbf{R}$
$\mathfrak{su}(2p,2q)$	$\mathfrak{su}^*(2(p+q))$	$\mathfrak{sp}(p,q)$
$\mathfrak{so}^*(2n)$	$\mathfrak{so}(n,n)$	$\mathfrak{so}(n;\mathbf{C})$
$\mathfrak{so}^*(4n)$	$\mathfrak{so}^*(4n)$	$\mathfrak{su}^*(2n) \oplus \mathbf{R}$
$\mathfrak{so}(2,p+q)$	$\mathfrak{so}(p+1,q+1)$	$\mathfrak{so}(p,1) \times \mathfrak{so}(1,q)$
$\mathfrak{sp}(n,\mathbf{R})$	$\mathfrak{sp}(n,\mathbf{R})$	$\mathfrak{sl}(n;\mathbf{R}) \oplus \mathbf{R}$
$\mathfrak{sp}(2n,\mathbf{R})$	$\mathfrak{sp}(n,n)$	$\mathfrak{sp}(n,\mathbf{C})$
$\mathfrak{e}_{6(-14)}$	$\mathfrak{e}_{6(6)}$	$\mathfrak{sp}(2,2)$
$\mathfrak{e}_{6(-14)}$	$\mathfrak{e}_{6(-26)}$	$\mathfrak{f}_{4(-20)}$
$\mathfrak{e}_{7(-25)}$	$\mathfrak{e}_{7(-25)}$	$\mathfrak{e}_{6(-26)} \oplus \mathbf{R}$
$\mathfrak{e}_{7(-25)}$	$\mathfrak{e}_{7(7)}$	$\mathfrak{su}^*(8)$

9.1. HERMITIAN SYMMETRIC SPACES

In this section we recall some structure theory, mainly following [33] and [18], Ch. VIII. Table 1, which is taken from [17], plays an important role in this and the subsequent sections.

Let $\mathfrak{g}$ be a non-compact simple real Lie algebra with complexification $\mathfrak{g}_c$. Let $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ be a Cartan decomposition and let θ denote the corresponding Cartan involution. Let $\mathfrak{z}$ denote the center of $\mathfrak{k}$. The Lie algebra $\mathfrak{g}$ is said to be Hermitian if the centralizer of $\mathfrak{z}$ in $\mathfrak{g}$ is equal to $\mathfrak{k}$. The center of $\mathfrak{k}$ is then one-dimensional and there is an element $Z_0 \in \mathfrak{z}$ such that $(adZ_0)^2 = -1$ on $\mathfrak{p}$. Fixing i a square root of -1, one has $\mathfrak{p}_c = \mathfrak{p} + i\mathfrak{p} = \mathfrak{p}_+ + \mathfrak{p}_-$ where $adZ_0|_{\mathfrak{p}_+} = i$, $adZ_0|_{\mathfrak{p}_-} = -i$. Then

$$\mathfrak{g}_c = \mathfrak{p}_+ \oplus \mathfrak{k}_c \oplus \mathfrak{p}_- \tag{17}$$

and $[\mathfrak{p}_\pm, \mathfrak{p}_\pm] = 0, [\mathfrak{p}_+, \mathfrak{p}_-] = \mathfrak{k}_c$ and $[\mathfrak{k}_c, \mathfrak{p}_\pm] = \mathfrak{p}_\pm$.

Let G_c be a connected, simply connected Lie group with Lie algebra $\mathfrak{g}_c$ and K_c, P_+, P_-, G, K the analytic subgroups corresponding to $\mathfrak{k}_c, \mathfrak{p}_+, \mathfrak{p}_-, \mathfrak{g}$ and $\mathfrak{k}$ respectively. Then K_cP_- (and K_cP_+) is a maximal parabolic subgroup of G_c with split component $A = \exp i\mathbb{R}Z_0$. G is closed in G_c.

Moreover, the exponential mapping is a diffeomorphism of $\mathfrak{p}_-$ onto P_- and of $\mathfrak{p}_+$ onto P_+ ([18] Ch. VIII, Lemma 7.8). Furthermore:

Lemma 12 (see [18], Ch. VIII, Lemmæ 7.9 and 7.10). a. *The mapping $(q, k, p) \mapsto qkp$ is a diffeomorphism of $P_+ \times K_c \times P_-$ onto an open dense submanifold of G_c containing G.*
b. *The set GK_cP_- is open in $P_+K_cP_-$ and $G \cap K_cP_- = K$.*

Thus G/K is mapped on an open, bounded domain $\mathcal{D}$ in $\mathfrak{p}_+$. The group G acts on $\mathcal{D}$ via holomorphic transformations.

Example. Let $\mathfrak{g} = \mathfrak{su}(1,1)$. Then $G_c = SL(2,\mathbb{C})$ and $G = SU(1,1)$. Clearly $Z_0 = \frac{1}{2}\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$, $\mathfrak{p}_\pm$ are one-dimensional and generated by $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ and $\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ respectively. Let $g = \begin{pmatrix} \alpha & \beta \\ \bar\beta & \bar\alpha \end{pmatrix} \in G$. Then the decomposition $g = qkp$ (see Lemma 12.a) is given by

$$g = \begin{pmatrix} 1 & \beta\bar\alpha^{-1} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \bar\alpha^{-1} & 0 \\ 0 & \bar\alpha \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \bar\alpha^{-1}\bar\beta & 1 \end{pmatrix}$$

where $|\alpha|^2 - |\beta|^2 = 1$. The embedding of G/K into $\mathbb{C}$ is given by

$$g \mapsto \beta\bar\alpha^{-1} = \zeta.$$

Since $|\alpha|^2 - |\beta|^2 = 1$, it follows $|\zeta| < 1$. Conversely, let $|\zeta| < 1$. Take then α such that $|\alpha|^2 = (1 - |\zeta|^2)^{-1}$ and let $\beta = \zeta\bar\alpha$. Then $\begin{pmatrix} \alpha & \beta \\ \bar\beta & \bar\alpha \end{pmatrix}$ is mapped onto ζ. So $\mathcal{D}$ is the unit disc "$|\zeta| < 1$". G acts on $\mathcal{D}$ by means of fractional linear transformations

$$g.\zeta = \frac{\alpha\zeta + \beta}{\bar\beta\zeta + \bar\alpha}, \qquad g = \begin{pmatrix} \alpha & \beta \\ \bar\beta & \bar\alpha \end{pmatrix} \in G.$$

Everywhere we shall denote $\bar{g}$ the complex conjugate of $g \in G_c$ with respect to G. So, for example, if $g = \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} \in SL(2, \mathbb{C})$, then its conjugate with respect to $SU(1,1)$ is given by $\bar{g} = \begin{pmatrix} \bar{a}^{-1} & 0 \\ 0 & \bar{a} \end{pmatrix}$. Notice that $\bar{P}_+ = P_-$.

For $g \in P_+ K_c P_-$ we shall write $g = (g)_+ (g)_0 (g)_-$, where $(g)_\pm \in P_\pm$, $(g)_0 \in K_c$. For $g \in G_c$, $z \in \mathfrak{p}_+$ such that $g.\exp z \in P_+ K_c P_-$ we define

$$\exp g(z) = (g.\exp z)_+ \tag{18}$$

$$J(g,z) = (g.\exp z)_0. \tag{19}$$

$J(g,z) \in K_c$ is called the *canonical automorphic factor* of G_c (terminology of Satake).

Lemma 13 (see [33], Ch. II, Lemma 5.1). *J satisfies*

(i) $J(g, o) = (g)_0$, *for* $g \in P_+ K_c P_-$,

(ii) $J(k, z) = k$ *for* $k \in K_c, z \in \mathfrak{p}_+$.

If for $g_1, g_2 \in G_c$ and $z \in \mathfrak{p}_+$, $g_1(g_2(z))$ and $g_2(z)$ are defined, then $(g_1 g_2)(z)$ is also defined and

(iii) $J(g_1 g_2, z) = J(g_1, g_2(z))\, J(g_2, z)$.

For $z, w \in \mathfrak{p}_+$ satisfying $(\exp \bar{w})^{-1}.\exp z \in P_+ K_c P_-$ we define

$$K(z,w) = J((\exp \bar{w})^{-1}, z)^{-1} \tag{20}$$

$$= ((\exp \bar{w})^{-1}.\exp z)_0^{-1}. \tag{21}$$

This expression is always defined for $z, w \in \mathcal{D}$, for then

$$(\exp \bar{w})^{-1}.\exp z \in \overline{(GK_cP_-)}^{-1} GK_cP_- = P_+K_cGK_cP_- = P_+K_cP_-.$$

$K(z,w)$, defined on $\mathcal{D} \times \mathcal{D}$, is called the *canonical kernel* on $\mathcal{D}$ (by Satake). $K(z,w)$ is holomorphic in z, anti-holomorphic in w, with values in K_c. Here are a few properties:

Lemma 14 (see [33], Ch. II, Lemma 5.2). (i) $K(z,w) = \overline{K(w,z)}^{-1}$ *if $K(z,w)$ is defined,*

(ii) $K(o, w) = K(z, o) = 1$ *for* $z, w \in \mathfrak{p}_+$.

If $g(z), \bar{g}(w)$ and $K(z,w)$ are defined, then $K(g(z), \bar{g}(w))$ is also defined and one has:

(iii) $K(g(z), \bar{g}(w)) = J(g,z)\, K(z,w)\, \overline{J(\bar{g}, w)}^{-1}$,

Lemma 15 (see [33], Ch. II, Lemma 5.3). *For $g \in G_c$ the Jacobian of the holomorphic mapping*
$z \mapsto g(z)$, when it is defined, is given by

$$\mathrm{Jac}\,(z \mapsto g(z)) = Ad_{\mathfrak{p}_+}(J(g,z)).$$

For any holomorphic character $\chi : K_c \mapsto \mathbb{C}$ we define:

$$\begin{aligned} j_\chi(g,z) &= \chi(J(g,z)), & (22)\\ k_\chi(z,w) &= \chi(K(z,w)). & (23) \end{aligned}$$

Since $\chi(\bar{k}) = \overline{\chi(k)}^{-1}$ we have :

$$\begin{aligned} k_\chi(z,w) &= \overline{k_\chi(w,z)}, & (24)\\ k_\chi(g(z),\bar{g}(w)) &= j_\chi(g,z)k_\chi(z,w)\overline{j_\chi(\bar{g},w)} & (25) \end{aligned}$$

in place of Lemma 14 (i) and (iii).

The character $\chi_1(k) = \det \mathrm{Ad}_{\mathfrak{p}_+}(k)$, $(k \in K_c)$ is of particular importance. We call the corresponding j_{χ_1}, k_{χ_1}: j_1 and k_1. Notice that

$$j_1(g,z) = \det(\mathrm{Jac}\,(z \mapsto g(z))). \tag{26}$$

Example. $\mathfrak{g} = \mathfrak{su}(1,1)$. For $g = \begin{pmatrix} \alpha & \beta \\ \bar{\beta} & \bar{\alpha} \end{pmatrix}$ in $SU(1,1)$ one has

$$J(g,z) = \begin{pmatrix} (\bar{\beta}z+\bar{\alpha})^{-1} & 0 \\ 0 & (\bar{\beta}z+\bar{\alpha}) \end{pmatrix}, \ K(z,w) = \begin{pmatrix} (1-z\bar{w}) & 0 \\ 0 & (1-z\bar{w})^{-1} \end{pmatrix}$$

and $\chi_1 \begin{pmatrix} \alpha & 0 \\ 0 & \alpha^{-1} \end{pmatrix} = \alpha^2$, $(\alpha \in \mathbb{C}^*)$, so

$$j_1(g,z) = (\bar{\beta}z+\bar{\alpha})^{-2}, \qquad k_1(z,w) = (1-\bar{z}w)^2.$$

Because of (26), $|k_1(z,z)|^{-1}d\mu(z)$, where $d\mu(z)$ is the Euclidean measure on $\mathfrak{p}_+$, is a G-invariant measure on $\mathcal{D}$. Indeed:

$$\begin{aligned} d\mu(g(z)) &= |j_1(g,z)|^2 d\mu(z), \\ k_1(g(z),g(z)) &= j_1(g,z)\,k_1(z,z)\,\overline{j_1(g,z)}, \qquad \text{for } g \in G. \end{aligned}$$

(see (25)). One can actually show that $k_1(z,z) > 0$ on $\mathcal{D}$. ([33], Ch. II, Lemma 5.8).

For the list of groups G considered here, we refer to the upper part of Table I, right column.

We now work in G_c. Let us choose the $\sigma\theta$-stable parabolic subgroup

$$P = K_c P_- \tag{27}$$

of G_c. Here σ is the conjugation in G_c with respect to G and θ the Cartan involution with fixed points $\mathfrak{k}+i\mathfrak{p}$. Set $A = \exp i\mathbb{R}Z_0$. Let $Z(K)$ be the analytic subgroup with Lie algebra $\mathfrak{z}$, and, K_s^c the subgroup with Lie algebra $[\mathfrak{k}_c, \mathfrak{k}_c]$. Set $M = K_s^c Z(K)$. Then $P = MAP_-$ is the Langlands decomposition of P.

Let $n = \dim_{\mathbb{C}} \mathfrak{p}_+$. Then

$$a^\rho = \mathrm{e}^{nt} \tag{28}$$

if $a = \exp itZ_0$, $t \in \mathbb{R}$. We choose δ to be one-dimensional and unitary:

$$\delta(m) = \left(\det_{\mathbb{C}}\left(Ad(m)\Big|_{\mathfrak{p}_+}\right)\right)^l \tag{29}$$

for $l \in \mathbb{Z}$. So the endomorphism J, defined in Section 6, is equal to the identity here.

We then get:

$$\Psi^P_{\delta,\lambda}(h) = k_1(h\cdot o,\, h\cdot o)^{\frac{\lambda-n}{2n}+\frac{l}{2}}\, j_1(h,o)^{-l} \tag{30}$$

($h \in G$). Observe that $H \cap P = G \cap P = K$.

For $l = 0$ we get the classical Berezin kernel, see [12]. $\Psi^P_{\delta,\lambda}(h)$ is defined here for all λ and all $h \in G$. It is holomorphic in λ and C^∞ in h.

9.2. SYMMETRIC SPACES OF HERMITIAN TYPE

Let $\mathfrak{g}, \mathfrak{g}_c, G, G_c, \ldots$ be as in section 9.1. We add to $\mathfrak{g}$ an involutive automorphism σ, commuting with the Cartan involution θ. Let $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{q}$ be the decomposition of $\mathfrak{g}$ into $+1$ and -1 eigenspaces of σ.

The Lie algebra $\mathfrak{g}$ is said to be of *Hermitian type* if $\mathfrak{g}$ is Hermitian and, in addition, $Z_0 \in \mathfrak{q} \cap \mathfrak{k}$. There are several other terminologies in use; the most closely related to us is: $\mathfrak{g}$ is a compactly causal Lie algebra.

The involution σ is extended to $\mathfrak{g}_c$ and G_c and leaves G invariant. Let H denote the closed subgroup of G consisting of the fixed points of σ. The Lie algebra of H is $\mathfrak{h}$.

Now observe that, since $\sigma(Z_0) = -Z_0$, $\sigma(\mathfrak{p}_+) = \mathfrak{p}_-$. Since $\bar{\mathfrak{p}}_+ = \mathfrak{p}_-$, we see that $\bar\sigma$ leaves $\mathfrak{p}_+$ and $\mathfrak{p}_-$ invariant. Set

$$\mathfrak{p}^{\bar\sigma}_{\pm} = \{X \in \mathfrak{p}_\pm \,:\, \bar\sigma(X) = X\}.$$

Then clearly $\dim_{\mathbb{R}} \mathfrak{p}^{\bar\sigma}_+ = \dim_{\mathbb{R}} \mathfrak{p}^{\bar\sigma}_- = \dim_{\mathbb{C}} \mathfrak{p}_+$, since $\bar\sigma$ is a conjugation.

It is clear that $\bar\sigma(\mathcal{D}) = \mathcal{D}$. Set $\mathcal{D}^{\bar\sigma}$ for the set of fixed points of $\bar\sigma$ in $\mathcal{D}$. Since $\bar\sigma(H) = H$ it easily follows that $H/H \cap K$ can be identified with an open submanifold of $\mathcal{D}^{\bar\sigma}$. The proof is according to the same lines as in Lemma 12. The real "ball" $\mathcal{D}^{\bar\sigma}$ is an interesting object; one can actually show that H acts *transitively* on it.

Example. $\mathfrak{g} = \mathfrak{su}(1,1)$, $\sigma(X) = \bar X, \mathfrak{h} = \mathfrak{so}(1,1)$. $H = SO(1,1)$, $h \in H$ is of the form $\begin{pmatrix} \cosh t & \sinh t \\ \sinh t & \cosh t \end{pmatrix}$, $t \in \mathbb{R}$.

Now $\mathcal{D} = \{z \in \mathbb{C} \,:\, |z| < 1\}$, so $\mathcal{D}^{\bar\sigma} = (-1,1) \subset \mathbb{R}$. This is clearly the same as $H.o = \{\tanh t \,:\, t \in \mathbb{R}\}$.

It is clear that $|k_1(z,z)|^{-1/2} d\nu(z)$, where $d\nu(z)$ is a Euclidean measure on $\mathcal{D}^{\bar\sigma}$, is a H-invariant measure on $\mathcal{D}^{\bar\sigma}$. The proof is along the same line as in section 9.1.

For the spaces we are talking about, see Table I (lower part).

Set

$$\mathfrak{g}^c = \mathfrak{h} + i\mathfrak{q}. \tag{31}$$

This Lie algebra is usually called the Cartan dual of $\mathfrak{g}$ with respect to $\mathfrak{h}$ (or to σ). Clearly $\mathfrak{g}^c$ is the set of fixed points of the anti-involution $\overline{\sigma}$ of $\mathfrak{g}_c$. We refer again to Table 1 for the possible $\mathfrak{g}^c$.

Let G^c be the connected Lie subgroup of G_c with Lie algebra $\mathfrak{g}^c$. G^c is a closed subgroup. Set $\mathfrak{l} = \mathfrak{k}_c^{\overline{\sigma}} = \mathfrak{h} \cap \mathfrak{k} + i\mathfrak{q} \cap \mathfrak{k}$ and let $L = G^c \cap K_c$. The Lie algebra of L is $\mathfrak{l}$. Notice that $\mathfrak{p}_{\pm}^{\overline{\sigma}}$ is contained in $\mathfrak{g}_c$, even in $\mathfrak{h} \cap \mathfrak{p} + i\mathfrak{q} \cap \mathfrak{p}$. Set $\mathfrak{n}_{\pm} = \mathfrak{p}_{\pm}^{\overline{\sigma}}$ and $N_{\pm} = \exp \mathfrak{n}_{\pm}$.

Then $Q^+ = LN_+$ and $Q^- = LN_-$ are standard maximal parabolic subgroups of G^c. In both cases $A = \exp i\mathbb{R}Z_0$ is the split component. Furthermore,

$$H \cap Q^+ = H \cap Q^- = H \cap L^{\sigma} = H \cap K. \tag{32}$$

For the definition of the Berezin distribution we proceed here similarly to Section 9.1 taking the parabolic subgroup Q^- with split component A. The group Q^- is $\sigma\theta$-invariant, where σ and θ have to be understood as restrictions from G_c to G^c. We obtain a formula similar to (30), now taking $h \in H$. Observe that in both cases (9.1 and 9.2) the parabolic subgroup we have chosen is maximal.

9.3. MATRIX-VALUED BEREZIN DISTRIBUTIONS

Return to the context of 9.1 and let us take τ to be an irreducible holomorphic representation of K_c on a finite-dimensional complex vector space with scalar product such that $\tau | K$ is unitary.

Lemma 16. *One has $\tau^*(k) = \tau(\overline{k})$ for $k \in K_c$.*

This follows easily by writing $k = k_0 \exp iX$ with $k_0 \in K$, $X \in \mathfrak{k}$ and using that $\tau | K$ is unitary.

Set δ the restriction of τ to M (see 9.1). Then δ remains irreducible. Notice that J can be taken the identity map. We get:

$$\Psi^P_{\delta,\lambda}(h) = k_1(h \cdot o, h \cdot o)^{\frac{\lambda - n}{2n}} \cdot \delta(m(h))^* \quad (h \in G). \tag{33}$$

A typical example is given in [13]. Let $n = \dim \mathfrak{p}_-$. We then can take for $k \in K_c$ the holomorphic representations

$$(i) \qquad \tau_n(k) = \det_{\mathbb{C}} \mathrm{Ad}(k)_{|\mathfrak{p}_-} \quad \text{(scalar valued)},$$

$$(ii) \qquad \tau_1(k) = \mathrm{Ad}(k)_{|\mathfrak{p}_-} \quad \text{on } \mathfrak{p}_-,$$

$$(iii) \qquad \tau_r(k) = \bigwedge^r \mathrm{Ad}(k)_{|\mathfrak{p}_-} \quad \text{on } \bigwedge^r \mathfrak{p}_-, \quad (1 \le r \le n).$$

The representations τ_n and τ_1 are irreducible, while τ_r certainly is in case $G = SU(1, n, \mathbb{C})$.

Notes. Very often $\Psi^P_{\delta,\lambda}$ is positive-definite in 9.1–9.3. For example, if $\delta = 1$, then $\Psi^P_{\delta,\lambda}$ is positive-definite if λ belongs to the Berezin–Wallach set (see [16]). It gives then, in a natural way, rise to a unitary representation of H, which is usually called a “canonical” representation. For a typical matrix-valued situation, with $G = SU(1, n, \mathbb{C})$ or $G = SO_0(1, n)$, see [8]. The decomposition of the canonical representation into irreducible unitary representations is also obtained in [8].

9.4. COMPLEX HYPERBOLIC SPACES

Here $G = \mathrm{SL}(n, \mathbb{C})$, $H = \mathrm{SU}(p, q)$, $(p + q = n)$, and P is the group of lower triangular block matrices of type $(1, n-1)$. So σ is given by $\sigma(g) = I_{pq}(g^{-1})^* I_{pq}$ where $I_{pq} = \mathrm{diag}(-1, \ldots, -1, +1, \ldots, +1)$ with p minus-signs and q plus-signs. Furthermore, the Cartan involution is $\theta(g) = (g^{-1})^*$.

Observe that σ and θ commute. For $p \in P$, $p = \begin{pmatrix} a & 0 \\ c & b \end{pmatrix}$, with $a \in \mathbb{C}^*$, $b \in \mathrm{GL}(n-1, \mathbb{C})$, $a \det b = 1$, we take

$$\omega_{\delta,\lambda}(p) = |a|^{\lambda - \rho} \left(\frac{a}{|a|} \right)^{\delta}$$

with $\delta \in \mathbb{Z}$, $\rho = n - 1$.

We refer to [25] for more details about the form of $\Psi^P_{\delta,\lambda}$ and its expansion into extremal positive-definite $\mathrm{S}(\mathrm{U}(1) \times \mathrm{U}(p-1, q))$–spherical distributions. See in particular [25], Sections 10, 11, 12.

9.5. COMPACT HERMITIAN SYMMETRIC SPACES

Let us return to 9.1 and use the notation of 9.1.

We now consider in G_c the two commuting involutions σ and θ with σ equal to θ, the Cartan involution. We still take $P = P_-$, which is certainly $\sigma\theta$–stable. Instead of $H = G$ we have in this situation $H = U$, the analytic subgroup of G_c with Lie algebra $\mathfrak{u} = \mathfrak{k} + i\mathfrak{p}$. Observe that $U \cap P_- = K$ again. So $\Psi^P_{\delta,\lambda}$ is a distribution on U. If $\delta = 1$, it is even a distribution on U/K, a compact Hermitian symmetric space.

The decomposition of $\Psi^P_{\delta,\lambda}$ for $\delta = 1$ can be obtained as follows. There are in fact two ways: by direct computation (only known in concrete situations) or by “analytic continuation” of the decomposition of $\Psi^P_{\delta,\lambda}$ for the case G/K (case 9.1). Notice that U/K is a “compact form” of G/K. This second method is due to Neretin (see [27]) and quite recent. As to the first method, an interesting important case is $U = \mathrm{SU}(p+q)$, $K = \mathrm{S}(\mathrm{U}(p) \times \mathrm{U}(q))$, $G = \mathrm{SU}(p, q)$. In this situation the decomposition is computed by D. Pickrell [31].

In Section 10 we shall explain the second method on the same example, but with $p = 1$. Neretin claims that this second method works in general, which we beleive to be true. There is however, up to now, no conceptual proof. Neretin’s proof is “case by case”.

A similar observation as above applies to symmetric spaces of Hermitian type with U replaced by the maximal compact subgroup K^c of G^c (the Cartan dual of G). Then $K^c/K \cap H$ and $H/H \cap K$ have to be related. Of course K^c is the analytic subgroup with Lie algebra $\mathfrak{h} \cap \mathfrak{k} + i\mathfrak{q} \cap \mathfrak{p}$. For P we take Q^- (see 9.2).

An important question is therefore how to relate spherical functions on K^c and H (with respect to $K \cap H$). Set $k_0 = \exp i\frac{\pi}{2} Z_0$ (see 9.2 for notation). Then, according to [30], Lemma 1.5, $Ad(k_0)$ is a linear isomorphism between $\mathfrak{h} \cap \mathfrak{p}$ and $i\mathfrak{q} \cap \mathfrak{p}$ and hence a surjective Lie homomorphism of H onto K^c, leaving $K \cap H$ fixed. In this way spherical functions on K^c correspond naturally with certain spherical functions on H: if φ is a $K \cap H$-spherical function on K^c, then $\varphi \circ Ad(k_0)$ is a spherical function on H.

In Section 10 we shall discuss details on some spaces of Hermitian type, namely:

$$\begin{aligned} &G^c = SL(n+1, \mathbb{R}) \qquad H = SO(1,n) \\ &\qquad\qquad K = S(O(1) \times O(n)) \qquad K^c = SO(n+1, \mathbb{R}) \\ &G^c = SL(n+1, \mathbb{H}) \qquad H = Sp(1,n) \\ &\qquad\qquad K = Sp(1) \times Sp(n) \qquad K^c = Sp(n+1, \mathbb{C}) \end{aligned}$$

Notice that $SL(n+1, \mathbb{H}) \simeq SU^*(2n+2)$ (see also Table I).

In a similar way we can treat $G^c = E_{6(-26)}$, $H = F_{4(-20)}$, $K = SO(9)$, $K^c = F_{4(-52)}$, being the same kind of groups over the octonions.

$G = SO_0(p, q+1)$, $H = SO_0(p,q)$, P

9.6. IS A PARABOLIC ASSOCIATED WITH A CONE

As usual, $SO_0(p,q)$ denotes the identity component of the group of real $(p+q) \times (p+q)$ matrices leaving the quadratic form

$$[x,y] = x_1 y_1 + \cdots x_p y_p - x_{p+1} y_{p+1} - \cdots - x_n y_n \quad (n = p+q)$$

invariant. Set $H = SO_0(p,q)$, and similarly, $G = SO_0(p, q+1)$. H is also the connected component of the fixed point group of the involution σ on G defined by $\sigma(x) = IxI$ with $I = I_{n,1}$. Set $\xi^0 = (1, 0, \ldots, 0, 1) \in \mathbb{R}^{n+1}$ and let P be the stabilizer of the line $\mathbb{R}\xi^0$ in G. Observe that $G\xi^0$ is a cone, so P is called a parabolic subgroup associated with a cone. Let θ denote the usual Cartan involution of G given by $\theta(x) = {}^t x^{-1}$. Then clearly P is $\sigma\theta$-stable.

Let $G' = \overline{N}MAN$ ($P = MAN$), and write $g \in G'$ as $g = \overline{n}\, m\, \underline{a}(g)\, n$ with $\underline{a}(g)$ of the form

$$\begin{pmatrix} \cosh t & 0 & \sinh t \\ 0 & I_{n-1} & 0 \\ \sinh t & 0 & \cosh t \end{pmatrix}.$$

One easily gets $e^t = \frac{1}{2}\big|[I\xi^0,\, g\xi^0]\big|$. We choose δ the trivial representation of M. Then we get for $h \in H$:

$$\Psi^P_{\delta,\lambda}(h) = \left|\frac{[he_1,\, e_1] + 1}{2}\right|^{\lambda-\rho}.$$

with $\rho = \frac{1}{2}(n-1)$.

Actually, $\Psi^P_{\delta,\lambda}$ is defined on

$$X = \{x \in \mathbb{R}^n : [x,x] = 1\} \simeq SO_0(p,q)/SO_0(p-1,q)$$

by

$$\Psi^P_{\delta,\lambda}(x) = \left|\frac{x_1+1}{2}\right|^{\lambda-\rho}.$$

This function is close to the function studied in [23] by Molchanov and can be treated in the same way. $\Psi^P_{\delta,\lambda}$ is however not positive-definite in general.

9.7. ASSOCIATED WITH A CONE

Taking $\sigma = \theta$ in example 9.6 with $G = SO_0(p,q)$ now, and $H = SO(p) \times SO(q)$, but P as before, namely the stabilizer of $\mathbb{R}(1,0,\dots,0,1)$ in $\mathbb{R}^n$ $(n = p+q)$, leads to $(\delta = 1)$:

$$\Psi^P_{\delta,\lambda}(s_1,\, s_2) = (1/2)^{\lambda-\rho}\,|s_{11} + s_{nn}|^{\lambda-\rho}$$

with $s_1 \in S^{p-1}$, $s_2 \in S^{q-1}$. $\Psi^P_{\delta,\lambda}$ is $SO(p) \times SO(q)$ invariant. $\Psi^P_{\delta,\lambda}$ can be decomposed into spherical functions (Gegenbauer polynomials). The integrals we obtain ressemble those in Molchanov's paper [22].

9.8. PARA-HERMITIAN SYMMETRIC SPACES

The following table contains the list of symmetric Lie algebras $(\mathfrak{g}, \mathfrak{h})$ which correspond to para-Hermitian spaces G/H with G simple (cf. [24]). Here $G_{pq}(\mathbf{F})$ denotes the Grassmann manifold of p-planes in $\mathbf{F}^n$ where $\mathbf{F} = \mathbf{R}$ or $\mathbf{H}$; S^{m-1} is the unit sphere in $\mathbf{R}^m$; $P_2(\mathbf{O})$ denotes the octonion projective plane; $n = p+q$.

TABLE II.

$\mathfrak{g}$	$\mathfrak{h}$	G/H
$\mathfrak{sl}(n.\mathbf{R})$	$\mathfrak{sl}(p,\mathbf{R}) + \mathfrak{sl}(q,\mathbf{R}) + \mathbf{R}$	$G_{pq}(\mathbf{R})$
$\mathfrak{su}^*(2n)$	$\mathfrak{su}^*(2p) + \mathfrak{su}^*(2q) + \mathbf{R}$	$G_{pq}(\mathbf{H})$
$\mathfrak{su}(n,n)$	$\mathfrak{sl}(n,\mathbf{C}) + \mathbf{R}$	$U(n)$
$\mathfrak{so}^*(4n)$	$\mathfrak{su}^*(2n) + \mathbf{R}$	$U(2n)/Sp(n)$
$\mathfrak{so}(n,n)$	$\mathfrak{sl}(n,\mathbf{R}) + \mathbf{R}$	$SO(n)$
$\mathfrak{so}(p,q)$	$\mathfrak{so}(p-1,q-1) + \mathbf{R}$	$(S^{p-1} \times S^{q-1})/\mathbf{Z}_2$
$\mathfrak{sp}(n,\mathbf{R})$	$\mathfrak{sl}(n,\mathbf{R}) + \mathbf{R}$	$U(n)/O(n)$
$\mathfrak{sp}(n,n)$	$\mathfrak{su}^*(2n) + \mathbf{R}$	$Sp(n)$
$\mathfrak{e}_{6(6)}$	$\mathfrak{so}(5,5) + \mathbf{R}$	$G_{22}(\mathbf{H})/\mathbf{Z}_2$
$\mathfrak{e}_{6(-26)}$	$\mathfrak{so}(1,9) + \mathbf{R}$	$P_2(\mathbf{O})$
$\mathfrak{e}_{7(7)}$	$\mathfrak{e}_{6(6)} + \mathbf{R}$	$SU(8)/Sp(4) \cdot \mathbf{Z})_2$
$\mathfrak{e}_{7(-25)}$	$\mathfrak{e}_{6(-26)} + \mathbf{R}$	$S^1 \cdot E_6/F_4$

The general approach to Berezin kernels in these cases is described in [24]. It leads to Berezin distributions on G/H. The rank one example

$$G = SL(n,\mathbb{R}),\ \ H = S(GL(1,\mathbb{R}) \times GL(n-1,\mathbb{R})),$$
$$P:\ \text{upper triangular matrices of type } (1, \mathrm{n}-1)$$

is extensively studied in two papers by Molchanov and Van Dijk [9], [10]. Extension of these results to line bundles over G/H is a next step.

10. The new approach in action

We will demonstrate how maximal degenerate representations of $SL(n+1,\mathbb{F})$ unite canonical representations of $SU(1,n,\mathbb{F})$ and $SU(n+1,\mathbb{F})$ for $\mathbb{F}=\mathbb{R}$, $\mathbb{C}$ and $\mathbb{H}$ (the quaternions). For $\mathbb{F}=\mathbb{C}$ we are in the Hermitian symmetric case described in 9.1 (non-compact) and 9.5 (compact), the other cases, where we deal with spaces of Hermitian type are described in 9.2 and again 9.5. How to unite? We therefore start, in order to be absolutely transparant, from the beginning again. We take $\delta=1$, for symplicity. Let d be the dimension of $\mathbb{F}$ over $\mathbb{R}$, so $d=1,2$ or 4 respectively.

10.1. MAXIMAL DEGENERATE REPRESENTATIONS OF $SL(n+1,\mathbb{F})$

10.1.1. *Definition of the representations*
On $\mathbb{F}^{n+1}$, regarded as a right vector space over $\mathbb{F}$, we consider the Hermitian form

$$[x,y]=\overline{y}_0\,x_0-\overline{y}_1\,x_1-\cdots-\overline{y}_n\,x_n.$$

Let $G=SU(1,n,\mathbb{F})$ be the group of $(n+1)\times(n+1)$ matrices with coefficients in $\mathbb{F}$ which preserve this Hermitian form. Set $G_{\mathbb{F}}=SL(n+1,\mathbb{F})$. Denote by $K_{\mathbb{F}}$ the subgroup

$$K_{\mathbb{F}}=S(GL(1,\mathbb{F})\times GL(n,\mathbb{F}))$$

of $G_{\mathbb{F}}$ and set $U=SU(n+1,\mathbb{F}), K=S(U(1,\mathbb{F})\times U(n,\mathbb{F}))$. Let $P^{\pm}$ be the two maximal parabolic subgroups of $G_{\mathbb{F}}$ consisting of upper and lower block matrices respectively:

$$P^{+}:\begin{pmatrix} a & b\\ 0 & c\end{pmatrix},\quad P^{-}:\begin{pmatrix} a & 0\\ b & c\end{pmatrix} \tag{34}$$

with $a\in\mathbb{F}^*$, $\begin{pmatrix} a & 0\\ 0 & c\end{pmatrix}\in K_{\mathbb{F}}$, b a row (column) vector in $\mathbb{F}^n$. For $\mu\in\mathbb{C}$, define the character ω_μ of $P^{\pm}$ by the formula:

$$\omega_\mu(p)=|a|^\mu,$$

where $p\in P^{\pm}$ has one of the forms (34). Consider the representations $\pi_\mu^{\pm}$ of $G_{\mathbb{F}}$ induced from $P^{\pm}$:

$$\pi_\mu^{\pm}=\operatorname{Ind}\omega_{\mp\mu}. \tag{35}$$

Let us describe these representations in the "compact picture". One has the following decompositions:

$$G=UP^{+}=UP^{-}, \tag{36}$$

which we call Iwasawa type decompositions. For the corresponding decompositions $g=up$ of an element $g\in G_{\mathbb{F}}$, the factors p and u are defined up to an element of the subgroup

$$U\cap P^{+}=U\cap P^{-}=U\cap K_{\mathbb{F}}=K.$$

The coset spaces $G_{\mathbb{F}}/P^{\pm}$ can be identified with the coset space U/K. Set

$$S = \{z \in \mathbb{F}^{n+1} \ : \ \|z\|^2 = 1\},$$

which clearly can be identified with $\mathrm{SU}(n+1, \mathbb{F})/\mathrm{SU}(n, \mathbb{F})$ via $u \to ue_0$ ($u \in \mathrm{SU}(n+1, \mathbb{F})$).

Let us denote by $C_0(S)$ the vector space of C^∞-functions φ on S satisfying

$$\varphi(s\lambda) = \varphi(s) \tag{37}$$

for all $\lambda \in \mathbb{F}$ with $|\lambda| = 1$.

$C_0(S)$ can be seen as the representation space of both π_μ^+ and π_μ^-. In fact $\pi_\mu^+ = \pi_\mu^- \circ \tau$ where τ is the Cartan involution of $G_{\mathbb{F}}$: $\tau(g) = (\overline{g}^t)^{-1}$.

The group $G_{\mathbb{F}}$ acts on S; denote by $g \cdot s$ ($g \in G_{\mathbb{F}}$, $s \in S$) the action of g on s:

$$g \cdot s = \frac{g(s)}{\|g(s)\|}. \tag{38}$$

We have for $\varphi \in C_0(S)$:

$$\pi_\mu^-(g)\varphi(s) = \varphi(g^{-1} \cdot s)\, \|g^{-1}(s)\|^\mu. \tag{39}$$

In a similar way we have:

$$\pi_\mu^+(g)\varphi(s) = \varphi(\tau(g^{-1}) \cdot s)\, \|\tau(g^{-1})s\|^\mu. \tag{40}$$

Let $(\ |\ )$ denote the standard inner product on $L^2(S)$,

$$(\varphi|\psi) = \int_S \varphi(s)\, \overline{\psi(s)}\, ds. \tag{41}$$

Here ds is the normalized U-invariant measure on S. This measure ds is transformed by the action of $g \in G_{\mathbb{F}}$ as follows:

$$d\widetilde{s} = \|g(s)\|^{-d(n+1)} ds, \quad \widetilde{s} = g \cdot s. \tag{42}$$

It implies that the Hermitian form (41) is invariant with respect to the pairs

$$(\pi_\mu^-, \pi^-_{-\overline{\mu}-d(n+1)}) \quad \text{and} \quad (\pi_\mu^+, \pi^+_{-\overline{\mu}-d(n+1)}) \tag{43}$$

Therefore, if $\operatorname{Re}\mu = -(n+1)d/2$, then the representations $\pi_\mu^\pm$ are unitarizable, the inner product being (41).

10.1.2. *Intertwining operators and irreducibility*

It turns out that $\pi_\mu^\pm$ is at least irreducible if $\mu \notin \mathbb{Z}$.

Define the operator A_μ on $C_0(S)$ by the formula

$$A_\mu\varphi(s) = \int_S |(s,t)|^{-\mu-d(n+1)} \varphi(t)\, dt. \tag{44}$$

This integral converges absolutely for $\mathrm{Re}\,\mu < -dn - 1$ and can be analytically extended to the whole μ-plane as a mereomorphic function. It is easily checked that A_μ is an intertwining operator between π_μ^+ and $\pi_{\mu'}^-$:

$$A_\mu \, \pi_\mu^\pm(g) = \pi_{\mu'}^\mp(g)\, A_\mu, \; g \in G_{\mathbb{F}}, \tag{45}$$

with $\mu' = -\mu - d(n+1)$.

10.1.3. *Restriction to G*
Consider the diagonal matrix $I = \mathrm{diag}\,\{1, -1, \ldots, -1\}$. Then

$$G = \{g \in G_{\mathbb{F}} \; : \; \overline{g}^t = Ig^{-1}I\}. \tag{46}$$

So the Cartan involution τ of $G_{\mathbb{F}}$ restricted to G is given by $\tau(g) = IgI$ $(g \in G)$. Consequently, π_μ^+ is equivalent to π_μ^- on G: the equivalence is given by $\varphi \to E\varphi$ with

$$E\varphi(s) = \varphi(Is) \tag{47}$$

for $\varphi \in C_0(S)$.

Now consider the action of G on S given by (38). There are 3 orbits, given by

$$[s,s] > 0, \qquad [s,s] = 0 \qquad \text{and} \qquad [s,s] < 0. \tag{48}$$

All three orbits are invariant under $s \to s\lambda$ with $\lambda \in \mathbb{F}$, $|\lambda| = 1$. Call O_1, O_2, O_3 the corresponding G-orbits on $S/\sim$ where $s \sim s'$ if and only if $s = s'\lambda$ for some $\lambda \in \mathbb{F}$, $|\lambda| = 1$. Then we have, with e_i denoting the i-th unit vector in $\mathbb{F}^{n+1}$ $(0 \le i \le n)$:

$$O_1 \simeq G/K \qquad \text{via} \qquad g \to g \cdot e_0, \tag{49}$$

$$O_2 \simeq G/P_{min} \qquad \text{via} \qquad g \to g \cdot (e_0 + e_n), \tag{50}$$

$$O_3 \simeq G/\mathrm{S}(\mathrm{U}(1,n-1) \times \mathrm{U}(1)) \qquad \text{via} \qquad g \to g \cdot e_n. \tag{51}$$

The subgroup P_{min} is a minimal parabolic subgroup of G.

Let φ be a C^∞-function with compact support on $[s,s] \neq 0$, satisfying (37). Set

$$\psi(s) = \varphi(s)\, |[s,s]|^{-\mu/2}. \tag{52}$$

Then ψ satisfies the same condition (37). Moreover,

$$\begin{aligned} \psi(g^{-1} \cdot s) &= \varphi(g^{-1} \cdot s)\, ||g^{-1}(s)||^\mu \, |[s,s]|^{-\mu/2} \\ &= \pi_\mu^-(g)\varphi(s)\, |[s,s]|^{-\mu/2} \end{aligned}$$

So the linear map $\varphi \to \psi$ intertwines the restriction of π_μ^- to G with the left regular representation of G on $\mathcal{D}(G/K)$ and $\mathcal{D}(G/H)$ respectively where $H = \mathrm{S}(\mathrm{U}(1, n-1, \mathbb{F}) \times \mathrm{U}(1, \mathbb{F}))$.

A G-invariant measure on $O_1 \cup O_3$ is given by

$$d\nu(s) = \frac{ds}{|[s,s]|^{\frac{d(n+1)}{2}}}. \tag{53}$$

So, if we provide $\mathcal{D}(S)$ with the inner product on $[s,s] \neq 0$ given by

$$\langle \varphi_1, \varphi_2 \rangle = \int_S \varphi_1(s)\,\overline{\varphi_2(s)}\,|[s,s]|^{-\operatorname{Re}\mu - \frac{d(n+1)}{2}}\,ds, \tag{54}$$

then π_μ^- becomes unitary, if we restrict it to G.

From now on we shall only consider the restriction of π_μ^- to G and call it R_μ. Clearly R_μ is equivalent to $R_{\mu'}$. The intertwinig operator becomes

$$A_\mu \varphi(s) = \int_S |[s,t]|^{-\mu - d(n+1)} \varphi(t)\,dt. \tag{55}$$

Observe that A_μ is defined on $[s,s] > 0$ for all μ, provided φ has compact support in this open set. Then $A_\mu\varphi$ is a C^∞-function on this set (non-necessarily with compact support). On $[s,s] < 0$ one still has to deal with analytic continuation in μ, since convergence of the integral is not garantueed for all μ.

For $\varphi_1, \varphi_2 \in C_0(S)$ and $\mu \in \mathbb{R}$, consider the Hermitian form

$$\langle \varphi_1, A_\mu \varphi_2 \rangle = \int_S \int_S |[s,t]|^{-\mu - d(n+1)} \varphi_1(s)\,\overline{\varphi_2(t)}\,ds\,dt. \tag{56}$$

This form is clearly invariant with respect to R_μ. Applying the linear transformation (40) on the open set $[s,s] \neq 0$, we get the following:

$$\langle \psi_1, \psi_2 \rangle = \int_S \int_S \psi_1(s)\,\overline{\psi_2(t)} \left| \frac{[s,s]\,[t,t]}{[s,t]\,[t,s]} \right|^{\frac{\mu}{2} + \frac{d(n+1)}{2}} d\nu(s)\,d\nu(t). \tag{57}$$

Now restrict to $[s,s] > 0$. Then the kernels in (57), namely

$$B_\lambda(s,t) = \left\{ \frac{[s,s]\,[t,t]}{[s,t]\,[t,s]} \right\}^\lambda \tag{58}$$

($\lambda \in \mathbb{C}$) are defined on $\mathcal{O}_1 \times \mathcal{O}_1$ and coincide with the Berezin kernels on G/K, given in (30). Indeed, if we identify $\mathcal{O}_1$ with the unit ball $\mathcal{D}$ in $\mathbb{F}^n$ via

$$(z_0, z_1, \dots, z_n) \to \left(\frac{z_1}{z_0}, \dots, \frac{z_n}{z_0} \right) \tag{59}$$

then G acts by means of fractional linear transformations on $\mathcal{D}$ and B_λ gets into

$$\widetilde{B}_\lambda(z,w) = \left\{ \frac{(1 - (\|z\|^2)(1 - \|w\|^2)}{[1 - (z,w)][1 - (w,z)]} \right\}^\lambda \tag{60}$$

where $(\ ,\ )$ is the standard scalar product on $\mathbb{F}^n$. For $\Psi_{\delta,\lambda}^{P^-}$ we get

$$\Psi_{\delta,\lambda}^{P^-}(g) = (1 - \|z\|^2)^\lambda \qquad \text{if } z = g \cdot 0. \tag{61}$$

We shall simply write Ψ_λ for $\Psi_{\delta,\lambda}^{P^-}$ since $\delta = 1$ and P^- is fixed. It is known that $\widetilde{B}_\lambda$ is a positive-definite kernel if λ is real and $\lambda > 0$ when $\mathbb{F} = \mathbb{R}$, $\mathbb{C}$ and $\lambda > 1$ when

$\mathbb{F} = \mathbb{H}$ (see [8]). Let for those λ, τ_λ be the unitary representation of G defined by $\widetilde{B}_\lambda$. It is called a *canonical representation* of G, see [36].

10.2. DECOMPOSITION OF THE CANONICAL REPRESENTATIONS

10.2.1. *For* $\lambda > \rho/2$

Decomposition of τ_λ into irreducible unitary representations is equivalent with the expansion of Ψ_λ into positive-definite spherical functions φ_s.

Let $\rho = \frac{1}{2}d(n+1) - 1$. Then according to [8], Section 8, one has:

Theorem 17.

$$\Psi_\lambda = \frac{2^{2\rho-2}\Gamma(\frac{dn}{2})}{\pi^{dn/2+1}} \int_0^\infty a_\lambda(\mu)\,\varphi_{i\mu}\,\frac{d\mu}{|c(i\mu)|^2} \tag{62}$$

in the sense of distributions on G (or $\mathcal{D}$), for $\operatorname{Re}\lambda > \rho/2$, *where*

$$a_\lambda(\mu) = \pi^{\frac{dn}{2}} \frac{\Gamma(\lambda + \frac{i\mu-\rho}{2})\,\Gamma(\lambda + \frac{-i\mu-\rho}{2})}{\Gamma(\lambda)\,\Gamma(\lambda - \frac{d}{2} + 1)} \tag{63}$$

$$c(s) = \Gamma(\frac{dn}{2})\,2^{\rho-s}\,\frac{\Gamma(s)}{\Gamma(\frac{s+\rho}{2})\Gamma(\frac{s+\rho-d+2}{2})} \tag{64}$$

$$\varphi_s(a_t) = {}_2F_1(\frac{s+\rho}{2}, \frac{-s+\rho}{2}; \frac{dn}{2}; -\sinh^2 t) \tag{65}$$

and

$$a_t = \begin{pmatrix} \cosh t & 0 & \sinh t \\ 0 & I & 0 \\ \sinh t & 0 & \cosh t \end{pmatrix}.$$

10.2.2. *For* $\frac{\rho}{2} - k - 1 < \operatorname{Re}\lambda \leq \frac{\rho}{2} - k$

We are going to perform analytic continuation in λ. There are at least two ways to follow, one is described in [7] and the other one in [8]. The result is as follows (see again [7], [8]). Let k be a natural number. For $\frac{\rho}{2} - k - 1 < \operatorname{Re}\lambda < \frac{\rho}{2} - k$, we get $k+1$ extra discrete terms in the decomposition (62), namely

$$\sum_{l=0}^{k} R_l(\lambda)\,\varphi_{\rho-2\lambda-2l} \tag{66}$$

with

$$R_l(\lambda) = \frac{1}{\Gamma(\frac{dn}{2})}(\rho - 2\lambda - 2l)\cdot$$

$$\cdot \frac{\Gamma(1-\lambda)\Gamma(\frac{d}{2}-\lambda)\Gamma(\rho-\lambda-l)\Gamma(\rho-\lambda-l+1-\frac{d}{2})}{l!\Gamma(\frac{d}{2}-\lambda-l)\Gamma(1-\lambda-l)\Gamma(\rho-2\lambda-l+1)}. \tag{67}$$

So we have

Theorem 18. *If* $\frac{\rho}{2} - k - 1 < \operatorname{Re}\lambda < \frac{\rho}{2} - k$ $(k = 0, 1, 2, \dots)$, *then*

$$\Psi_\lambda = \frac{2^{2\rho-2}\Gamma(\frac{dn}{2})}{\pi^{dn/2+1}} \int_0^\infty a_\lambda(\mu)\,\varphi_{i\mu}\,\frac{d\mu}{|c(i\mu)|^2} + \sum_{l=0}^{k} R_l(\lambda)\,\varphi_{\rho-2\lambda-2l} \tag{68}$$

in the sense of distributions on $\mathcal{D}$.

We notice that this formula also holds in $\lambda = \frac{\rho}{2} - k$ by taking the left=right limit for λ to $\frac{\rho}{2} - k$ in $\mathbb{R}$. Moreover, since Ψ_λ is positive-definite for $\lambda > 0$ ($\lambda > 1$ for $\mathbb{F} = \mathbb{H}$), we have $R_l(\lambda) \geq 0$ for such l for which $\rho - 2\lambda - 2l > 0$ (> 1 for $\mathbb{F} = \mathbb{H}$). The spherical functions $\varphi_{\rho-2\lambda-2l}$ belong then to the complementary series of $SU(1, n, \mathbb{F})$.

10.2.3. *Asymptotic behaviour for* $\lambda \to \infty$

We normalize the Berezin distribution Ψ_λ by demanding that the integral of Ψ_λ over $\mathcal{D}$ is equal to one, for λ real and large. Then clearly Ψ_λ tends to the delta-function δ at the origin in $\mathcal{D}$, in the distribution sense, when λ tends to infinity. One has

$$c_\lambda = \int_{\mathcal{D}} \Psi_\lambda(x)dx = \int_{\mathcal{D}} (1 - ||z||^2)^{\lambda - \frac{d}{2}(n+1)} d\mu(z) = \mathrm{vol}(S) \cdot \frac{\Gamma(\frac{dn}{2} + 1)\Gamma(\lambda - \rho)}{\Gamma(\lambda - \frac{d}{2} + 1)}.$$

Here $d\mu(z)$ is the Euclidean measure on $\mathbb{F}^n$ and $(1 - ||z||^2)^{-\frac{d}{2}(n+1)} d\mu(z)$ is a G-invariant measure on $\mathcal{D}$.

We now compute the asymptotic behaviour of $c_\lambda^{-1}\Psi_\lambda$ for $\lambda \to \infty$. We therefore apply Theorem 17. We have

$$c_\lambda^{-1}\, a_\lambda(\mu) = \frac{1}{vol(S)} \frac{\pi^{\frac{dn}{2}}}{\Gamma(\frac{dn}{2} + 1)} \frac{\Gamma(\lambda + \frac{i\mu - \rho}{2})\Gamma(\lambda + \frac{-i\mu - \rho}{2})}{\Gamma(\lambda)\Gamma(\lambda - \rho)}.$$

Applying the well-known asymptotic formula for the quotient for two Gamma functions:

$$\frac{\Gamma(z + a)}{\Gamma(z + b)} = z^{a-b} \left\{ 1 + \frac{1}{2z}(a - b)(a + b - 1) + \mathcal{O}(z^{-2}) \right\}, \quad z \to \infty$$

([14], p. 46, (1)), we get

$$c_\lambda^{-1}\, a_\lambda(\mu) = 1 + \frac{1}{\lambda}\left(\frac{-\mu^2 - \rho^2}{4}\right) + \mathcal{O}\left(\frac{1}{\lambda^2}\right) \quad (\lambda \to \infty). \tag{69}$$

In other words:

$$c_\lambda^{-1}\Psi_\lambda = \delta + \frac{\Box\delta}{4\lambda} + \mathcal{O}(\frac{1}{\lambda^2}) \quad (\lambda \to \infty).$$

where $\Box$ is the Laplace–Beltrami operator on $\mathcal{D}$. This implies that the "correspondence principle" holds (in all cases $d = 1, 2, 4$), see [8]. In particular τ_λ tends to the regular representation of G on $L^2(G/K)$ when λ tends to infinity.

10.2.4. *Berezin kernel of* $SU(n+1,\mathbb{F})$
Restricting $\pi_\mu^\pm$ to $U = SU(n+1,\mathbb{F})$ instead of G, and following the same procedure as in 10.1.1–10.1.3, we get the Berezin kernel for U:

$$B_\lambda^U(s,t) = |(s,t)|^{-2\lambda} \quad (s,t \in S/\sim). \tag{70}$$

The associated function Ψ_λ^U is:

$$\Psi_\lambda^U(u) = \Psi_\lambda^U(s) = |s_0|^{-2\lambda}, \; s = u\cdot e_0, \quad (s \in S/\sim, u \in U). \tag{71}$$

This is indeed a well-defined function for negative real λ. Set

$$\int_{S/\sim} \Psi_\lambda^U(s)ds = c_\lambda^U.$$

It is easily checked that $(c_\lambda^U)^{-1}\Psi_\lambda^U \to \delta$ on $S/\sim$ for $\lambda \to -\infty$ where δ is the delta-function at the origin e_0 of $S/\sim$. Moreover, with the normalization of ds (such that $vol(S) = 1$) of 10.1.1, we have

$$c_\lambda^U = \frac{\Gamma(\frac{d}{2}(n+1))\Gamma(\frac{d}{2}-\lambda)}{\Gamma(\frac{d}{2})\Gamma(-\lambda+\frac{d}{2}(n+1))}. \tag{72}$$

Observe that $S/\sim$ is isomorphic to U/K.

Let us now take λ negative integer. Then clearly B_λ^U is a positive-definite kernel and defines a "canonical" representation τ_λ^U of U.

Notice that in (68) the integral disappears for λ negative integer, so only the finite sum

$$\sum_{l=0}^{k} R_l(\lambda)\,\varphi_{\rho-2\lambda-2l}$$

remains. Here k is such that $\frac{\rho}{2}-k-1 < \lambda \le \frac{\rho}{2}-k$. Moreover, $R_l(\lambda) = 0$ if $\lambda + l \ge 1$, which follows from (67).

For the other l, namely $l \le -\lambda$,

$$\varphi_{\rho-2\lambda-2l}(a_t) = {}_2F_1(\rho-\lambda-l, \lambda+l; \frac{dn}{2}; -\sinh^2 t)$$

is a polynomial in $-\sinh^2 t$.

Hence we get:

$$\Psi_\lambda(a_t) = (\cosh t)^{-2\lambda} = \sum_{l=0}^{-\lambda} R_l(\lambda)\,{}_2F_1(\rho-\lambda-l, \lambda+l; \frac{dn}{2}; -\sinh^2 t)$$

Replacing t by it ($i = \sqrt{-1}$) we obtain:

$$\Psi_\lambda^U(b_t) = \sum_{l=o}^{-\lambda} R_l(\lambda)\,{}_2F_1(\rho-\lambda-l, \lambda+l; \frac{dn}{2}; \sin^2 t) \tag{73}$$

with $b_t = \begin{pmatrix} \cos t & 0 & i\sin t \\ 0 & I & 0 \\ i\sin t & 0 & \cos t \end{pmatrix} \in U$, so we get the decomposition of τ_λ^U (λ negative integer), since ${}_2F_1(\rho-\lambda-l, \lambda+l; \frac{dn}{2}; \sin^2 t)$ is the radial part of a spherical function of U with respect to K. We have also shown that (after normalization of the inner product) τ_λ^U tends to the regular representation of U on $L^2(U/K)$ when $\lambda \to -\infty$, λ integer.

10.2.5. *Asymptotic behaviour for $\lambda \to -\infty$ (λ integer)*
We normalize Ψ_λ^U and determine the asymptotic behaviour of $(c_\lambda^U)^{-1}\Psi_\lambda^U$ for λ integer, $\lambda \to -\infty$ more precisely.

Set $v = \rho - 2\lambda - 2l$ in (67) and set $r_v(\lambda) = R_l(\lambda)$. We have

$$\begin{aligned}\Gamma(\frac{dn}{2})\, r_v(\lambda) &= v \cdot \frac{\Gamma(\frac{v+\rho}{2})\Gamma(\frac{v+\rho+2-d}{2})}{\Gamma(\frac{v-\rho+d}{2})\Gamma(\frac{v-\rho+2}{2})} \cdot \\ &\quad \cdot \frac{\Gamma(1-\lambda)\Gamma(\frac{d}{2}-\lambda)}{\Gamma(\frac{\rho}{2}-\lambda-\frac{v}{2}+1)\Gamma(\frac{\rho}{2}-\lambda+\frac{v}{2}+1)}\end{aligned}$$

for $v = \rho + 2m$ with $m = 0, 1, 2, \ldots$.

We get, applying the asymptotic formula for the quotient of two Gamma functions:

$$\begin{aligned}(c_\lambda^U)^{-1}\, r_v(\lambda) &= v \cdot \frac{\Gamma(\frac{d}{2})\Gamma(\frac{v+\rho}{2})\Gamma(\frac{v+\rho+2-d}{2})}{\Gamma(\rho+1-\frac{d}{2})\Gamma(\rho+1)\Gamma(\frac{v-\rho+d}{2})\Gamma(\frac{v-\rho+2}{2})} \\ &\quad \cdot \{1 + \frac{v^2-\rho^2}{4\lambda} + \mathcal{O}(\frac{1}{\lambda^2})\} \qquad (\lambda \to -\infty,\ \lambda \text{ integer}). \qquad (74)\end{aligned}$$

Observe that (74) implies that again the "correspondence principle" holds. The factor in front of the braces is, for each v, the dimension of a space of harmonic polynomials (see e.g. [5]). This is easily seen, eventually applying Legendre's duplication formula of the Gamma function.

10.2.6. *Remarks*
One has for λ negative integer:

$$\Psi_\lambda^U = \sum_{m=0}^{\infty} r_{\rho+2m}(\lambda)\, \varphi_{\rho+2m}^U \qquad (75)$$

since $r_{\rho+2m}(\lambda) = 0$ for $m > -\lambda$. Here $\varphi_{\rho+2m}^U$ is the spherical function on U with respect to K with radial part ${}_2F_1(\rho+m, -m; \frac{dn}{2}; \sin^2 t)$. The expression $r_{\rho+2m}(\lambda)$ is defined for complex λ and is analytic in λ for $\operatorname{Re}\lambda < 1/2$. It is easily checked that (75) remains true for all complex λ with $\operatorname{Re}\lambda < 1/2$ in L^2-sense, by computing inner products and using formulae from [15]. Neretin applies alternatively Carlson' s theorem ([27], Theorem 9.6) which is very convenient here. Clearly (74) now holds without the restriction "λ integer". Formula (75) is a special case of a result by Pickrell [31].

References

1. van den Ban, E. P. (1988) The principal series for a reductive symmetric space I. H-fixed distribution vectors, *Ann. Sci. Ec. Norm. Sup.*, **21**, 359–412.
2. Berezin, F. A. (1974) Quantization, *Math. USSR Izv.*, **8**, 1109–1165.
3. Berezin, F. A. (1975) Quantization in complex symmetric space, *Math. USSR Izv.*, **9**, 341–379.
4. Berezin, F. A. (1978) Connection between co- and contravariant symbols of operators on the classical complex symmetric spaces, *Dokl. Akad. Nauk. USSR*, **241**, 15–17; English transl.: (1978) *Soviet Math. Dokl.*, **19**, 786–789.
5. Bröcker, T. and tom Dieck, T. (1985) *Representations of Compact Lie Groups*, Graduate Texts in Mathematics, vol. 98, Springer–Verlag, New York.
6. van Dijk, G. (1993) Group representations on spaces of distributions, *Russian J. Math. Phys.*, **2**, 57–68.
7. van Dijk, G. (2001) On canonical representations and Berezin kernels, Preprint Leiden University. (Notes of lectures given at the Workshop on Lie Groups and Lie Algebras, Bedlewo (Polen), 4–15 September 2000).
8. van Dijk, G. and Hille, S. C. (1997) Canonical representations related to hyperbolic spaces, *J. Funct. Anal.*, **147**, 109–139.
9. van Dijk, G. and Molchanov, V. F. (1998) The Berezin form for rank one para-Hermitian symmetric spaces, *J. Math. Pures Appl.*, **77**, 747–799
10. van Dijk, G. and Molchanov, V. F. (1999) Tensor products of maximal degenerate series representations of the group $SL(n, \mathbf{R})$, *J. Math. Pures Appl.*, **78**, 99–119.
11. van Dijk, G. and Pasquale, A. A. (1999) Canonical representations of $Sp(1, n)$ associated with representations of $Sp(1)$, *Comm. Math. Phys.*, **202**, 651–667.
12. van Dijk, G. and Pevzner, M. (2001) Berezin kernels on tube domains, *J. Funct. Anal.*, **181**, 189–208.
13. van Dijk, G. and Pevzner, M. (2000) Matrix-valued Berezin kernels, Report MI 2000-37, University of Leiden.
14. Erdelyi, A. et al. (1953) *Higher Transcendental Functions*, vol. I, McGraw-Hill, New York.
15. Erdelyi, A. et al., (1954) *Tables of integral transforms*, vol. II, McGraw-Hill, New York.
16. Faraut, J. and Koranyi, A. (1994) *Analysis on symmetric cones*, Oxford University Press.
17. Faraut, J. and Olafsson, G., (1995) Causal semisimple symmetric spaces, the geometry and harmonic analysis, in Hofmann, Lawson and Vinberg (Eds.), *Semigroups in Algebra, Geometry and Analysis*, De Gruyter, Berlin.
18. Helgason, S. (1978) *Differential geometry and symmetric spaces*, Academic Press, New York.
19. Helminck, A. G. and Helminck, G. F. (1998) A class of parabolic k-subgroups associated with symmetric K-varieties, *Trans. A.M.S.*, **350**, no. 11, 4669–4693.
20. Hille, S. C. (1999) Thesis, Leiden University.
21. Jorgensen, P. E. T. and Olafsson, G. (2000) Unitary representations and Osterwalder-Schrader duality, in *The Mathematical Legacy of Harish-Chandra. A Celebration of Representation Theory and Harmonic Analysis. PSPM 68.*
22. Molchanov, V. F. (1970) Representations of pseudo-orthogonal groups associated with a cone, *Math. USSR Sbornik*, **10**, 333–347.
23. Molchanov, V. F. (1972) The restriction of a representation of the complementary series of a pseudo-orthogonal group to a pseudo-orthogonal group of lower dimension, *Soviet Math. Dokl.*, **18**, 1493–1497
24. Molchanov, V. F. (1996) Quantization on para-Hermitian symmetric spaces, *Adv. in Math. Sci., AMS Transl. Ser. 2*, **175**, 81–96.
25. Molchanov, V. F. and van Dijk, G. (2001) Berezin forms on line bundles over complex hyperbolic spaces, Report MI 2001-02, Leiden University.

26. Neretin, Yu. A. (2000) Matrix analogs of B-function and Plancherel formula for Berezin kernel representations, *Mat. Sbornik*, **191**, 67–100; preprint version is available via `http://arXiv.org/abs/math/9905045`.
27. Neretin, Yu. A. (2000) Matrix balls, radial analysis of Berezin kernels, and hypergeometric determinants, `http://xxx.lanl.gov/math.RT/0012200v2`.
28. Neretin, Yu. A. (2001) Index hypergeometric transform and imitation of analysis of Berezin kernels on hyperbolic spaces, `http://xxx.lanl.gov/math.CA/0104035`.
29. Ørsted, B. and Zhang, G. (1997) Tensor products of analytic continuations of holomorphic discrete series, *Can. J. Math.*, **49**, 1224–1241.
30. Olafsson, G. (1991) Symmetric spaces of Hermitian type, *Diff. Geom. Appl.*, **1**, 195–233.
31. Pickrell, D. (1987) Measures on infinite-dimensional Grassmann manifolds, *J. Funct. Anal.*, **70**, 323–356.
32. Repka, J. (1979) Tensor products of holomorphic discrete series representations, *Can. J. Math.*, **31**, 836–844.
33. Satake, I. (1980) *Algebraic structures of symmetric domains*, Iwanami Shoten, Publishers and Princeton University press.
34. Schwartz, L. (1964) Sousespaces Hilbertiens d'espaces vectoriels topologiques et noyaux associés (noyaux reproduisants), *J. Analyse Math.*, XIII, 115–256.
35. Unterberger, A. and Upmeier, H. (1994) Berezin transform and invariant differential operators, *Comm. Math. Phys.*, **164**, 563–598.
36. Vershik, A. M., Gel'fand, I .M. and Graev, M. I. (1973) Representations of the group $\mathrm{SL}(2, \mathbf{R})$ where $\mathbf{R}$ is a ring of functions, *Uspekhi Mat. Nauk*, **28** (5), 83–128; English transl. (1973) *Russian Math. Surv.*, **28** (5), 87–132.
37. Wallach, N. R. (1992) *Real Reductive Groups* II, Academic Press, Boston.
38. Zhang, G. (2000) Berezin transform on line bundles over bounded symmetric domains, *J. Lie Theory*, **10**, 111–126.

THETA HYPERGEOMETRIC SERIES

V. P. SPIRIDONOV* (svp@thsun1.jinr.ru)
Bogoliubov Laboratory of Theoretical Physics
Joint Institute for Nuclear Research
Dubna, Moscow Region 141980
Russia

Abstract. We formulate general principles of building hypergeometric type series from the Jacobi theta functions that generalize the plain and basic hypergeometric series. Single and multivariable elliptic hypergeometric series are considered in detail. A characterization theorem for a single variable totally elliptic hypergeometric series is proved.

1. Introduction

This note is a mostly conceptual work reflecting partially the content of a lecture on special functions of hypergeometric type associated with elliptic beta integrals presented by the author at the NATO Advanced Study Institute "Asymptotic Combinatorics with Applications to Mathematical Physics" (St. Petersburg, July 9–23, 2001). It precedes a forthcoming extended and technically more elaborate review [27]. Here we describe general principles of building hypergeometric type series associated with the Jacobi theta functions [21]. Some other essential results of [27] were briefly presented in [25, 26]. We discuss only single variable and multivariable series built out of the Jacobi theta functions though some generalizations based upon the multidimensional Riemann theta functions are possible. Moreover, the main attention will be paid to such series obeying certain ellipticity conditions, i.e. to elliptic hypergeometric series. We start from a description of the Jacobi theta functions properties [1, 32].

Let us take two complex variables p and q lying inside the unit disk, i.e. $|p|, |q| < 1$. The *modular parameters* σ, $\mathrm{Im}(\sigma) > 0$, and τ, $\mathrm{Im}(\tau) > 0$, are introduced through an exponential representation

$$p = e^{2\pi i\tau}, \qquad q = e^{2\pi i\sigma}. \tag{1}$$

* Work supported in part by the Russian Foundation for Basic Research (RFBR) grant no. 00-01-00299.

V. A. Malyshev and A. M. Vershik (eds.),
Asymptotic Combinatorics with Application to Mathematical Physics, 307–327.

Define the p-shifted factorials [14]

$$(a;p)_\infty = \prod_{n=0}^{\infty}(1-ap^n), \qquad (a;p)_s = \frac{(a;p)_\infty}{(ap^s;p)_\infty}.$$

For a positive integer n one has

$$(a;p)_n = (1-a)(1-ap)\cdots(1-ap^{n-1})$$

and

$$(a;p)_{-n} = \frac{1}{(ap^{-n};p)_n}.$$

It is convenient to use the following shorthand notations

$$(a_1,\ldots,a_k;p)_\infty \equiv (a_1;p)_\infty\cdots(a_k;p)_\infty.$$

Let us introduce a Jacobi-type theta function

$$\theta(z;p) = (z, pz^{-1};p)_\infty. \tag{2}$$

It obeys the following simple transformation properties

$$\theta(pz;p) = \theta(z^{-1};p) = -z^{-1}\theta(z;p). \tag{3}$$

One has also $\theta(p^{-1}z;p) = -p^{-1}z\,\theta(z;p)$. Evidently, $\theta(z;p) = 0$ for $z = p^{-M}$, $M \in \mathbb{Z}$, and $\theta(z;0) = 1-z$.

The standard Jacobi's θ_1-function [32] is expressed through $\theta(z;p)$ as follows

$$\begin{aligned}
\theta_1(u;\sigma,\tau) &= -i\sum_{n=-\infty}^{\infty}(-1)^n p^{(2n+1)^2/8} q^{(n+1/2)u} \\
&= 2\sum_{n=0}^{\infty}(-1)^n e^{\pi i\tau(n+1/2)^2}\sin\pi(2n+1)\sigma u \\
&= 2p^{1/8}\sin\pi\sigma u\,(p, pe^{2\pi i\sigma u}, pe^{-2\pi i\sigma u};p)_\infty \\
&= p^{1/8} i q^{-u/2}\,(p;p)_\infty\,\theta(q^u;p), \quad u\in\mathbb{C}.
\end{aligned} \tag{4}$$

We have introduced artificially the second modular parameter σ into the definition of θ_1-function—the variable u will often take integer values and it is convenient to make an appropriate rescaling from the very beginning. Note that other Jacobi theta functions $\theta_{2,3,4}(u)$ can be obtained from $\theta_1(u)$ by a simple shift of the variable u [32], i.e. these functions structurally do not differ much from $\theta_1(u)$.

In the following considerations we shall be employing convenient notations by replacing θ_1-symbol in favor of the elliptic numbers $[u]$ used in [6]:

$$[u] \equiv \theta_1(u) \qquad \text{or} \qquad [u;\sigma,\tau] \equiv \theta_1(u;\sigma,\tau),$$

$$[u_0,\ldots,u_k] \equiv \prod_{m=0}^{k}[u_m].$$

Dependence on σ and τ will be indicated explicitly only if it is necessary. The function $[u]$ is entire, odd $[-u] = -[u]$, and doubly quasiperiodic

$$\begin{aligned} &[u + \sigma^{-1}] = -[u], \\ &[u + \tau\sigma^{-1}] = -e^{-\pi i \tau - 2\pi i \sigma u}[u]. \end{aligned} \tag{5}$$

It is well-known that the theta function $[u]$ can be derived uniquely (up to a constant factor) from the transformation properties (5) and the demand of entireness.

Modular transformations are described by the following $SL(2, \mathbb{Z})$ group action upon the modular parameters σ and τ

$$\tau \to \frac{a\tau + b}{c\tau + d}, \qquad \sigma \to \frac{\sigma}{c\tau + d}, \tag{6}$$

where $a, b, c, d \in \mathbb{Z}$ and $ad - bc = 1$. This group is generated by two simple transformations: $\tau \to \tau + 1$, $\sigma \to \sigma$, and $\tau \to -\tau^{-1}$, $\sigma \to \sigma\tau^{-1}$. In these two cases one has

$$[u; \sigma, \tau + 1] = e^{\pi i/4}[u; \sigma, \tau], \tag{7}$$

$$[u; \sigma/\tau, -1/\tau] = i(-i\tau)^{1/2} e^{\pi i \sigma^2 u^2/\tau}[u; \sigma, \tau], \tag{8}$$

where the square root sign of $(-i\tau)^{1/2}$ is fixed from the condition that the real part of this expression is positive.

2. Theta hypergeometric series ${}_rE_s$ and ${}_rG_s$

Now we are going to introduce formal power series ${}_rE_s$ and ${}_rG_s$ built out from the Jacobi theta functions (we will not consider their convergence properties here). They generalize the plain hypergeometric series ${}_rF_s$ and basic hypergeometric series ${}_r\Phi_s$ together with their bilateral partners ${}_rH_s$ and ${}_r\Psi_s$. The definitions given below follow the general spirit of qualitative (but constructive) definitions of the plain and basic hypergeometric series going back to Pochhammer and Horn [2, 14, 15]. In order to define theta hypergeometric series we use as a key property the quasiperiodicity of Jacobi theta functions (5).

Definition 1. The series $\sum_{n\in\mathbb{N}} c_n$ and $\sum_{n\in\mathbb{Z}} c_n$ are called *theta hypergeometric series of elliptic type* if the function $h(n) = c_{n+1}/c_n$ is a meromorphic doubly quasiperiodic function of n considered as a complex variable. More precisely, for $x \in \mathbb{C}$ the function $h(x)$ should obey the following properties:

$$h(x + \sigma^{-1}) = a h(x), \qquad h(x + \tau\sigma^{-1}) = b e^{2\pi i \sigma \gamma x} h(x), \tag{9}$$

where $\sigma^{-1}, \tau\sigma^{-1}$ are quasiperiods of the theta function $[u]$ (5) and a, b, γ are some complex numbers.

Theorem 2. *Let a meromorphic function $h(n)$ satisfies the properties* (9). *Then it has the following general form in terms of the Jacobi theta functions*

$$h(n) = \frac{[n + u_1, \ldots, n + u_r]}{[n + v_1, \ldots, n + v_s]} q^{\beta n} y, \tag{10}$$

where r,s are arbitrary non-negative integers and $u_1,\ldots,u_r,v_1,\ldots,v_s,\beta,y$ are arbitrary complex parameters restricted by the condition of non-singularity of $h(n)$ and related to the quasiperiodicity multipliers a,b,γ as follows:

$$a=(-1)^{r-s}e^{2\pi i\beta},\qquad \gamma=s-r,$$
$$b=(-1)^{r-s}e^{\pi i\tau(s-r+2\beta)}e^{2\pi i\sigma\left(\sum_{m=1}^{s}v_m-\sum_{m=1}^{r}u_m\right)}. \tag{11}$$

Proof. Let us tile the complex plane of n by parallelograms whose edges are formed by the theta function quasiperiods σ^{-1} and $\tau\sigma^{-1}$. Because the multipliers in the quasiperiodicity conditions (9) are entire functions of n without zeros, the meromorphic function $h(n)$ has the same finite number of zeros and poles in each parallelogram on the plane. Let us denote as $-u_1,\ldots,-u_r$ the zeros of $h(n)$ in one of such parallelograms and as $-v_1,\ldots,-v_s$ its poles. For simplicity we assume that these zeros and poles are simple— for creating multiple zeros or poles it is sufficient to set some of the numbers u_m or v_m to be equal to each other.

Let us represent the ratio of theta hypergeometric series coefficients as follows: $c_{n+1}/c_n=h(n)g(n)$, where $h(n)$ has the form (10) with some unfixed parameter β. Since all the zeros and poles of c_{n+1}/c_n are sitting in $h(n)$, the function $g(n)$ must be an entire function without zeros satisfying the constraints $g(n+\sigma^{-1})=a'g(n)$, $g(n+\tau\sigma^{-1})=b'e^{2\pi i\sigma\gamma' n}g(n)$ for some complex numbers a',b',γ'. However, the only function satisfying such demands is the exponential $q^{\beta' n}$, where β' is a free parameter. Since a factor of such type is already present in (10) we may set $g(n)=1$ and this proves that the most general c_{n+1}/c_n has the form (10) (note that y is just an arbitrary proportionality constant). Direct application of the properties (5) yields the connection between multipliers a,b,γ and the parameters u_m, $m=1,\ldots,r$, v_k, $k=1,\ldots,s$, β as stated in (11). □

Resolving the first order recurrence relation for the series coefficients c_n and normalizing $c_0=1$ we get the following explicit "additive" expression for the general theta hypergeometric series of elliptic type

$$\sum_{n\in\mathbb{N}\,\text{or}\,\mathbb{Z}}\frac{[u_1,\ldots,u_r]_n}{[v_1,\ldots,v_s]_n}\,q^{\beta n(n-1)/2}y^n, \tag{12}$$

where we used the elliptic shifted factorials defined for $n\in\mathbb{N}$ as follows:

$$[u_1,\ldots,u_k]_{\pm n}=\prod_{m=1}^{k}[u_m]_{\pm n},$$
$$[u]_n=[u][u+1]\cdots[u+n-1],\qquad [u]_{-n}=\frac{1}{[u-n]_n}. \tag{13}$$

In order to simplify the trigonometric degeneration limit $\mathrm{Im}(\tau)\to+\infty$ (or $p\to 0$) in the series (12) we renormalize y and introduce as a main series argument another variable z:

$$y\equiv(ip^{1/8})^{s-r}q^{(u_1+\ldots+u_r-v_1-\ldots-v_s)/2}z.$$

Let us replace the parameter β by another parameter α as well through the relation $\beta \equiv \alpha + (r-s)/2$. Then, we can rewrite the function (10) in the following "multiplicative" form using the functions $\theta(tq^n;p)$:

$$h(n) = \frac{\theta(t_1q^n, \ldots, t_rq^n; p)}{\theta(w_1q^n, \ldots, w_sq^n; p)} q^{\alpha n} z, \tag{14}$$

where $t_m = q^{u_m}$, $m = 1, \ldots, r$, $w_k = q^{v_k}$, $k = 1, \ldots, s$, and the following shorthand notations are employed:

$$\theta(t_1, \ldots, t_k; p) = \prod_{m=1}^{k} \theta(t_m; p).$$

Now we are in a position to introduce the unilateral theta hypergeometric series ${}_rE_s$. In its definition we follow the standard plain and basic hypergeometric series conventions. Namely, in the expression (14) we replace s by $s+1$ and set $u_r \equiv u_0$ and $v_{s+1} = 1$. This does not restrict generality of consideration since one can remove such a constraint by fixing one of the numerator parameters u_m to be equal to 1. Then we fix

$${}_rE_s\left(\begin{matrix} t_0, \ldots, t_{r-1} \\ w_1, \ldots, w_s \end{matrix}; q, p; \alpha, z\right) = \sum_{n=0}^{\infty} \frac{\theta(t_0, t_1, \ldots, t_{r-1}; p; q)_n}{\theta(q, w_1, \ldots, w_s; p; q)_n} q^{\alpha n(n-1)/2} z^n, \tag{15}$$

where we have introduced new notations for the elliptic shifted factorials

$$\theta(t; p; q)_n = \prod_{m=0}^{n-1} \theta(tq^m; p)$$

and

$$\theta(t_0, \ldots, t_k; p; q)_n = \prod_{m=0}^{k} \theta(t_m; p; q)_n.$$

We draw attention to the particular ordering of q, p used in the notations for theta hypergeometric series (in the previous papers we were ordering q after p which does not match with the ordering in terms of modular parameters σ, τ in $[u; \sigma, \tau]$).

An important fact is that theta hypergeometric series do not admit confluence limits. Indeed, because of the quasiperiodicity of theta functions the limits of parameters $t_m, w_m \to 0$ or $t_m, w_m \to \infty$ are not well defined and it is not possible to pass in this way from ${}_rE_s$-series to similar series with smaller values of indices r and s.

For the bilateral theta hypergeometric series we introduce different notations:

$${}_rG_s\left(\begin{matrix} t_1, \ldots, t_r \\ w_1, \ldots, w_s \end{matrix}; q, p; \alpha, z\right) = \sum_{n=-\infty}^{\infty} \frac{\theta(t_1, \ldots, t_r; p; q)_n}{\theta(w_1, \ldots, w_s; p; q)_n} q^{\alpha n(n-1)/2} z^n. \tag{16}$$

This expression was derived with the help of (14) without any changes. The elliptic shifted factorials for negative indices are defined in the following way:

$$\theta(t; p; q)_{-n} = \frac{1}{\theta(tq^{-n}; p; q)_n}, \quad n \in \mathbb{N}.$$

Due to the property $\theta(q;p;q)_{-n} = 0$ (or $[1]_{-n} = 0$) for $n > 0$, the choice $t_{s+1} = q$ (or $v_{s+1} = 1$) in the ${}_rG_{s+1}$ series leads to its termination from one side. After denoting $t_r \equiv t_0$ (or $u_r \equiv u_0$) one gets in this way the general ${}_rE_s$-series. Since the bilateral series are more general than the unilateral ones, it is sufficient to prove key properties of theta hypergeometric series in the bilateral case without further specification to the unilateral one.

Consider the limit $\mathrm{Im}(\tau) \to +\infty$ or $p \to 0$. In a straightforward manner one gets

$$\lim_{p\to 0} {}_rE_s = {}_r\Phi_s \left(\begin{matrix} t_0, t_1, \ldots, t_{r-1} \\ w_1, \ldots, w_s \end{matrix} ; q; \alpha, z \right) = \sum_{n=0}^{\infty} \frac{(t_0, t_1, \ldots, t_{r-1}; q)_n}{(q, w_1, \ldots, w_s; q)_n} q^{\alpha n(n-1)/2} z^n. \tag{17}$$

This basic hypergeometric series is different from the standard one by the presence of an additional parameter α. The definition of ${}_r\Phi_s$ series suggested in [23] uses $\alpha = 0$. The definition given in [14] looks as follows

$${}_r\Phi_s = \sum_{n=0}^{\infty} \frac{(t_0, t_1, \ldots, t_{r-1}; q)_n}{(q, w_1, \ldots, w_s; q)_n} \left((-1)^n q^{n(n-1)/2}\right)^{s+1-r} z^n, \tag{18}$$

which matches with (17) for $\alpha = s+1-r$ after the replacement of z by $(-1)^{s+1-r}z$. Actually, the $\alpha = 0$ and $\alpha = s+1-r$ choices are related to each other through the inversion transformation $q \to q^{-1}$ with the subsequent redefinition of parameters t_m, w_m, z. One of the characterizations of the basic hypergeometric series $\sum_n c_n$ consists in the demand for c_{n+1}/c_n to be a general rational function of q^n which is satisfied by (17) only for integer α. In order to get in the limit $p \to 0$ the standard q-hypergeometric series we fix $\alpha = 0$. It is not clear at the moment whether this choice is the most "natural" one or it does not play a fundamental role—this question can be answered only after the discovery of good applications for the series ${}_rE_s$ in pure mathematical or mathematical physics problems. From the point of view of elliptic beta integrals [25, 27] this is the most natural choice indeed.

In the bilateral case we fix $\alpha = 0$ as well, so that in the $p \to 0$ limit the ${}_rG_s$ series are reduced to the general ${}_r\Psi_s$-series:

$${}_r\Psi_s \left(\begin{matrix} t_1, \ldots, t_r \\ w_1, \ldots, w_s \end{matrix} ; q; z \right) = \sum_{n=-\infty}^{\infty} \frac{(t_1, \ldots, t_r; q)_n}{(w_1, \ldots, w_s; q)_n} z^n. \tag{19}$$

Definition 3. The series ${}_{r+1}E_r$ and ${}_rG_r$ are called *balanced* if their parameters satisfy the constraints, in the additive form,

$$u_0 + \ldots + u_r = 1 + v_1 + \ldots + v_r \tag{20}$$

and

$$u_1 + \ldots + u_r = v_1 + \ldots + v_r \tag{21}$$

respectively. In the multiplicative form these restrictions look as follows: $\prod_{m=0}^r t_m = q \prod_{k=1}^r w_k$ and $\prod_{m=1}^r t_m = \prod_{k=1}^r w_k$ respectively.

Remark 4. In the limit $p \to 0$ the series ${}_{r+1}E_r$ goes to ${}_{r+1}\Phi_r$ provided the parameters u_m (or t_m), $m = 0, \dots, r$, and v_k (or w_k), $k = 1, \dots, r$, *remain fixed.* Then our condition of balancing does not coincide with the one given in [14], where ${}_{r+1}\Phi_r$ is called balanced provided $q \prod_{m=0}^{r} t_m = \prod_{k=1}^{r} w_k$ (simultaneously one usually assumes also that $z = q$, but we drop this requirement). A discrepancy in these definitions will be resolved after imposing some additional constraints upon the series parameters (see the very-well-poisedness condition below).

3. Elliptic hypergeometric series

From the author's point of view the following definition plays a fundamental role for the whole theory of hypergeometric type series since it explains origins of some known peculiarities of the plain and basic hypergeometric series.

Definition 5. The series $\sum_{n\in\mathbb{N}} c_n$ and $\sum_{n\in\mathbb{Z}} c_n$ are called *elliptic hypergeometric series* if $h(n) = c_{n+1}/c_n$ is an elliptic function of the argument n which is considered as a complex variable, i.e. $h(x)$ is a meromorphic double periodic function of $x \in \mathbb{C}$.

Theorem 6. *Let σ^{-1} and $\tau\sigma^{-1}$ be two periods of the elliptic function $h(x)$, i.e. $h(x + \sigma^{-1}) = h(x)$ and $h(x + \tau\sigma^{-1}) = h(x)$. Let $r + 1$ be the order of the elliptic function $h(x)$, i.e. the number of its poles (or zeros) in the parallelogram of periods. Then the unilateral (or bilateral) elliptic hypergeometric series coincides with the balanced theta hypergeometric series ${}_{r+1}E_r$ (or ${}_{r+1}G_{r+1}$).*

Proof. It is well known that any elliptic function $h(x)$, $x \in \mathbb{C}$, of the order $r+1$ with the periods σ^{-1} and $\tau\sigma^{-1}$ can be written as a ratio of θ_1-functions as follows [32]:

$$h(x) = z \prod_{m=0}^{r} \frac{[x + \alpha_m; \sigma, \tau]}{[x + \beta_m; \sigma, \tau]}, \tag{22}$$

where the zeros $\alpha_0, \dots, \alpha_r$ and the poles $\beta_0, \dots, \beta_r$ satisfy the following constraint:

$$\sum_{m=0}^{r} \alpha_m = \sum_{m=0}^{r} \beta_m. \tag{23}$$

Now the identification of the unilateral elliptic hypergeometric series with the balanced ${}_{r+1}E_r$-series is evident. One just has to shift $x \to x - \beta_0 + 1$, set $x \in \mathbb{N}$, denote $u_m = \alpha_m - \beta_0 + 1$, $v_m = \beta_m - \beta_0 + 1$, and resolve the recurrence relation $c_{n+1} = h(n)c_n$. After this, the condition (23) becomes the balancing condition for the ${}_{r+1}E_r$ series. A similar situation takes place, evidently, in the bilateral series case ${}_{r+1}G_{r+1}$, when α_m and β_m just coincide with u_m and v_m respectively.

Note that because of the balancing condition (23) the function (22) can be rewritten as a simple ratio of $\theta(t;p)$-functions:

$$h(x) = z \prod_{m=0}^{r} \frac{\theta(t_m q^x; p)}{\theta(w_m q^x; p)},$$

where $t_m = q^{\alpha_m}$ and $w_m = q^{\beta_m}$. □

Definition 7. Theta hypergeometric series of elliptic type are called *modular hypergeometric series* if they are invariant with respect to the $SL(2,\mathbb{Z})$ group action (6).

Consider what kind of constraints upon the parameters of ${}_rE_s$ and ${}_rG_s$ series one has to impose in order to get the modular hypergeometric series. Evidently, it is sufficient to establish modularity of the function $h(n) = c_{n+1}/c_n$. From its explicit form (10) and the transformation laws (7) and (8) it is easy to see that in the unilateral case one must have

$$\sum_{m=0}^{r-1}(x+u_m)^2 = (x+1)^2 + \sum_{m=0}^{s}(x+v_m)^2,$$

which is possible only if a) $s = r-1$, b) the parameters satisfy the balancing condition (20), and c) the following constraint is valid:

$$u_0^2 + \ldots + u_{r-1}^2 = 1 + v_1^2 + \ldots + v_{r-1}^2. \tag{24}$$

Under these conditions the ${}_rE_{r-1}$-series becomes modular invariant. Let us note that modularity of theta hypergeometric series assumes their ellipticity. The opposite is not correct, but a more strong demand of ellipticity, to be formulated below, automatically leads to modular invariance. Modular hypergeometric series represent particular examples of Jacobi modular functions in the sense of Eichler and Zagier [12].

In the bilateral case one must have $r = s$, the balancing condition (21), and the constraint

$$u_1^2 + \ldots + u_r^2 = v_1^2 + \ldots + v_r^2 \tag{25}$$

for the ${}_rG_r$-series to be modular invariant.

Definition 8. The theta hypergeometric series ${}_{r+1}E_r$ is called *well-poised* if its parameters satisfy the following constraints

$$u_0 + 1 = u_1 + v_1 = \ldots = u_r + v_r \tag{26}$$

in the additive form or

$$qt_0 = t_1 w_1 = \ldots = t_r w_r \tag{27}$$

in the multiplicative form. Similarly, the series ${}_rG_r$ is called well-poised if $u_1 + v_1 = \ldots = u_r + v_r$ or $t_1 w_1 = \ldots = t_r w_r$.

This definition of well-poised series matches with the one used in the theory of plain and basic hypergeometric series [14]. Note that it does not imply the balancing condition.

Definition 9. The series ${}_{r+1}E_r$ is called *very-well-poised* if, in addition to the constraints (26) or (27), one imposes the restrictions

$$\begin{aligned} u_{r-3} &= \frac{1}{2}u_0 + 1, \quad u_{r-2} = \frac{1}{2}u_0 + 1 - \frac{1}{2\sigma}, \\ u_{r-1} &= \frac{1}{2}u_0 + 1 - \frac{\tau}{2\sigma}, \quad u_r = \frac{1}{2}u_0 + 1 + \frac{1+\tau}{2\sigma}, \end{aligned} \tag{28}$$

or, in the multiplicative form,

$$t_{r-3} = t_0^{1/2} q, \quad t_{r-2} = -t_0^{1/2} q,$$
$$t_{r-1} = t_0^{1/2} q p^{-1/2}, \quad t_r = -t_0^{1/2} q p^{1/2}. \tag{29}$$

Let us derive a simplified form of the very-well-poised series. First, we notice that

$$\theta(zp^{-1/2}; p) = -zp^{-1/2}\theta(zp^{1/2}; p)$$

and

$$\theta(z, -z, zp^{1/2}, -zp^{1/2}; p) = \theta(z^2; p).$$

After application of these relations, one can find that

$$\frac{\theta(t_{r-3}, \ldots, t_r; p; q)_n}{\theta(qt_0/t_{r-3}, \ldots, qt_0/t_r; p; q)_n} = \frac{\theta(t_0 q^{2n}; p)}{\theta(t_0; p)} (-q)^n.$$

As a result, one gets

$$\begin{aligned} &{}_{r+1}E_r \left(\begin{matrix} t_0, t_1, \ldots, t_{r-4}, qt_0^{1/2}, -qt_0^{1/2}, qp^{-1/2}t_0^{1/2}, -qp^{1/2}t_0^{1/2} \\ qt_0/t_1, \ldots, qt_0/t_{r-4}, t_0^{1/2}, -t_0^{1/2}, p^{1/2}t_0^{1/2}, -p^{-1/2}t_0^{1/2} \end{matrix} ; q, p; z \right) \\ &\qquad = \sum_{n=0}^{\infty} \frac{\theta(t_0 q^{2n}; p)}{\theta(t_0; p)} \prod_{m=0}^{r-4} \frac{\theta(t_m; p; q)_n}{\theta(qt_0/t_m; p; q)_n} (-qz)^n. \end{aligned} \tag{30}$$

For convenience we introduce separate notations for the very-well-poised series, since they contain an essentially smaller number of parameters than the general theta hypergeometric series ${}_{r+1}E_r$. For this we replace z by $-z$ and all the parameters t_m, $m = 0, \ldots, r-4$, by $t_0 t_m$ (in particular, this replaces t_0 by t_0^2). Then we write

$${}_{r+1}E_r(t_0; t_1, \ldots, t_{r-4}; q, p; z) \equiv \sum_{n=0}^{\infty} \frac{\theta(t_0^2 q^{2n}; p)}{\theta(t_0^2; p)} \prod_{m=0}^{r-4} \frac{\theta(t_0 t_m; p; q)_n}{\theta(qt_0 t_m^{-1}; p; q)_n} (qz)^n. \tag{31}$$

In terms of the elliptic numbers this series takes the form:

$$\begin{aligned} &{}_{r+1}E_r(u_0; u_1, \ldots, u_{r-4}; \sigma, \tau; z) \\ &\qquad \equiv \sum_{n=0}^{\infty} \frac{[2u_0 + 2n]}{[2u_0]} \prod_{m=0}^{r-4} \frac{[u_0 + u_m]_n}{[u_0 + 1 - u_m]_n} z^n q^{n(\sum_{m=0}^{r-4} u_m - (r-7)/2)}. \end{aligned} \tag{32}$$

We use the same symbol ${}_{r+1}E_r$ in (31) and (32) since these series can be easily distinguished from the general ${}_{r+1}E_r$ series by the number of free parameters.

For $p = 0$ theta hypergeometric series (31) are reduced to the very-well-poised basic hypergeometric series:

$${}_{r-1}\Phi_{r-2}(t_0; t_1, \ldots, t_{r-4}; q; qz) = \sum_{n=0}^{\infty} \frac{1 - t_0^2 q^{2n}}{1 - t_0^2} \prod_{m=0}^{r-4} \frac{(t_0 t_m; q)_n}{(qt_0 t_m^{-1}; q)_n} (qz)^n,$$

which are different from the corresponding partners in [14] by the replacement of z by qz and t_m by $t_0 t_m$ (a standard notation for these series would be ${}_{r-1}W_{r-2}$ but we are not using it here).

Remind that the balancing condition is not involved into the definition of the very-well-poised theta hypergeometric series. Imposing the corresponding constraint

$$\sum_{m=0}^{r}(u_0+u_m)=1+\sum_{m=1}^{r}(u_0+1-u_m)$$

upon (32) we get

$$\sum_{m=0}^{r-4}u_m=\frac{r-7}{2}.$$

In the multiplicative form this condition takes the form

$$\prod_{m=0}^{r-4}t_m=q^{(r-7)/2}.$$

But this is precisely the balancing condition for the very-well-poised series appearing in the theory of basic hypergeometric series [14]. Thus for the very-well-poised series there is no discrepancy in the definitions of balancing condition given in [14] and in this paper. This happens because the constraints (28) taken separately are not well defined in the limit $\mathrm{Im}(\tau)\to+\infty$. Note that for the balanced series an extra factor standing in (32) to the right of z^n disappears. Summarizing this consideration we conclude that a very natural condition of ellipticity of the function $h(n)=c_{n+1}/c_n$ in the theta hypergeometric series provides a substantial meaning to the (innatural) balancing condition for the standard basic hypergeometric series.

If we impose balancing condition in the multiplicative form then there appears an ambiguity. Indeed, substituting into the condition

$$\prod_{m=0}^{r}(t_0t_m)=q\prod_{m=1}^{r}(qt_0/t_m)$$

the constraints $t_{r-3}=q$, $t_{r-2}=-q$, $t_{r-1}=qp^{-1/2}$, $t_r=-qp^{1/2}$ (these are the restrictions (29) after the shift $t_m\to t_0t_m$) we get $\prod_{m=0}^{r-4}t_m^2=q^{r-7}$ which yields $\prod_{m=0}^{r-4}t_m=\pm q^{(r-7)/2}$ and it is known that only the plus sign corresponds to the correct balancing condition for odd r (the even r cases remain ambiguous, but even r do not appear in known examples of summation formulae of basic hypergeometric series).

In the same way, the bilateral theta hypergeometric series ${}_rG_r$ are called very-well-poised if the constraints (28) or (29) are satisfied, where u_0 or $t_0=q^{u_0}$ is a free parameter. Following the unilateral series case, we replace z by $-z$, shift the parameters $t_m\to t_0t_m$, $m=0,\dots,r-4$, and introduce the following shorthand notations for the simplified form of these series:

$${}_rG_r(t_0;t_1,\dots,t_{r-4};q,p;z)=\sum_{n=-\infty}^{\infty}\frac{\theta(t_0^2q^{2n};p)}{\theta(t_0^2;p)}\prod_{m=1}^{r-4}\frac{\theta(t_0t_m;p;q)_n}{\theta(qt_0t_m^{-1};p;q)_n}(qz)^n \qquad (33)$$

or in terms of the elliptic numbers

$$
{}_rG_r(u_0; u_1, \ldots, u_{r-4}; \sigma, \tau; z)
= \sum_{n=-\infty}^{\infty} \frac{[2u_0+2n]}{[2u_0]} \prod_{m=1}^{r-4} \frac{[u_0+u_m]_n}{[u_0+1-u_m]_n} z^n q^{n(\sum_{m=1}^{r-4} u_m - (r-8)/2)}. \quad (34)
$$

Repeating considerations for the bilateral series we find the following compact form for the balancing condition for the ${}_rG_r$-series (33):

$$
\sum_{m=1}^{r-4} u_m = \frac{r-8}{2}, \quad \text{or} \quad \prod_{m=1}^{r-4} t_m = q^{(r-8)/2}.
$$

Under the constraint $u_{r-3} = u_0$ (or $t_{r-3} = t_0$) the ${}_{r+1}G_{r+1}$-series is converted into the ${}_{r+1}E_r$ series. A general connection between the very-well-poised series of E and G types looks as follows:

$$
\begin{aligned}
{}_rG_r(t_0; t_1, \ldots, t_{r-4}; q, p; z) &= {}_{r+2}E_{r+1}(t_0; t_1, \ldots, t_{r-4}, qt_0^{-1}; q, p; z) \\
&\quad + \frac{q^{r-7}}{z\prod_{m=1}^{r-4} t_m^2} \frac{\theta(t_0^{-2}q^2; p)}{\theta(t_0^{-2}; p)} \prod_{m=1}^{r-4} \frac{\theta(t_m t_0^{-1}; p)}{\theta(q t_0^{-1} t_m^{-1}; p)} \\
&\quad \times {}_{r+2}E_{r+1}\left(\frac{q}{t_0}; t_1, \ldots, t_{r-4}, t_0; q, p; \frac{q^{r-8}}{z\prod_{m=1}^{r-4} t_m^2}\right). \quad (35)
\end{aligned}
$$

Remark 10. Within our classifications, an elliptic generalization of basic hypergeometric series ${}_{r+1}\Phi_r$ introduced by Frenkel and Turaev in [13] coincides with the very-well-poised balanced theta hypergeometric series ${}_{r+1}E_r$ of the unit argument $z = 1$. Such series have their origins in elliptic solutions of the Yang–Baxter equation [5, 3, 6, 13] and biorthogonal rational functions with self-similar spectral properties [28–30, 26].

Definition 11. The series $\sum_{n\in\mathbb{N}} c_n$ and $\sum_{n\in\mathbb{Z}} c_n$ are called *totally elliptic* hypergeometric series if $h(n) = c_{n+1}/c_n$ is an elliptic function of *all free parameters* entering it (except of the parameter z by which one can always multiply $h(n)$) with equal periods of double periodicity.

Theorem 12. *The most general (in the sense of a maximal number of independent free parameters) totally elliptic theta hypergeometric series coincide with the well-poised balanced theta hypergeometric series ${}_rE_{r-1}$ (in the unilateral case) and ${}_rG_r$ (in the bilateral case) for $r > 2$. Totally elliptic series are automatically modular invariant.*

Proof. It is sufficient to prove this theorem for the bilateral series since the unilateral series can be obtained afterwards by a simple reduction. Ellipticity in n leads to $h(n)$ of the form

$$
h(n) = \frac{[n+u_1, \ldots, n+u_r]}{[n+v_1, \ldots, n+v_r]} z,
$$

with the free parameters $u_1, \ldots, u_r$ and $v_1, \ldots, v_r$ satisfying the balancing condition $u_1+\ldots+u_r = v_1+\ldots+v_r$. From such a representation it is evident that there is a freedom in the shift of parameters by an arbitrary constant: $u_m \to u_m + u_0$, $v_m \to v_m + u_0$, $m = 1, \ldots, r$, which does not spoil the balancing condition.

Let us determine now the maximal possible number of independent variables in the totally elliptic hypergeometric series. In general one can denote as a_l, $l = 1, \ldots, L$, a set of free parameters of the elliptic hypergeometric series in which the series is doubly periodic with some periods. Then $u_m = \sum_{k=1}^{L} \alpha_{mk} a_k + \beta_m$ and $v_m = \sum_{k=1}^{L} \gamma_{mk} a_k + \delta_m$ are some linear combinations of a_l with integer coefficients α_{mk}, γ_{mk}. However, because of the possibilities to change variables we can take a number of u_m starting, say, from u_1 and v_m as $a_1, \ldots, a_L$ and demand the double periodicity in these parameters themselves. Since the minimal order of elliptic function is equal to 2, the function $h(n)$ should have at least two zeros and two poles (or one double zero or pole) in u_1. Double zeros and poles ask for additional constraints, i.e. to a reduction of the number of free parameters, and we discard such a possibility.

Let us assume that u_r depends linearly on u_1 and suppose that $u_1, \ldots, u_{r-1}$ are independent variables. Then it is evident that all denominator parameters v_m, $m = 1, \ldots, r$, cannot contain additional independent variables. Indeed, if it would be so, then, inevitably, this parameter should show up at least in one θ-function in the numerator, which cannot happen by the assumption.

So, $L = r - 1$ is the maximal possible number of independent variables in the totally elliptic hypergeometric series and u_r together with v_m, $m = 1, \ldots, r$, depend linearly on u_k, $k = 1, \ldots, r-1$. Because of the permutational invariance in the latter variables one must have $u_r = \alpha \sum_{k=1}^{r-1} u_k + \beta$, where α, β are some numerical coefficients to be determined (evidently, α must be an integer). As to the choice of v_m the unique option guaranteeing permutational invariance of the product $\prod_{m=1}^{r} [x + v_m]$ in $u_1, \ldots, u_{r-1}$ is the following one

$$v_m = \gamma \sum_{k=1}^{r-1} u_k + \delta u_m + \rho, \quad m = 1, \ldots, r-1,$$

and $v_r = \mu \sum_{k=1}^{r-1} u_k + \nu$, where $\gamma, \delta, \rho, \mu, \nu$ are some numerical parameters (γ, δ, μ must be integers). This is the most general choice of v_m since all other permutationally invariant combinations of u_m require products of more than r theta functions. Substitution of the taken ansatz into the balancing condition yields $1 + \alpha = (r-1)\gamma + \delta + \mu$ and $\beta = (r-1)\rho + \nu$, which guarantees invariance of $h(x)$ under the shift $u_k \to u_k + \sigma^{-1}$ and cancels the sign factor emerging from the shift $u_k \to u_k + \tau\sigma^{-1}$. A bit cumbersome but technically straightforward analysis of the condition of cancellation of the factors of the form $e^{-2\pi i \sigma u}$ yields the equations

$$\delta^2 = 1, \qquad \alpha^2 = (r-1)\gamma^2 + 2\gamma\delta + \mu^2$$

and

$$\alpha\beta = (r-1)\gamma\rho + \rho\delta + \mu\nu.$$

The constraint generated by the cancellation of the factors of the form $e^{-\pi i\tau}$ appears to be irrelevant and it will not be indicated.

Let $\delta = 1$. Then two equations upon the coefficients α, γ, μ yield that either $\gamma = 0$ or $\gamma(r-1)(r-2)/2 + \mu(r-1) = 1$. Since γ and μ are integers, the second case cannot be valid (the integers on the left hand side are proportional to $r-1$ whereas on the right hand such proportionality does not take place for $r > 2$). The choice $\gamma = 0$ leads to $\alpha = \mu$ and, from other two equations, one gets $\rho = 0$ and $\beta = \nu$. As a result, $h(n) = 1$, i.e. we get a trivial solution which is discarded.

Let $\delta = -1$. Solution of the taken equations gives uniquely

$$\alpha = \frac{\gamma r}{2}, \quad \mu = 1 - \frac{\gamma(r-2)}{2}, \quad \beta = \frac{\rho r}{2}, \quad \nu = -\frac{\rho(r-2)}{2},$$

where γ is an integer (for odd r it must be an even number) and ρ is an arbitrary parameter. Arbitrariness of ρ seems to contradict the statement that there are no new independent parameters in the variable u_r. This paradox is resolved in the following way. Let us first make the shift $\rho \to \rho - \gamma \sum_{k=1}^{r-1} u_k$. It is easy to see that this leads to the removal of γ from $h(n)$, i.e. it is a fake parameter. Denote now $\rho \equiv 2u_0$ and make the shifts $u_m \to u_m + u_0$, $m = 1, \ldots, r-1$. As a result, one gets

$$h(n) = \prod_{m=1}^{r-1} \frac{[n + u_0 + u_m]}{[n + u_0 - u_m]} \frac{[n + u_0 - \sum_{k=1}^{r-1} u_k]}{[n + u_0 + \sum_{k=1}^{r-1} u_k]} z, \qquad (36)$$

i.e. the parameter u_0 plays the same role as n and ellipticity in it is evident. It is not difficult to recognize in (36) the most general expression for c_{n+1}/c_n of well-poised and balanced theta hypergeometric series. Thus we have proved that ellipticity in all free parameters in $h(n)$ leads uniquely to the balancing and well-poisedness conditions.

Let us prove now that the totally elliptic hypergeometric series are automatically modular invariant. For this is it sufficient to check that the sum of squares of the parameters u entering elliptic numbers $[n+u]$ in the numerator of $h(n)$ (36) and denominator coincide. The numerator parameters generate the sum

$$\sum_{k=1}^{r-1} u_k^2 + \Big(-\sum_{k=1}^{r-1} u_k\Big)^2$$

which is trivially equal to the sum appearing from the denominator:

$$\sum_{k=1}^{r-1} (-u_k)^2 + \Big(\sum_{k=1}^{r-1} u_k\Big)^2,$$

i.e. modular invariance is automatic. (Note that well-poised theta hypergeometric series without balancing condition are not modular invariant).

All the considerations given above were designed for the bilateral ${}_rG_r$ series case, but a passage to the unilateral well-poised and balanced ${}_rE_{r-1}$-series is done by a simple specification of one of the parameters. One has to set $u_{r-1} = u_0 - 1$

and then shift $u_0 \to u_0 + 1/2$, $u_m \to u_m - 1/2$, $m = 1, \dots, r-2$. This brings $h(n)$ to the form $h(n) = z \prod_{m=0}^{r-1} [n + u_0 + u_m]/[n + 1 + u_0 - u_m]$, where we introduced anew the u_{r-1} parameter through the relation $\sum_{m=0}^{r-1} u_m = r/2$. □

Remark 13. We have replaced t_0 by t_0^2 in the definition of very-well-poised elliptic hypergeometric series ${}_{r+1}E_r$ in order to have p-shift invariance in the variable t_0 (or ellipticity in u_0). Otherwise there would be p-shift invariance in $t_0^{1/2}$.

Thus we have found interesting origins of the balancing and well-poisedness conditions for the plain and basic hypergeometric series calling for a revision of these notions. However, origins of the very-well-poisedness condition remain unknown. Probably elliptic functions $h(n)$ obeying such a constraint have some particular arithmetic properties. An indication on this is given by the transformation and summation formulas for some of the very-well-poised balanced theta hypergeometric series of the unit argument.

The theta hypergeometric series ${}_rE_s$ and ${}_rG_s$ are defined as formal infinite series. However, because of the quasiperiodicity of the theta functions it is not a simple task to determine their convergence and this problem will not be considered here. It can be shown that for some choice of parameters the radius of convergence R of the balanced ${}_{r+1}E_r$-series is equal to 1. If R is the radius of convergence of the very-well-poised ${}_{r+2}E_{r+1}$-series without balancing condition, i.e. if these infinite series are well defined for $|z| < R$, then from the representation (35) it follows that the ${}_rG_r$-series converge for $\left| q^{r-8}/R \prod_{m=1}^{r-4} t_m^2 \right| < |z| < R$. A rigorous meaning to the ${}_rE_s$-series can be given by imposing some truncation conditions. The theta hypergeometric series truncates if for some m,

$$u_m = -N - K\sigma^{-1} - M\tau\sigma^{-1}, \qquad N \in \mathbb{N}, \quad K, M \in \mathbb{Z}. \tag{37}$$

or in the multiplicative form

$$t_m = q^{-N} p^{-M}, \qquad N \in \mathbb{N}, \quad M \in \mathbb{Z}. \tag{38}$$

The well-poised elliptic hypergeometric series are double periodic in their parameters with the periods σ^{-1} and $\tau\sigma^{-1}$. Therefore, these truncated series do not depend on the integers K, M.

The top-level identity in the theory of basic hypergeometric series is the four term Bailey identity for non-terminating ${}_{10}\Phi_9$ very-well-poised balanced series of the unit argument [14]. In the terminating case there remains only two terms. In [13] Frenkel and Turaev have proved an elliptic generalization of the Bailey identity in the terminating case. In our notations it looks as follows

$$\begin{aligned} {}_{12}E_{11}(t_0; t_1, \dots, t_7; q, p; 1) &= \frac{\theta(qt_0^2, qs_0/s_4, qs_0/s_5, q/t_4t_5; p; q)_N}{\theta(qs_0^2, qt_0/t_4, qt_0/t_5, q/s_4s_5; p; q)_N} \\ &\quad \times {}_{12}E_{11}(s_0; s_1, \dots, s_7; q, p; 1), \end{aligned} \tag{39}$$

where it is assumed that $\prod_{m=0}^{7} t_m = q^2$, $t_0 t_6 = q^{-N}$, $N \in \mathbb{N}$, and

$$s_0^2 = \frac{qt_0}{t_1 t_2 t_3}, \quad s_1 = \frac{s_0 t_1}{t_0}, \quad s_2 = \frac{s_0 t_2}{t_0}, \quad s_3 = \frac{s_0 t_3}{t_0},$$
$$s_4 = \frac{t_0 t_4}{s_0}, \quad s_5 = \frac{t_0 t_5}{s_0}, \quad s_6 = \frac{t_0 t_6}{s_0}, \quad s_7 = \frac{t_0 t_7}{s_0}.$$

If one sets $t_2 t_3 = q$, then the left-hand side of (39) becomes a terminating ${}_{10}E_9$-series, and in the series on the right-hand side one gets $s_1 = 1$, i.e. only its first term is different from zero. This gives the Frenkel-Turaev sum—an elliptic generalization of the Jackson's sum for terminating very-well-poised balanced ${}_8\Phi_7$-series [14]. After diminishing the indices of $t_{4,5,6,7}$ by two it takes the following form:

$$ {}_{10}E_9(t_0; t_1, \ldots, t_5; q, p; 1) = \frac{\theta(qt_0^2; p; q)_N \prod_{1 \leq r < s \leq 3} \theta(q/t_r t_s; p; q)_N}{\theta(q/t_0 t_1 t_2 t_3; p; q)_N \prod_{r=1}^{3} \theta(qt_0/t_r; p; q)_N}, \tag{40}$$

where the parameters t_r are assumed to satisfy the balancing condition $\prod_{r=0}^{5} t_r = q$ and the truncation condition $t_0 t_4 = q^{-N}$, $N \in \mathbb{N}$.

Remark 14. Due to the clarification of the relation of very-well-poisedness condition with the general structure of theta hypergeometric series, starting from this paper we change notations for the elliptic hypergeometric series in the generalizations of Bailey and Jackson identities. The symbols ${}_{10}E_9$ and ${}_8E_7$ in the papers [28–30, 24, 26, 8–11] read in the current notations as ${}_{12}E_{11}$ and ${}_{10}E_9$ respectively.

Remark 15. Despite of the double periodicity, infinite totally elliptic hypergeometric functions are not elliptic functions of u_r since they have infinitely many poles in the parallelogram of periods. Indeed, some of the poles in t_s, $s = 1, \ldots, r-4$, are located at $t_s = t_0 q^{n+1} p^m$, where $n \in \mathbb{N}$ and $m \in \mathbb{Z}$. If $q^k \neq p^l$ for any $k, l \in \mathbb{N}$, then, evidently, there are infinitely many integers n and m such that t_s stays, say, within the bounds $|p| < |t_s| < 1$. This means that there are infinitely many poles in the parameters t_s in this annulus.

Remark 16. The symbols ${}_rE_s$ and ${}_rG_s$ were chosen for denoting the one-sided and bilateral theta hypergeometric series in order to make them as close as possible to the standard notations ${}_rF_s$ and ${}_rH_s$ used for one-sided and bilateral plain hypergeometric series respectively. The letter "E" refers also to the word "elliptic". To the author's taste greek symbols ${}_r\Phi_s$ and ${}_r\Psi_s$ used for denoting basic hypergeometric series fit well enough into the aesthetics created by the sequence of letters E, Φ, F, G, Ψ, H.

4. Multiple elliptic hypergeometric series

Following the definition of theta hypergeometric series of a single variable one can consider formal multiple sums of quasiperiodic combinations of Jacobi θ_1-functions depending of more than one summation index. However, we shall limit ourselves only to the multiple elliptic hypergeometric series.

Definition 17. The formal series

$$\sum_{\lambda_1,\ldots,\lambda_n=0}^{\infty} c(\lambda_1,\ldots,\lambda_n) \quad \text{or} \quad \sum_{\lambda_1,\ldots,\lambda_n=-\infty}^{\infty} c(\lambda_1,\ldots,\lambda_n),$$

and

$$\sum_{\substack{\lambda_1,\ldots,\lambda_n=0\\ \lambda_1\leq\cdots\leq\lambda_n}}^{\infty} c(\lambda_1,\ldots,\lambda_n) \quad \text{or} \quad \sum_{\substack{\lambda_1,\ldots,\lambda_n=-\infty\\ \lambda_1\leq\cdots\leq\lambda_n}}^{\infty} c(\lambda_1,\ldots,\lambda_n)$$

are called *multiple elliptic hypergeometric series* if: a) the coefficients $c(\lambda)$ are symmetric with respect to an action of permutation group $\mathcal{S}_n$ upon the summation variables $\lambda_1,\ldots,\lambda_n$ and the free parameters entering $c(\lambda)$; b) for all $k=1,\ldots,n$ the functions

$$h_k(\lambda)=\frac{c(\lambda_1,\ldots,\lambda_k+1,\ldots,\lambda_n)}{c(\lambda_1,\ldots,\lambda_k,\ldots,\lambda_n)},$$

are elliptic in λ_k, $k=1,\ldots,n$, considered as complex variables. These series are called totally elliptic if, in addition, the functions $h_k(\lambda)$ are elliptic in all free parameters except of the free multiplication factors.

Suppose that $h_k(\lambda)$ is symmetric in $\lambda_1,\ldots,\lambda_{k-1},\lambda_{k+1},\ldots,\lambda_n$. Then, using the results of the one variable analysis, it is not difficult to see that the most general expression for the coefficients $c(\lambda)$ is:

$$c(\lambda)=\prod_{k=1}^{n}\left(\prod_{1\leq i_1<\ldots<i_k\leq n}\prod_{m=1}^{r_k}\frac{[u_{km}]_{\lambda_{i_1}+\ldots+\lambda_{i_k}}}{[v_{km}]_{\lambda_{i_1}+\ldots+\lambda_{i_k}}}\right)z_1^{\lambda_1}\cdots z_n^{\lambda_n}, \tag{41}$$

where

$$\sum_{k=1}^{n}C_{n-1}^{k-1}\sum_{m=1}^{r_k}(u_{km}-v_{km})=0.$$

However, if the action of $\mathcal{S}_n$ permutes $\lambda_1,\ldots,\lambda_n$ and simultaneously free parameters entering $c(\lambda)$ other than $z_1,\ldots,z_n$, then the situation is richer, e.g. more general combinations of λ_k are allowed than it is indicated in (41). We shall not go into further analysis of general situation but pass to some particular examples.

Currently there are two known examples of multiple elliptic hypergeometric series leading to some constructive identities (multivariable analogues of the Frenkel-Turaev ${}_{10}E_9$-summation formula). The first one corresponds to an elliptic extension of the Aomoto-Ito-Macdonald type of series [4, 18, 19]. Its structure is read off from the following multivariable generalization of the Frenkel-Turaev summation formula considered in [31, 8, 22]. Let $N\in\mathbb{N}$ and the parameters $t,t_r\in\mathbb{C}$, $r=0,\ldots,5$, are constrained by the balancing condition $t^{2n-2}\prod_{r=0}^{5}t_r=q$ and the truncation condition $t^{n-1}t_0t_4=q^{-N}$. Then one has the following theta-functions identity

$$\sum_{0\le\lambda_1\le\lambda_2\le\cdots\le\lambda_n\le N} q^{\sum_{j=1}^n \lambda_j} t^{2\sum_{j=1}^n (n-j)\lambda_j}$$
$$\times \prod_{1\le j<k\le n} \left(\frac{\theta(\tau_k\tau_j q^{\lambda_k+\lambda_j}, \tau_k\tau_j^{-1} q^{\lambda_k-\lambda_j}; p)}{\theta(\tau_k\tau_j, \tau_k\tau_j^{-1}; p)} \right.$$
$$\left. \times \frac{\theta(t\tau_k\tau_j; p; q)_{\lambda_k+\lambda_j}}{\theta(qt^{-1}\tau_k\tau_j; p; q)_{\lambda_k+\lambda_j}} \frac{\theta(t\tau_k\tau_j^{-1}; p; q)_{\lambda_k-\lambda_j}}{\theta(qt^{-1}\tau_k\tau_j^{-1}; p; q)_{\lambda_k-\lambda_j}} \right)$$
$$\times \prod_{j=1}^n \left(\frac{\theta(\tau_j^2 q^{2\lambda_j}; p)}{\theta(\tau_j^2; p)} \prod_{r=0}^5 \frac{\theta(t_r\tau_j; p; q)_{\lambda_j}}{\theta(qt_r^{-1}\tau_j; p; q)_{\lambda_j}} \right)$$
$$= \frac{\theta(qt^{n+j-2}t_0^2; p; q)_N \prod_{1\le r<s\le 3} \theta(qt^{1-j}t_r^{-1}t_s^{-1}; p; q)_N}{\theta(qt^{2-n-j}\prod_{r=0}^3 t_r^{-1}; p; q)_N \prod_{r=1}^3 \theta(qt^{j-1}t_0 t_r^{-1}; p; q)_N}. \tag{42}$$

Here the parameters τ_j are related to t_0 and t as follows: $\tau_j = t_0 t^{j-1}$, $j = 1, \ldots, n$. Note that the series coefficients $c(\lambda)$ are symmetric with respect to the simultaneous permutation of the variables λ_j and λ_k and the parameters τ_j and τ_k for arbitrary $j \neq k$ (for this one has to assume that τ_j are independent variables).

Theorem 18. *The series standing on the left hand side of* (42) *is a totally elliptic multiple hypergeometric series.*

Proof. Ratios of the series coefficients yield

$$h_l(\lambda) = \prod_{j=1}^{l-1} \frac{\theta(\tau_j\tau_l q^{\lambda_j+\lambda_l+1}, \tau_j^{-1}\tau_l q^{\lambda_l+1-\lambda_j}, t\tau_j\tau_l q^{\lambda_j+\lambda_l}, t\tau_j^{-1}\tau_l q^{\lambda_l-\lambda_j}; p)}{\theta(\tau_j\tau_l q^{\lambda_j+\lambda_l}, \tau_j^{-1}\tau_l q^{\lambda_l-\lambda_j}, t^{-1}\tau_j\tau_l q^{\lambda_j+\lambda_l+1}, t^{-1}\tau_j^{-1}\tau_l q^{\lambda_l+1-\lambda_j}; p)}$$
$$\times \prod_{k=l+1}^{n} \frac{\theta(\tau_k\tau_l q^{\lambda_k+\lambda_l+1}, \tau_k\tau_l^{-1} q^{\lambda_k-\lambda_l-1}, t\tau_k\tau_l q^{\lambda_k+\lambda_l}, t^{-1}\tau_k\tau_l^{-1} q^{\lambda_k-\lambda_l}; p)}{\theta(\tau_k\tau_l q^{\lambda_k+\lambda_l}, \tau_k\tau_l^{-1} q^{\lambda_k-\lambda_l}, t^{-1}\tau_k\tau_l q^{\lambda_k+\lambda_l+1}, t\tau_k\tau_l^{-1} q^{\lambda_k-\lambda_l-1}; p)}$$
$$\times qt^{2(n-l)} \frac{\theta(\tau_l^2 q^{2\lambda_l+2}; p)}{\theta(\tau_l^2; p)} \prod_{m=0}^5 \frac{\theta(t_m\tau_l q^{\lambda_l}; p)}{\theta(t_m^{-1}\tau_l q^{\lambda_l+1}; p)}. \tag{43}$$

Using the equalities (3) one can easily check the ellipticity of this $h_l(\lambda)$ in λ_i for $i < l$ and $i > l$ (for this it is simply necessary to see that h_l does not change after the replacement of q^{λ_i} by pq^{λ_i}). For the ellipticity in λ_l itself one has an essentially longer computation. The replacement of q^{λ_l} by pq^{λ_l} in the product $\prod_{j=1}^{l-1}$ yields a multiplier $t^{-4(l-1)}$. The product $\prod_{k=l+1}^{n}$ yields the multiplier $t^{-4(n-l)}$. The remaining part of h_l generates the factor $q^{-4}\prod_{m=0}^5 qt_m^{-2}$. The product of all these three factors takes the form $q^2 t^{-4(n-1)}\prod_{m=0}^5 t_m^{-2}$ and it is equal to 1 due to the balancing condition.

So, we have found that the taken series is indeed a multiple elliptic hypergeometric series. Let us prove now its total ellipticity or p-shift invariance in the parameters t_m, $m = 0, \ldots, 4$ and t. The p-shift invariance in the parameters $t_1, \ldots, t_4$ follows from the balancing condition in the same way as in the single variable series case. Consider the $t_0 \to pt_0$ shift. The product $\prod_{j=1}^{l-1}$ yields a multiplier $t^{-4(l-1)}$ and the product $\prod_{k=l+1}^{n}$ yields the factor $t^{-4(n-l)}$. The remaining part of h_l generates the factor $q^2 \prod_{m=0}^{5} t_m^{-2}$. The product of all these three multipliers is equal to 1.

Finally, the shift $t \to pt$ calls for the most complicated computation. The product $\prod_{j=1}^{l-1}$ yields a complicated multiplier $(q^{2\lambda_l+1} t^{2(l-1)} \tau_l^2 p^{2l-3})^{2(1-l)}$. The product $\prod_{k=l+1}^{n}$ generates no less complicated expression $(q^{2\lambda_l+1} t^{2(l-1)} \tau_l^2 p^{2(l-1)})^{2(l-n)}$. The remaining part of h_l leads to the following factor (after the use of the balancing condition): $(q^{2\lambda_l+1} t^{2(l-1)} \tau_l^2)^{2(n-1)} p^{(4n-6)(l-1)}$. The product of all these three multipliers yields 1. Thus we have proved the total ellipticity of the taken type of series. □

The second example of multiple series corresponds to an elliptic generalization of the Milne–Gustafson type multiple basic hypergeometric series [20, 16, 7], which are, in turn, q-analogues of the Hollman, Biedenharn, and Louck plain multiple hypergeometric series [17]. Its structure is read off from the following summation formula suggested in [10]. Let $q^n \neq p^m$ for $n, m \in \mathbb{N}$. Then for the parameters $t_0, \ldots, t_{2n+3}$ subject to the balancing condition $q^{-1} \prod_{r=0}^{2n+3} t_r = 1$ and the truncation conditions $q^{N_j} t_j t_{n+j} = 1$, $j = 1, \ldots, n$, where $N_j \in \mathbb{N}$, one has the identity

$$
\begin{aligned}
&\sum_{\substack{0 \le \lambda_j \le N_j \\ j=1,\ldots,n}} q^{\sum_{j=1}^{n} j\lambda_j} \prod_{1 \le j < k \le n} \frac{\theta(t_j t_k q^{\lambda_j+\lambda_k}, t_j t_k^{-1} q^{\lambda_j-\lambda_k}; p)}{\theta(t_j t_k, t_j t_k^{-1}; p)} \\
&\qquad \times \prod_{1 \le j \le n} \left(\frac{\theta(t_j^2 q^{2\lambda_j}; p)}{\theta(t_j^2; p)} \prod_{0 \le r \le 2n+3} \frac{\theta(t_j t_r; p; q)_{\lambda_j}}{\theta(q t_j t_r^{-1}; p; q)_{\lambda_j}} \right) \\
&= \theta(qa^{-1}b^{-1}, qa^{-1}c^{-1}, qb^{-1}c^{-1}; p; q)_{N_1+\cdots+N_n} \\
&\qquad \times \prod_{1 \le j < k \le n} \frac{\theta(q t_j t_k; p; q)_{N_j} \theta(q t_j t_k; p; q)_{N_k}}{\theta(q t_j t_k; p; q)_{N_j+N_k}} \\
&\qquad \times \prod_{1 \le j \le n} \frac{\theta(q t_j^2; p; q)_{N_j}}{\theta(q t_j a^{-1}, q t_j b^{-1}, q t_j c^{-1}, q^{1+N_1+\cdots+N_n-N_j} t_j^{-1} a^{-1} b^{-1} c^{-1}; p; q)_{N_j}},
\end{aligned}
\tag{44}
$$

where $a \equiv t_{2n+1}$, $b \equiv t_{2n+2}$, $c \equiv t_{2n+3}$. Note that this series coefficients $c(\lambda)$ are symmetric with respect to simultaneous permutation of the variables λ_j and λ_k together with the parameters t_j and t_k for arbitrary $j, k = 1, \ldots, n$, $j \neq k$.

Theorem 19. *The series standing on the left hand side of* (44) *is a totally elliptic hypergeometric series.*

Proof. Ratios of the successive series coefficients yield

$$h_l(\lambda)=\prod_{j=1}^{l-1}\frac{\theta(t_jt_lq^{\lambda_j+\lambda_l+1},t_jt_l^{-1}q^{\lambda_j-\lambda_l-1};p)}{\theta(t_jt_lq^{\lambda_j+\lambda_l},t_jt_l^{-1}q^{\lambda_j-\lambda_l};p)}$$
$$\times\prod_{k=l+1}^{n}\frac{\theta(t_lt_kq^{\lambda_l+\lambda_k+1},t_lt_k^{-1}q^{\lambda_l+1-\lambda_k};p)}{\theta(t_lt_kq^{\lambda_l+\lambda_k},t_lt_k^{-1}q^{\lambda_l-\lambda_k};p)}$$
$$\times q^l\frac{\theta(t_l^2q^{2\lambda_l+2};p)}{\theta(t_l^2q^{2\lambda_l};p)}\prod_{m=0}^{2n+3}\frac{\theta(t_lt_mq^{\lambda_l};p)}{\theta(t_lt_m^{-1}q^{\lambda_l+1};p)}. \tag{45}$$

It is easy to check the ellipticity of this expression in λ_j for $j<l$ and $j>l$. Ellipticity in λ_l itself follows from a more complicated computation. Namely, the change of q^{λ_j} to pq^{λ_j} leads to additional multipliers in $h_l(\lambda)$: $q^{2(1-l)}$—from the product $\prod_{j=1}^{l-1}$, $q^{2(l-n)}$—from the product $\prod_{k=l+1}^{n}$, and $q^{-4}\prod_{m=0}^{2n+3}qt_m^{-2}$—from the rest of $h_l(\lambda)$. Multiplication of these three expressions gives 1 due to the balancing condition.

Ellipticity in the parameters t_m, $m=0,\dots,2n+3$, is checked separately for $m<l, m>l$ and $m=l$. The first two cases are easy enough and do not worth of special consideration. The replacement $t_l\to pt_l$ leads to the following multipliers: $q^{2(1-l)}$—from the product $\prod_{j=1}^{l-1}$, $q^{2(l-n)}$—from the product $\prod_{k=l+1}^{n}$, and $q^{-2n}\prod_{m=0}^{2n+3}t_m^{-2}$—from the rest of $h_l(\lambda)$. The balancing condition guarantees again that the total multiplier is equal to 1. Thus we have proved total ellipticity of this series as well. □

Remark 20. Modular invariance of the series (42) and (44) has been established in [8] and [10] respectively. We conjecture that all totally elliptic multiple hypergeometric series are automatically modular invariant similar to the one-variable series situation. One can introduce general notions of well-poised and very-well-poised multiple theta hypergeometric series, with the given above examples being counted as very-well-poised series, but we shall not discuss this topic in the present paper.

The author is indebted to J. F. van Diejen and A. S. Zhedanov for a collaboration in the work on elliptic hypergeometric series and for useful discussions of this paper. Valuable comments and encouragement from G. E. Andrews, R. Askey, A. Berkovich are highly appreciated as well.

References

1. Abramowitz, M. and Stegun, I. A., Eds. (1964) ***Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables***, National Bureau of Standards Applied Mathematics Series, Vol. 55, US Government Printing Office, Washington, DC.
2. Andrews, G. E., Askey, R. and Roy, R. (1999) ***Special Functions***, Encyclopedia of Math. Appl. **71**, Cambridge Univ. Press, Cambridge.
3. Andrews, G. E., Baxter, R. J. and Forrester, P. J. (1984) Eight-vertex SOS model and generalized Rogers–Ramanujan type identities, *J. Stat. Phys.*, **35**, 412–485.
4. Aomoto, K. (1998) On elliptic product formulas for Jackson integrals associated with reduced root systems, *J. Alg. Comb.*, **8**, 115–126.

5. Baxter, R. J. (1982) *Exactly Solved Models in Statistical Mechanics*, Academic Press, London, 1982.
6. Date, E., Jimbo, M., Kuniba, A., Miwa, T. and Okado, M. (1987) Exactly solvable SOS models: Local height probabilities and theta function identities, *Nucl. Phys.* B **290**, 231–273; (1988) II: Proof of the star-triangle relation and combinatorial identities, in M. Jimbo, T. Miwa, and A. Tsuchiya (Eds.) *Conformal Field Theory and Lattice Models*, Advanced Studies in Pure Math. **16**, 17–122.
7. Denis R. Y. and Gustafson, R. A. (1992) An $SU(n)$ q-beta integral transformation and multiple hypergeometric series identities, *SIAM J. Math. Anal.* **23**, 552–561.
8. van Diejen, J. F. and Spiridonov, V. P. (2000) An elliptic Macdonald–Morris conjecture and multiple modular hypergeometric sums, *Math. Res. Letters* **7**, 729–746.
9. van Diejen, J. F. and Spiridonov, V. P. (2001) Elliptic Selberg integrals, *Internat. Math. Res. Notices*, no. 20, 1083–1110.
10. van Diejen, J. F. and Spiridonov, V. P. (in press) Modular hypergeometric residue sums of elliptic Selberg integrals, *Lett. Math. Phys.*
11. van Diejen, J. F. and Spiridonov, V. P. (in press) Elliptic beta integrals and modular hypergeometric sums: an overview, *Proc. NSF Conference "Special Functions 2000: Current Perspective and Future Directions"*, (Tempe, Arizona, May 29–June 9, 2000), Rocky Mountain J. Math.
12. Eichler, M. and Zagier, D. (1985) *The Theory of Jacobi Forms*, Progress in Math. **55**, Birkhäuser, Boston.
13. Frenkel, I. B. and Turaev, V. G. (1997) Elliptic solutions of the Yang–Baxter equation and modular hypergeometric functions, in V. I. Arnold, I. M. Gelfand, V. S. Retakh, and M. Smirnov (Eds.) *The Arnold–Gelfand Mathematical Seminars*, Birkhäuser, Boston, 171–204.
14. Gasper, G. and Rahman, M. (1990) *Basic Hypergeometric Series*, Encyclopedia of Mathematics and its Applications **35**, Cambridge Univ. Press, Cambridge.
15. Gelfand, I. M., Graev, M. I. and Retakh, V. S. (1992) General hypergeometric systems of equations and series of hypergeometric type, *Uspekhi Mat. Nauk* (Russ. Math. Surveys), **47**, 3–82 (in Russian).
16. Gustafson, R. A. (1989) The Macdonald identities for affine root systems of classical type and hypergeometric series very-well-poised on semisimple Lie algebras, in (Eds. N. K. Thakare, K. C. Sharma, T. T. Raghunathan), *Ramanujan International Symposium on Analysis*, Macmillan of India, New Dehli, 185–224.
17. Hollman III, W. J., Biedenharn, L. C. and Louck, J. D. (1976) On hypergeometric series well-poised in $SU(n)$, *SIAM J. Math. Anal.*, **7**, 529–541.
18. Ito, M. (1997) On a theta product formula for Jackson integrals associated with root systems of rank two, *J. Math. Anal. Appl.*, **216**, 122–163.
19. Macdonald, I. G. (1998) Constant term identities, orthogonal polynomials, and affine Hecke algebras, *Doc. Math.*, DMV Extra Volume ICM I, 303–317.
20. Milne, S. C. (1989) The multidimensional ${}_1\Psi_1$ sum and Macdonald identities for $A_l^{(1)}$, in L. Ehrenpreis and R. C. Gunning (Eds.), *Theta Functions Bowdoin 1987*, Proc. Symp. Pure Math. **49** (part 2), Amer. Math. Soc., Providence, R.I., 323–359.
21. Mumford, D. (1983, 1984) Tata Lectures on Theta I, II, *Progress in Math.* **28, 43**, Birkhäuser, Boston.
22. Rosengren, H. (in print) A proof of a multivariable elliptic summation formula conjectured by Warnaar, *Contemp. Math.*
23. Slater, L. J. (1966) *Generalized Hypergeometric Functions*, Cambridge Univ. Press, Cambridge.
24. Spiridonov, V. P. (2000) Solitons and Coulomb plasmas, similarity reductions and special functions, in C. Dunkl, M. E. H. Ismail, and R. Wong (Eds.), Proc. International Workshop "Special Functions" (Hong Kong, China, June 21–25, 1999), World Scientific, Singapore, 324–338.
25. Spiridonov, V. P. (in print) An elliptic beta integral, in S. Elaydi, J. Fenner-Lopez, and G. Ladas (Eds.), *Proc. Fifth International Conference on Difference Equations and Applications* (Temuco, Chile, January 3–7, 2000), Gordon and Breach; (2001) On an elliptic beta function, *Russian Math. Surveys*, **56**, no. 1, 185–186.
26. Spiridonov, V. P. (2001) Elliptic beta integrals and special functions of hypergeometric

type, in G. von Gehlen and S. Pakuliak (Eds.), *Proc. NATO ARW "Integrable Structures of Exactly Solvable Two-Dimensional Models of Quantum Field Theory"*, (Kiev, Ukraine, September 25-30, 2000), Kluwer Academic Publishers, Dordrecht, 305–313.

27. Spiridonov, V. P. (in preparation) Continuous biorthogonal functions, theta hypergeometric series, and elliptic beta integrals.
28. Spiridonov, V. P. and Zhedanov, A. S. (2000) Spectral transformation chains and some new biorthogonal rational functions, *Commun. Math. Phys.*, **210**, 49–83.
29. Spiridonov, V. P. and Zhedanov, A. S. (2000) Classical biorthogonal rational functions on elliptic grids, *C. R. Math. Rep. Acad. Sci. Canada*, **22**, no. 2, 70–76.
30. Spiridonov, V. P. and Zhedanov, A. S. (2001) Generalized eigenvalue problem and a new family of rational functions biorthogonal on elliptic grids, in J. Bustoz, M. E. H. Ismail, and S. K. Suslov (Eds.), *Proc. NATO ASI "Special functions-2000"* (Tempe, USA, May 29–June 9, 2000), Kluwer Academic Publishers, Dordrecht, 365–388.
31. Warnaar, S. O. (2000) Summation and transformation formulas for elliptic hypergeometric series, preprint .
32. Whittaker, E. T. and Watson, G. N. (1986) *A Course of Modern Analysis*, Cambridge Univ. Press, Cambridge.