

Ina Kersten,
Ralf Meyer (Eds.)



Symmetries in Algebra and Number Theory

Proceedings of the
Göttingen–Jerusalem
Conference
held October 27–30,
2008 in Göttingen

Including contributions
to the Satellite Conference
„On the Legacy of
Hermann Weyl“
held November 3–4, 2008



Universitätsverlag Göttingen

Ina Kersten, Ralf Meyer (Eds.)
Symmetries in Algebra and Number Theory

This work is licensed under the [Creative Commons](#) License 2.0 “by-nd”, allowing you to download, distribute and print the document in a few copies for private or educational use, given that the document stays unchanged and the creator is mentioned. You are not allowed to sell copies of the free version.



erschienen im Universitätsverlag Göttingen 2009

Ina Kersten,
Ralf Meyer (Eds.)

Symmetries in Algebra and
Number Theory
(SANT)

Proceedings of the Göttingen-
Jerusalem Conference held
October 27-30, 2008 in Göttingen

Including contributions to the
Satellite Conference
“On the Legacy of Hermann Weyl”,
held November 3-4, 2008



Universitätsverlag Göttingen
2009

Bibliographische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Daten sind im Internet über <http://dnb.ddb.de> abrufbar.

Address of the Editors

Ina Kersten
Mathematisches Institut
Georg - August Universität Göttingen
Bunsenstr. 3-5
37073 Göttingen
E-mail: kersten@uni-math.gwdg.de

Ralf Meyer
Mathematisches Institut
Georg - August Universität Göttingen
Bunsenstr. 3-5
37073 Göttingen
E-mail: r Meyer@uni-math.gwdg.de

This work is protected by German Intellectual Property Right Law.
It is also available as an Open Access version through the publisher's homepage and the Online Catalogue of the State and University Library of Goettingen (<http://www.sub.uni-goettingen.de>). Users of the free online version are invited to read, download and distribute it. Users may also print a small number for educational or private use. However they may not sell print versions of the online book.

Satz und Layout: Stefan Wiedmann
Umschlaggestaltung: Margo Bargheer
Titelabbildung: Göttinger Modellsammlung, Modell Nr. 302
„Fläche 5. Ordnung mit vielen Symmetrieeigenschaften“

© 2009 Universitätsverlag Göttingen
<http://univerlag.uni-goettingen.de>
ISBN: 978-3-940344-96-0

ACKNOWLEDGEMENT

The joint Göttingen-Jerusalem conference on “Symmetries in Algebra and Number Theory” (SANT) took place in the German-Israeli Year of Science and Technology 2008, and the German Year of Mathematics 2008.

Many thanks for supporting this conference go to the Ministry of Science and Culture in Niedersachsen, the University Göttingen, the newly founded Courant Center “Higher Order Structure in Mathematics” in Göttingen as well as to the Hebrew University Jerusalem.

Further thanks go to the Courant Center “Higher Order Structure in Mathematics”, and the Research Training Group „Mathematical Structures in Modern Quantum Physics“ for supporting the satellite conference on the Legacy of Herman Weyl.

I am very grateful to the coorganizer of the SANT conference David Kazhdan from Jerusalem, the organizer of the satellite conference Ralf Meyer from Göttingen, and to Claudia Gabler from Göttingen for supporting the organizers in a perfect way. Further thanks go to all participants, guests, lecturers, and authors contributing to the very success of both of the conferences.

February 2009

Ina Kersten

Participants of the SANT Conference

Organized by David Kazhdan and Ina Kersten

Dorothea Bahns (Göttingen)
Laurent Bartholdi (Göttingen)
Valentin Blomer (Toronto)
Ehud de Shalit (Jerusalem)
Norbert Hoffmann (Göttingen)
David Kazhdan (Jerusalem)
Ina Kersten (Göttingen)
Erez Lapid (Jerusalem)
Hannah Marwig (Göttingen)
Ralf Meyer (Göttingen)
Preda Mihailescu (Göttingen)
Julio Jose Moyano Fernandez (Osnabrück)
Samuel J. Patterson (Göttingen)
Pablo Ramacher (Göttingen)
Norbert Schappacher (Strasbourg)
Michael Schein (Jerusalem)
Jan van Geel (Gent)
Yakov Varshavsky (Jerusalem)
Ingo Witt (Göttingen)
Chengchang Zhu (Göttingen)

Participants of the Conference on the Legacy of Hermann Weyl

Organized by Ralf Meyer

Dennis Hejhal (Minnesota/Uppsala)
David Kazhdan (Jerusalem)
Russell Lyons (Bloomington)
Ralf Meyer (Göttingen)
Werner Müller (Bonn)
Samuel J. Patterson (Göttingen)
Norbert Straumann (Zürich)
Ingo Witt (Göttingen)

CONTENTS

Acknowledgement	v
Participants of the SANT Conference	vi
Participants of the Conference on the Legacy of Hermann Weyl	vi
Opening	xi
Welcome note Minister Stratmann	xi
Welcome note Vice-President Prof. Dr. Münch	xii
Welcome note Representative of the Hebrew University Dr. Eikenberg	xiv
Abstracts	xvii
LAURENT BARTHOLDI — <i>Self-similarity in group theory and algebra</i>	1
1. Self-similar sets	2
2. Self-similar groups	2
3. Self-similar associative algebras	6
4. Self-similar Lie algebras	7
References	9
VALENTIN BLOMER & GERGELY HARCOS — <i>L-functions, automorphic forms, and arithmetic</i>	11
1. <i>L</i> -functions	11
2. Automorphic forms on GL_2	14
3. Subconvexity for automorphic <i>L</i> -functions	17
4. Applications	21
References	23
EHUD DE SHALIT — <i>Bounded cohomology of the <i>p</i>-adic upper half plane</i>	27
1. Introduction	27
2. Differential forms and cohomology with coefficients in <i>M</i>	32
3. Bounded forms and bounded cohomology	39
References	47

NORBERT HOFFMANN — <i>On Arakelov vector bundles over arithmetic curves</i> . . .	49
Introduction	49
1. Arithmetic curves	50
2. Arakelov vector bundles on arithmetic curves	51
3. Riemann-Roch	52
4. The Narasimhan-Seshadri theorem	53
5. Faltings' criterion for semistability	54
References	56
TOBIAS FINIS & EREZ LAPID — <i>A conjectural non-commutative generalization of a volume formula of McMullen-Schneider</i>	59
1. Preliminaries	59
2. Piecewise polynomial functions	60
3. The setup	61
4. The conjectural formula	63
5. Remarks	63
References	64
HANNAH MARKWIG — <i>The j-invariant of a plane tropical cubic</i>	65
1. The algebraic setting	65
2. The tropical setting	66
3. The result	67
References	68
PREDA MIHĂILESCU — <i>Leopoldt's conjecture for some galois extensions</i>	69
1. Introduction	70
2. Baker theory and Leopoldt's conjecture	70
3. The solvable case and beyond	75
References	76
PREDA MIHĂILESCU — <i>On Leopoldt's conjecture and a special case of Greenberg's conjecture</i>	79
1. Introduction	80
2. Local theory	82
3. Auxiliary fields	84
4. Proof of Theorem 1	86
5. The Phantom Field	91
References	97

RALF MEYER — <i>Homological algebra for Schwartz algebras</i>	99
1. Introduction	99
2. Categories of representations	101
3. Why do we need bornological modules?	104
4. Some ideas from the proof of the main theorem	108
5. Some applications	111
6. Conclusion	113
References	113
MICHAEL M. SCHEIN — <i>Weights in generalizations of Serre's conjecture and the mod p local Langlands correspondence</i>	115
1. Introduction	115
2. Serre's conjecture for Hilbert modular forms	118
3. Representations of $GL_n(M)$, for M a p -adic field	124
4. Towards the mod p local Langlands correspondence	133
References	136
YAKOV VARSHAVSKY — <i>Towards Langlands correspondence over function fields for split reductive groups</i>	139
1. Langlands correspondence	139
2. Formulation of the result	141
3. Moduli spaces of F -bundles	142
4. The explicit form of the First Main theorem	144
5. Strategy of the proof of the Second Main Theorem	145
References	146
INGO WITT — <i>Maximal Sobolev regularity at radial points</i>	149
1. Introduction	149
2. Facts from microlocal analysis	152
3. Analytic properties of radial points	154
4. Sketch of proof of the main theorem	158
References	160
CHENCHANG ZHU — <i>Elliptic gamma function provides the Čech cocycle of a gerbe</i>	161
1. Introduction	161
2. Enlarged gamma function family	162
3. The geometry of the universal triptic curve	162
4. Main theorems	165
References	168

DAVID KAZHDAN — <i>The affine Macdonald's formula</i>	169
1. The affine Macdonald's formula	169
NORBERT STRAUMANN — <i>Hermann Weyl and the Early History of Gauge Theories</i>	173
1. Introduction	173
2. Weyl's Attempt to Unify Gravitation and Electromagnetism	174
3. Weyl's 1929 Classic: "Electron and Gravitation"	180
4. Concluding Remarks	187
References	188

OPENING

Welcome note Minister Stratmann

At the age of seven, Carl Friedrich Gauß added the numbers of 1 to 100 as fast as lightning and obtained the result: 5050. How did he solve the equation so quickly? Would we have been able to do this at such a young age? Can some of us do it now? These questions, among others, take centre stage in the Year of Mathematics 2008.

Apart from being the Year of Mathematics, 2008 is also the German-Israeli Year of Science and Technology. Both of these issues merge in the “Conference on Symmetries in Algebra and Number Theory” which is organized cooperatively by the Mathematics Faculty of the Georg-August-University Göttingen and the Einstein Institute of Mathematics of the Hebrew University Jerusalem and hosted in Göttingen from October 27 until October 30, 2008.

French mathematician Émile Borel once described his discipline as “poetry of ideas”. This characterization is validated by numerous prosaic, technical developments for which mathematics provides the fundamental condition. Neither the invention nor the production of computers, mp3-players or other machines would be possible without the various achievements generated in the field of mathematics. Technical terms such as fuzzy logic, chaos theory or computer simulation pervade our everyday language. Concurrently, the prevalence of these concerns demonstrates that since our world functions in ever more complex and mechanized modes, we become increasingly dependent on the regulative force of mathematics.

By assembling the specialist conference “Symmetries in Algebra and Number Theory”, scientists once again impressively prove that there is nothing more practical than a good theory.

On this note, I extend my best wishes to all participants for inspiring discussions and exciting insights that will advance science in Israel and in Germany via the power of mathematics.

Sincerely,



Lutz Stratmann
Minister of Science and Culture of Niedersachsen

Welcome note Vice-President Prof. Dr. Münch

Honourable Dr Eikenberg, Honourable Professor Kazhdan, Dear Professor Kersten, Honourable Guests!

It is a great pleasure for me to welcome you in the name of the presidential board of the Georg-August-Universität in Göttingen to the conference “Symmetries in Algebra and Number Theory”.

Let me begin with a little historical story: In January 1801 Giuseppe Piazzi, a priest and astronomer in Palermo, discovered the first asteroid Ceres. Piazzi observed Ceres from January 1st until February 11th. Then the weather became bad in Sicily and in summer Ceres was too close to the sun to be observed. Piazzi published the celestial positions of Ceres determined during the first 40 days of the year. Although these data were rather sparse, a few scientists tried to calculate the positions where Ceres should be found in November and December of 1801. One of them – 24 year old Carl Friedrich Gauss – predicted a position which was about 10 degrees off the other astronomers’ estimates. The astronomer von Zach in Gotha, who was the only observer to know Gauss’ calculation, found Ceres on December 7th. This success made Gauss famous in the scientific world.

This extraordinary achievement of a young mathematician is certainly quite spectacular. Nevertheless, in Göttingen we are reluctant to be proud of it. Our reluctance has two reasons. Firstly, in 1801 Gauss was not yet living in Göttingen but in his hometown Brunswick. It was the time between his student years in Göttingen and the almost 50 years when he was a professor of our university and director of our observatory. So we cannot claim that this was an achievement which is to be associated with Göttingen. The other reason for our reluctance is more serious. Being a member of the Georg-August-Universität and more generally being German, I find it hard to feel enough shame for the years between 33 and 45 of the last century in order to be allowed to feel pride for whatever brilliant achievements have been made in the decades and centuries before, namely in the field of mathematics. What nazi-barbarism meant for the development of science in Germany is adequately expressed by a quote of David Hilbert. In 1934 he was asked by the new Prussian secretary for education, whether his institute suffered any loss by the leave of Jewish scientists and of dissidents.

David Hilbert is reported to have answered: “The institute doesn’t even exist any more“. Hilbert knew that the loss of eminent minds like Landau, Courant, Bernstein, Noether, Blumenthal, and Weyl was beyond estimate. Like Ceres was lost in the daylight of the summer of 1801, Göttingen science was lost in the night of German history. More than 200 years ago, it took a mathematical hero to find the asteroid when the sky was clear again.

During the last 60 years we have struggled to find again the Göttingen University of former times known as a place for ideas, knowledge, and understanding. We have gone a long way, we have made many steps forward and sometimes found that our steps are yet too small. Since last year the German excellence program for universities is providing us with an extraordinary opportunity to take a leap forward in the search of excellence. 15 young researchers have been hired for our five new Courant Centres, which will change the face of Göttingen University. One of the Courant Centres is dedicated to "Higher Order Structures in Mathematics". Again we depend on scientific and scholarly heroes: not only those who are part of our University but also those who are part of the worldwide scientific community, who work together with our researchers and help Göttingen to be what it is best at: a place for the advancement of science.

Let me quote David Hilbert a second time. It is also reported he had said that one of his students became a novelist, because he lacked the force of imagination necessary for mathematics. I cannot say whether this is true or just made up. Anyhow, we are grateful to you for sharing our mathematicians' interests and for sharing your finest resource, namely your thoughts, ideas, and your force of imagination.

Ladies and Gentlemen, thank you for being here. I wish you a successful conference, new insights, an inspiring and pleasant stay in Göttingen.

Sincerely,



Joachim Münch
Vice-President of the Georg-August-Universität

Welcome note Representative of the Hebrew University Dr. Eikenberg

Shalom – Boka tov – Bruchim haba'im

Using these three well known Hebrew words, let me thank the organizers for inviting me as representative of the Hebrew University of Jerusalem in Germany to your prestigious bilateral Conference and for the honour to contribute with some remarks to this opening ceremony.

Looking at the media, at exhibitions, lectures and at other events such as your conference, I must say: 2008 as the year of mathematics has become even for a mathematical layperson like me more than just only a label which was attached by the Federal Minister of Education and Research to the current year in Germany. In any case, I can speak from own experience: there have been remarkably more points of contact with your field which were reaching me during this year 2008 than ever before. But to put in mathematical terms these contacts were not more than tangential points. To give you some examples:

First: In a recent article about a symposium dealing with the mathematical master pieces of Albrecht Dürer one of the lecturers was quoted saying, Mathematics is as difficult as the Eiger North Face, which is for climbing alpinists one of the most challenging mountains in the Alps. I don't know whether the lecturer was an art historian or a Mathematician. In any case he was in line with my respect for your discipline.

Second: More tempting was at first glance what I read when I visited an exhibition about Jewish Mathematicians in the German speaking academic community, currently presented in Bonn, covering the period from the 19th century until the first decades after the Second World War and travelling altogether to seven German cities. One of the exhibits was focussing on Edmund Landau who undertook with his book "Grundlagen der Analysis", published in 1930, to address students as well as mathematical laypersons. This was the reason, why he wrote two prefaces, starting with the phrase. "I only require from my readers logical thinking and the mastery of the German language, however nothing related to the so-called schoolboy mathematics or even to the high level mathematics." To be honest I never reached the top of the Eiger North Face nor did I continue the reading of Landau's book, in spite of the fact that one of the Hebrew University's Minerva Centres was named "Edmund Landau Minerva Center for Research in Mathematical Analysis".

Third: Only a few days ago I listened to one of the so-called Hirzebruch lectures at the University of Bonn, offered by the well-known German writer Hans-Magnus Enzensberger under the title "Metaphysical moods of Mathematics" ("Methaphysische Mucken der Mathematik"). To be honest again, I attended this lecture mainly out of interest for Enzensberger. Enzensberger – with a major interest for mathematics over all his life – focussed in his lecture on the philosophy of mathematics. I do

not claim I was able to follow him all the time. But he finished his remarks with the following phrase: “Most people do not have the slightest idea, to which extent their life depends on a long chain of Mathematicians”.

Fourth: Once, I got a clue of the impact mathematicians had and have in the field of arts when I was confronted for the first time with Marcel Duchamps, a Franco-American artist. His work, in particular his Ready-mades, ironizing inter alia the principle of gravity and influenced by an anti-scientific scepticism - expressed for example in a hat-stand, hanging down by a thread from the ceiling - were influenced by the mathematician and physicist Henri Poincaré who regarded the scientific laws only as subjective understandings between human beings.

What is the concluding message which I draw from these rather superficial contacts with your field? I would not be well advised to use my words for any remarks about the scientific objectives of your Conference. To be on the safe side I will end, therefore, my remarks with some general comments.

It is probably not by pure accident that your conference happens to take place this year. As I mentioned before, 2008 was declared in Germany the Year of Mathematics, i.a. to underline the importance of this scientific discipline, to win more young people to focus their studies in this field and to increase at least the understanding of the general public for the significance of mathematics.

In addition, I conclude from the logo on the website for your conference that your event is also linked to the German Israeli Year of Science and Technology, which is celebrated in 2008 and from which both countries expect new impulses for their cooperation, which is, as I assume, still possible even within the dense network of already ongoing bilateral scientific contacts.

Moreover, your conference is integrated in a chain of joint undertakings in various fields including joint workshops, joint bilateral or EU-financed projects between the Hebrew University of Jerusalem and the Georg-August- Universität Göttingen. I think, one can say without hesitation: the partnership between the two Universities is full of life.

My next remark refers to the role of Mathematics at Hebrew and Göttingen University. I do not want to carry coal to New Castle. You know much better than I do what Mathematics means to both of your universities in terms of scientific contributions and in terms of the role they play for the overall reputation of your research communities.

Of course, there is also the historical dimension. I am referring to the period until 1933 with Goettingen's reputation as the most distinguished scientific centre of mathematics. As I learnt again from the exhibition which I mentioned at the beginning, such reputation was highly based on at least a temporary, although never completely unchallenged German-Jewish synthesis – or should I say symbiosis? - and on the substantial contributions by Jewish mathematicians. I was informed

that Lorenz Abraham Stern, a mathematician, was around 1860 the first Jew to be appointed as a full professor in Germany - as a Jew who stuck to his traditional religious belief. And this happened in Göttingen. 1933 20 out of 94 full professors for mathematics in Germany were of Jewish origin. Beginning in that year according to a study 127 Jewish mathematicians were driven out of Germany, including 19 from Göttingen – an irreversible loss as in so many other scientific fields and areas outside the research area. 15 Jewish mathematicians from Germany perished during the Nazi regime. The figures, I am quoting, are derived from the exhibition and its catalogue, which also expresses a word of caution concerning their finality.

In the light of such past there is still enough reason - even more than 60 years after the end of the Nazi regime - to praise that it goes now without saying that Israeli and German scientists have been meeting each other for many years, are cooperating with each other and often have made personal friendship.

Let me add a last point. The website for your conference is mentioning the Einstein-Institute of Mathematics at the Hebrew University as a co-organizer. The name of Einstein leads me to another event which I would like to bring to your attention. “Einstein revisits Humboldt” is the title for the presentation of the Einstein Centre of the Hebrew University hosted by the Humboldt Universität zu Berlin in cooperation with the Max-Planck-Institute for the History of Science in Berlin on 5 November 2008. Within four working groups focussing on physics, on astrophysics, on Philosophy of Science and on International Relations Scientists from the Hebrew University will meet with their German partners and will present themselves during a festive event in the afternoon to the general public. I have brought with me a poster drawing the attention to the public part of the event and a couple of invitations together with detailed programs. I leave them with you in case of interest.

And last not least, mentioning Humboldt leads again to Goettingen and to its famous mathematician Carl Friedrich Gauss who also met Humboldt. Gauss and Alexander von Humboldt are the main figures in a recent novel by the Austrian writer Daniel Kehlmann. “Measuring the world” is a book full of humour, full of information, filled with short and precise phrases and scenes. In this case easy Mathematics for me!

But now I have moved already far away from the real topic of your Conference and I better stop here in order to leave this latter topic in your hands and not to steal more time from your deliberations. I am wishing you a successful conference inspired by the genius loci.

Thank you for your attention.



Henning Eikenberg

Representative of the Hebrew University for EU Research Framework Programmes

ABSTRACTS

Self-similarity in group theory and algebra
LAURENT BARTHOLDI 1

We describe avatars of self-similarity in the worlds of groups, associative and Lie algebras; in each case, we highlight some important examples and the questions or conjectures that they answer.

L-functions, automorphic forms, and arithmetic
VALENTIN BLOMER & GERGELY HARCOS 11

We give a short, informal survey on the role of automorphic L -functions in number theory. We present the strongest currently known subconvexity bounds for twisted L -functions over number fields due to the authors and give various arithmetic applications. This is based on a talk of the first author.

Bounded cohomology of the p -adic upper half plane
EHUD DE SHALIT 27

We consider the bounded cohomology of Drinfel'd's p -adic upper half plane, with values in a local system M . We show that it has a canonical splitting into the space of bounded one-forms with values in M . The paper is largely expository, and self-contained.

On Arakelov vector bundles over arithmetic curves
NORBERT HOFFMANN 49

Here 'arithmetic curve' means the set of all places X of a number field K . Under the classical analogy between number fields and function fields, Arakelov vector bundles over X correspond to algebraic vector bundles over a smooth projective curve C . The text discusses analogues over X of three results on such bundles over C : Riemann-Roch, the Narasimhan-Seshadri theorem, and Faltings' criterion for semistability.

<i>A conjectural non-commutative generalization of a volume formula of McMullen-Schneider</i>	
TOBIAS FINIS & EREZ LAPID	59

The Fourier transform of the characteristic function of a convex polytope is given by a localization formula due to Brion [2]. On the other hand, an argument by McMullen-Schneider expresses the mixed volume of d polytopes in a d -dimensional space in terms of volumes of parallelotopes formed by d -tuples of edges, one from each polytope [4]. When we specialize these formulas we get two seemingly different expressions for the volume of a polytope. We propose a non-commutative conjectural generalization of this identity. In the case of a root zonotope, such an identity is known and used in the study of the spectral side of Arthur's trace formula [3].

<i>The j-invariant of a plane tropical cubic</i>	
HANNAH MARKWIG (JOINT WITH ERIC KATZ AND THOMAS MARKWIG)	65

In tropical geometry, algebraic varieties are replaced by certain degenerations called tropical varieties. Tropical varieties are piece-wise linear objects that can be studied using combinatorics and linear algebra methods. We can use tropical geometry to prove theorems about algebraic geometry by means of these new methods. To reach this aim, we have to understand the connection between algebraic geometry and tropical geometry. In this talk, we want to understand the tropical analogue of the j -invariant of a smooth elliptic curve.

<i>Leopoldt's conjecture for some galois extensions</i>	
PREDA MIHĂILESCU	69

In this paper we use Baker theory for proving some special cases of Leopoldt's conjecture. Thus we show that if the conjecture is true for some field \mathbb{K} , then it is true for solvable extensions thereof. Also, if it holds for arbitrary extensions with simple groups, then it holds for any number field. The proofs are essentially based on the talk given at the SANT Conference.

<i>On Leopoldt's conjecture and a special case of Greenberg's conjecture</i>	
PREDA MIHĂILESCU	79

The conjecture of Leopoldt states that the p -adic regulator of a number field does not vanish. It was proved for the abelian case in 1967 by Brumer, using Baker theory. We show that when \mathbb{K} is a totally real extension in which the prime p is totally split, then Leopoldt's conjecture is equivalent to a special case of Greenberg's conjecture, which we prove. This case allows thus

a proof which does not use transcendence theory. At the end of the paper we give a new approach for the general case of Leopoldt's conjecture, which is currently under peer review. The two papers in these Proceedings gather various approaches for attacking the conjecture of Leopoldt.

Homological algebra for Schwartz algebras

RALF MEYER 99

Let G be a reductive group over a non-Archimedean local field. For two tempered smooth representations, it makes no difference for the Ext-groups whether we work in the category of tempered smooth representations of G or of all smooth representations of G . Similar results hold for certain discrete groups. We explain the basic ideas from functional analysis and geometric group theory that are needed to state this result correctly and prove it.

Weights in generalizations of Serre's conjecture and the mod p local Langlands correspondence

MICHAEL M. SCHEIN115

In this mostly expository article we give a survey of some of the generalizations of Serre's conjecture and results towards them that have been obtained in recent years. We also discuss recent progress towards a mod p local Langlands correspondence for p -adic fields and its connections with Serre's conjecture. A theorem describing the structure of some mod p Hecke algebras for GL_n is proved.

Towards Langlands correspondence over function fields for split reductive groups

YAKOV VARSHAVSKY (JOINT WITH DAVID KAZHDAN)139

In this note I will describe our joint work with David Kazhdan on the global Langlands correspondence over function fields for arbitrary split reductive groups.

Our main result asserts that for every pair (π, ω) , where π is a cuspidal representation of G one of whose local components is a cuspidal Deligne-Lusztig representation, and ω is a representation of the dual group, there exists a virtual Galois representation $\rho_{\pi, \omega}$, whose L -function equals the L -function of the pair (π, ω) .

<i>Maximal Sobolev regularity at radial points</i>	
INGO WITT	149

We establish precise regularity results for solutions to pseudodifferential equations with real principal symbols near radial points, microlocally in the scale of H^s Sobolev spaces. There is the new phenomenon of *maximal Sobolev regularity*. We clarify this point by discussing how the two distinguished parametrices that one has near a radial point, where the pseudodifferential operator under study is of real principal type, extend into that radial point, and we provide an analytic formula for maximal Sobolev regularity in terms of the principal and subprincipal symbols.

<i>Elliptic gamma function provides the Čech cocycle of a gerbe</i>	
CHENCHANG ZHU	161

G. Felder and A. Varchenko discovered certain modular formulas for elliptic gamma functions. These identities are generalized to an infinite set of identities for elliptic gamma functions associated to pairs of planes in 3-dimensional space in [2]. There we also use the language of stacks and gerbes to give a natural framework for a systematic description of these identities and their domain of validity. In this note I summarize the work in [2] with an emphasize on the Čech open covers.

<i>The affine Macdonald's formula</i>	
DAVID KAZHDAN	169

In a paper with A. Braverman we defined the affine spherical Hecke algebra. It is natural then to ask about the spherical eigenfunctionals. In the finite-dimensional case they are described by MacDonald. In this note we will talk about their affine analogue.

<i>Hermann Weyl and the Early History of Gauge Theories</i>	
NORBERT STRAUMANN	173

One of the major developments of twentieth century physics has been the gradual recognition that a common feature of the known fundamental interactions is their gauge structure. In this talk the early history of gauge theory is reviewed, emphasizing especially Weyl's seminal contributions of 1918 and 1929.

SELF-SIMILARITY IN GROUP THEORY AND ALGEBRA

Laurent Bartholdi

Mathematisches Institut, Georg-August Universität Göttingen
E-mail: laurent@uni-math.gwdg.de

Abstract. We describe avatars of self-similarity in the worlds of groups, associative and Lie algebras; in each case, we highlight some important examples and the questions or conjectures that they answer.

Humans have since their origins been fascinated with symmetry. To quote Pierre de la Harpe [18, page 3], it is “one way of coping with the frustrations of life’s limitations: we like to recognize symmetries which allow us to grasp more than what we can see”.

Symmetry is most conveniently expressed by a group of symmetries, and the bigger the group, the more symmetries one has at hand. Uncountable symmetry groups have two extremes (cf. Figure 1).

We shall be concerned in this text with objects whose symmetry group is totally disconnected; a good model to keep in mind is the fern; symmetries include the exchange of two neighbouring twigs, and all the replicas of this operation on sub-ferns.

November 2008.

This is an extended abstract of a talk given in the “Symmetries in Algebra and Number Theory” conference in Göttingen, October 27–31, 2008.



FIGURE 1. Spaces with uncountable connected (left) and totally disconnected (right) symmetry group.

1. Self-similar sets

A *self-similar set* is a set X with a family Σ of maps $X \rightarrow X$.

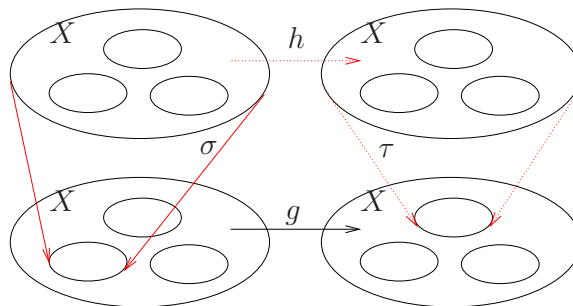
In interesting examples, these maps are injective; their images cover X ; and their intersections have empty interior. Furthermore, X could be a metric space, and the maps could contract the metric.

A particularly simple model for X is a Cantor set, with Σ the set of maps from X to its top-level clopens. More generally, a self-similarity structure on X with appropriate restrictions gives rise to a *Markov encoding* of X .

2. Self-similar groups

A group G acting on a self-similar set X is *self-similar* [23] if for every $g \in G$ and $\sigma \in \Sigma$ there exist unique $\tau \in \Sigma$ and $h \in G$ with

$$(2.1) \quad g\sigma = \tau h :$$



The self-similarity of G is conveniently encoded in a map $\phi : G \times \Sigma \rightarrow \Sigma \times G$, defined by $\phi(g, \sigma) = (\tau, h)$ using the above condition. This map contains the same information as a map $G \rightarrow (\Sigma \times G)^\Sigma = \Sigma^\Sigma \times G^\Sigma$; in fact its image lies in the semi-direct product $G^\Sigma \rtimes \text{Sym}(\Sigma)$ with $\text{Sym}(\Sigma)$ acting by permutation of the factors on G^Σ . This group is the *wreath product* of G and Σ , sometimes written $G \wr \Sigma$. Therefore, the self-similarity of G is also encoded by a map

$$\phi : G \rightarrow G \wr \Sigma := G^\Sigma \rtimes \text{Sym}(\Sigma).$$

It is not hard to see that ϕ is in fact a group homomorphism.

Given a self-similar group G , one may forget the space X on which it acts, and reconstruct an action from the self-similarity map. This action will be on the set Σ^* of words over Σ ; it may be extended by continuity to an action on Σ^∞ .

Given $g \in G$ and $a_1 \dots a_n \in \Sigma^*$, one sets $g_0 = g$; and for $i = 1, \dots, n$ one computes in turn $\phi(g_{i-1}, a_i) = (b_i, g_i)$. Then one defines the action of g by

$$g \cdot a_1 \dots a_n := b_1 \dots b_n.$$

2.1. Grigorchuk's group. One may proceed in reverse: one starts by a self-similarity map on a well-known group F , for example a free group; and defines a new group as the quotient of F that acts faithfully on Σ^* .

For example, let $F = \langle a, b, c, d \rangle$ be a free group of rank 4; take $\Sigma = \{0, 1\}$; and define $\phi : F \times \Sigma \rightarrow \Sigma \times F$ by

$$\begin{array}{llll} a0 \mapsto 1, & b0 \mapsto 0a, & c0 \mapsto 0a, & d0 \mapsto 0, \\ a1 \mapsto 0, & b1 \mapsto 1c, & c1 \mapsto 1d, & d1 \mapsto 1b. \end{array}$$

Let \mathcal{G} be the subgroup of $\text{Sym}(\Sigma^*)$ defined by ϕ . It is the first *Grigorchuk group* [14]. For an in-depth introduction to \mathcal{G} , see [18, Chapter 8].

2.2. Burnside's problem. Recall that a *torsion group* is one in which all elements have finite order; in a *p-group*, all elements furthermore have order a power of p . Burnside [10] asked in 1902 whether every finitely generated torsion group is finite, and this question has stimulated much research in group theory during the 20th century.

Variants of Burnside's question have been considered: the "Bounded Burnside Problem" asks whether there exists an infinite finitely generated group G and a bound m such that every element of G has torsion of exponent dividing m . For $m \leq 4$ or $m = 6$ this is impossible; while already the case $m = 5$ is open. On the other

hand, Adyan [1] (improving on [24] showed that for $m = 665$ the group could be infinite.

The “Restricted Burnside Problem” asks whether, for given k, m , there are finitely many finite groups with k generators and exponent m ; or, equivalently, whether there is a largest such group. This question was answered positively by Zelmanov [32, 33].

Golod [13] constructed in 1964 examples of infinite p -groups, settling Burnside’s original question. These examples are not easily manageable: Golod starts by an enumeration a_1, a_2, \dots of all elements without constant term of the associative algebra $\mathcal{A} = \mathbb{F}_p\{x, y\}$; he then considers $\mathcal{B} = \mathcal{A} / (a_i^{p^{n_i}})$, where the n_i are sufficiently fast growing that \mathcal{B} is infinite-dimensional — this is where the “Golod-Shafarevich condition” comes into play. The subgroup G of \mathcal{B}^\times generated by $\{1 + x, 1 + y\}$ is an infinite p -group.

Alëshin then produced [2] a more concrete family of torsion groups. The Grigorchuk group \mathcal{G} is, up to finite index inclusions, the first term in this family:

Theorem 2.1 (Grigorchuk [14]). *The Grigorchuk group \mathcal{G} is an infinite, finitely generated torsion group.*

Note that \mathcal{G} has unbounded exponent, by Zelmanov’s result, since it is residually finite. Other, similar torsion groups were constructed by Gupta and Sidki [17], but in this text we shall concentrate for simplicity on \mathcal{G} .

2.3. Growth of groups. Let G be a finitely generated group, with generating system S . The *growth function* of G is the function

$$\gamma(n) = \#\{g \in G \mid g = s_1 \dots s_m \text{ for } m \leq n \text{ and } s_i \in S\}.$$

The study of growth for finitely generated groups was introduced in the 1950’s in the former USSR [29] and in the 1960’s in the West [21]; its first appearance seems to be [20].

Note that, although γ depends on the choice of S , its asymptotic behaviour does not. A group has *polynomial growth* if $\gamma(n)$ is asymptotically polynomial; it has *exponential growth* if $\gamma(n)$ is asymptotically exponential.

It is easy to see that abelian groups have polynomial growth; for example, the growth of \mathbb{Z}^d is equivalent to n^d . The same holds for nilpotent groups. Gromov proved conversely:

Theorem 2.2 (Gromov [16]). *A finitely generated group has polynomial growth if and only if it contains a nilpotent subgroup of finite index.*

On the other hand, free groups, or more generally fundamental groups of negatively curved compact manifolds, have exponential growth.

There are many classes of groups for which the following alternative holds: either a group has polynomial growth, or it has exponential growth. This is the case for example for linear [30] and word-hyperbolic [12] groups.

Milnor asked in 1968 whether there exist groups of intermediate growth [22]. This was answered positively in the early 1980's, again by Grigorchuk:

Theorem 2.3 ([15]). *The Grigorchuk group \mathcal{G} has intermediate growth; more precisely,*

$$e^{n^\alpha} \lesssim \gamma(n) \lesssim e^{n^\beta},$$

with $\alpha = 0.515$ and $\beta = \log(2)/\log(2/\eta) \approx 0.767$, for $\eta \approx 0.811$ the real root of the polynomial $t^3 + t^2 + t - 2$.

The upper bound appeared in [4], and the lower bound appeared in [5]. I conjecture that the upper bound is sharp.

2.4. Dynamical systems. Apart from important examples such as the Grigorchuk group \mathcal{G} , there is a general construction giving a self-similar group, due to Nekrashevych [23].

Let \mathcal{M} be a topological space, and let $f : \mathcal{M} \rightarrow \mathcal{M}$ be a branched covering of \mathcal{M} , with critical locus C . As a simple but good example, consider $\mathcal{M} = \widehat{\mathbb{C}}$ and $f = z^2 - 1$, with $C = \{0, \infty\}$. Set

$$\mathcal{M}' = \mathcal{M} \setminus \bigcup_{n \geq 1} f^n(C),$$

and choose a basepoint $* \in \mathcal{M}'$. Then $\pi_1(\mathcal{M}', *)$ acts by monodromy on

$$T := \bigsqcup_{n \geq 0} f^{-n}(*).$$

To compute effectively this action, let $\Sigma = \{\sigma_x\}$ be a collection of paths in \mathcal{M}' from $*$ to $x \in f^{-1}(*)$. For every $g \in \pi_1(\mathcal{M}', *)$ and every $\sigma_x \in \Sigma$, there exists a unique f -lift \tilde{g} of g starting at x ; this lift will end at some $y \in f^{-1}(*)$. Define then

$$\phi(g, \sigma_x) = (\sigma_y, \sigma_y^{-1} g \sigma_x)$$

with concatenation of paths written in the right-to-left order. This defines a self-similarity structure on $\pi_1(\mathcal{M}', *)$, and transfers its action on T to an action on Σ^* .

I am grateful to Professor Stuhler for having suggested the following, more algebraic approach during my talk. One sets $X = \widetilde{\mathcal{M}'}$ the universal cover, on which $G = \pi_1(\mathcal{M}', *)$ acts by deck transformations; and one lets Σ be the branches of f at $*$, extended by analytic continuation to maps $X \rightarrow X$. Then for every $\sigma \in \Sigma$ and every $g \in G$ there are unique $\sigma' \in \Sigma$ and $g' \in G$ with $\sigma g = g' \sigma'$; this defines a self-similarity structure on G . Note however that the arrows for this self-similarity structure are in the opposite direction as in (2.1).

The *iterated monodromy group* of f is the quotient of $\pi_1(\mathcal{M}', *)$ acting faithfully on T . It serves both as a good source of examples (the group associated with $f = z^2 - 1$ is an example of amenable group that does not belong to any of the previously known classes [8]) and as a powerful invariant for the dynamical system defined by f up to isotopy [7]; this was crucial in solving an open problem by Hubbard [11].

3. Self-similar associative algebras

To move from groups to algebras, we may simply linearize the definition of a self-similar group: a *self-similar associative algebra* is an algebra \mathcal{A} , a vector space V , and a linear map

$$\phi : \mathcal{A} \otimes V \rightarrow V \otimes \mathcal{A},$$

called its *self-similarity map*. One may also write ϕ as an algebra homomorphism

$$\phi : \mathcal{A} \rightarrow \mathcal{A} \otimes \text{End}(V) = M_d(\mathcal{A}),$$

if V has dimension d . In a different language, such algebras were first considered by Sidki [28].

Let $\varepsilon : \mathcal{A} \rightarrow \mathbb{k}$ be a homomorphism (an *augmentation*). Then \mathcal{A} acts by linear transformations on the tensor algebra $T(V) = \bigoplus_{n \geq 0} V^{\otimes n}$; the action is defined as follows. Consider $a \in \mathcal{A}$ and $v_1 \otimes \cdots \otimes v_n \in T(V)$. Set $a_0 = a$; for $i = 1, \dots, n$ compute $\phi(a_{i-1} \otimes v_i) = \sum w_i \otimes a_i$; then

$$a \cdot v_1 \otimes \cdots \otimes v_n := \sum \sum w_1 \otimes \cdots \otimes w_n \varepsilon(a_n).$$

Self-similar algebras may be constructed as follows: let G be a self-similar group, with map $\phi : G \times \Sigma \rightarrow \Sigma \times G$. Let \mathbb{k} be a field. Set $\mathcal{A} = \mathbb{k}G$ and $V = \mathbb{k}\Sigma$. Then the linear extension of ϕ defines a self-similar associative algebra structure on \mathcal{A} .

Note that, even if the map ϕ is injective (as for example for a self-similar group *defined* by its self-similarity map), its linear extension may well fail to be injective.

As a concrete example, consider the associative algebra $\mathbb{k}G$ associated with the Grigorchuk group. Its self-similarity map may be defined by

$$a \mapsto \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad b \mapsto \begin{pmatrix} a & 0 \\ 0 & c \end{pmatrix}, \quad c \mapsto \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}, \quad d \mapsto \begin{pmatrix} 1 & 0 \\ 0 & b \end{pmatrix}.$$

Note now that $x := (d-1)a(d-1) \neq 0 \in \mathbb{k}G$, but $\phi(x) = 0$.

Following Sidki, we call the quotient $\mathcal{A}(G)$ of $\mathbb{k}G$ acting faithfully on $T(V)$ the *thinned algebra* of G . It is the self-similar algebra defined by the linear extension of G 's self-similarity map.

3.1. Goodearl's conjecture. Let now $\mathcal{A} = \mathcal{A}(\mathcal{G})$ be the thinned algebra of the Grigorchuk group.

Theorem 3.1 ([3]). *The algebra \mathcal{A} has quadratic growth, and therefore has Gelfand-Kirillov dimension 2. It contains a copy of Laurent polynomials $\mathbb{k}[t, t^{-1}]$.*

Assume that \mathbb{k} has characteristic 2. Then \mathcal{A} is graded, and its homogeneous elements are nil.

Assume furthermore that \mathbb{k} is an algebraic extension of \mathbb{F}_2 . Then the Jacobson radical of \mathcal{A} coincides with its augmentation ideal

$$\bar{\omega}(\mathcal{A}) = \ker(\varepsilon) = \langle a-1, b-1, c-1, d-1 \rangle.$$

(A finitely generated algebra has quadratic growth if the dimension of the span of $(\leq n)$ -fold products of generators grows as a quadratic function of n ; the Jacobson radical is the intersection of all maximal right ideals, and can be defined as those $x \in \mathcal{A}$ for which $1 + \alpha x \beta$ is invertible for all $\alpha, \beta \in \mathcal{A}$).

It turns out that these features contradict a conjecture attributed to Goodearl [9, Conjecture 3.1]. This conjecture stated that if \mathcal{A} is an algebra over a field and has Gelfand-Kirillov dimension 2, then its Jacobson radical is nil.

4. Self-similar Lie algebras

We finally consider self-similarity for Lie algebras. Let \mathcal{L} be a Lie algebra, and let \mathcal{C} be a commutative algebra. A *self-similar Lie algebra* is given by a self-similarity map

$$\phi : \mathcal{L} \rightarrow \mathcal{L} \wr \mathcal{C} := \mathcal{L} \otimes \mathcal{C} \rtimes \text{Der } \mathcal{C}.$$

In that case, \mathcal{L} acts by linear maps on $T(\mathcal{C})$; the action is defined as follows. First, $\mathcal{L} \cdot 1 = 0$. Consider then $l \in \mathcal{L}$ and $c \otimes d \in T(\mathcal{C})$, with $c \in \mathcal{C}$ and $d \in T(\mathcal{C})$. Compute

$\phi(l) = \sum l_i \otimes c_i + \partial$; then inductively set

$$l \cdot (c \otimes d) := \sum c_i c \otimes (l_i \cdot d) + \partial(c) \otimes d.$$

Again one may *construct* a Lie algebra by giving a self-similarity map on a free algebra, and by considering its quotient acting faithfully on $T(\mathcal{C})$.

Let \mathcal{L} be a self-similar Lie algebra. Then the associative envelope $\mathcal{A}(\mathcal{L})$ of its action on $T(\mathcal{C})$ is a self-similar associative algebra. Concretely, define $\phi : \mathcal{A}(\mathcal{L}) \rightarrow \mathcal{A}(\mathcal{L}) \otimes \text{End}_{\mathbb{k}}(\mathcal{C})$ by

$$\phi(l) = \sum \ell_i \otimes M_{c_i} + 1 \otimes \partial,$$

where M_c denotes the \mathbb{k} -endomorphism of \mathcal{C} defined by multiplication by c , and where $\phi(l) = \sum l_i \otimes c_i + \partial$ holds in the Lie algebra \mathcal{L} .

4.1. Zelmanov's conjecture. Let \mathbb{k} be a field of characteristic 2, and set $\mathcal{C} = \mathbb{k}[x]/(x^2)$. Consider the Lie algebra $\mathcal{L} = \langle A, B, C, D \rangle$, defined by its self-similarity map

$$A \mapsto \frac{\partial}{\partial x}, \quad B \mapsto A + C(1 + x), \quad C \mapsto A + D(1 + x), \quad D \mapsto B(1 + x).$$

Theorem 4.1. *The Lie algebra \mathcal{L} is graded, and has linear growth; more precisely, its homogeneous components of degree ≥ 2 have dimension $\in \{1, 2\}$.*

The associative envelope of $\mathcal{A}(\mathcal{L})$ of \mathcal{L} is the associative algebra $\mathcal{A} = \mathcal{A}(\mathcal{G})$ from §3.1. The Lie algebra \mathcal{L} coincides with the Lie algebra associated with the dimension series of the Grigorchuk group \mathcal{G} from §2.1.

Recall that the dimension series of a p -group G is the sequence of subgroups defined by $G_1 = G$ and $G_n = [G, G_{n-1}](G_{[n/p]})^p$ for $n \geq 2$. Commutation in G gives rise to a Lie algebra structure on $\mathcal{L}(G) := \bigoplus_{n \geq 1} \mathbb{k} \otimes_{\mathbb{F}_p} (G_n/G_{n+1})$. The structure of $\mathcal{L}(\mathcal{G})$ was computed in [6], based on ideas in [26].

The isomorphism between $\mathcal{A}(\mathcal{G})$ and $\mathcal{A}(\mathcal{L})$ is as follows: a corresponds to $A + 1$, etc.; the basis $\{0, 1\}$ of V corresponds to the basis $\{1, 1 + x\}$ of \mathcal{C} .

Lie algebras associated to groups have been extensively studied [19, 31]; a group has *finite width* if the ranks in the Lie algebra above are bounded. A conjecture attributed to Zelmanov [34] asserted that groups of finite width had to fall in precise classes: solvable groups, p -adic analytic groups, or (up to finite index) the positive part of a loop group or the Nottingham group $\text{Aut}^1(\mathbb{F}_p[[t]])$.

This conjecture is disproved by \mathcal{G} and \mathcal{L} , and the structure of groups of finite width is therefore considerably more complicated than one might have expected.

4.2. Petrogradsky-Shestakov-Zelmanov Lie algebras. It might have naively been believed that, if G is a p -torsion self-similar group (such as the Grigorchuk group) and \mathbb{k} is a field of characteristic p , then $\mathcal{O}(\mathcal{A}(G))$ is a nil ring. This is certainly not true, as we saw in §3.1. However, Shestakov and Zelmanov (following ideas by Petrogradsky [25]) construct non-nilpotent nil rings from Lie algebras, as follows.

Let \mathbb{k} be a field of characteristic p , and set $\mathcal{C} = \mathbb{k}[x]/(x^p)$. Define the self-similar Lie algebra $\mathfrak{L} = \langle t, u, v \rangle$ by

$$t \mapsto \frac{\partial}{\partial x}, \quad u \mapsto vx^{p-1} + t, \quad v \mapsto u.$$

Theorem 4.2 ([27]). *The subalgebra $\langle u, v \rangle$ of \mathfrak{L} is a finitely generated, nil Lie algebra, but it is not nilpotent. It has Gelfand-Kirillov dimension $\in (1, 2)$.*

The associative envelope of \mathfrak{L} may be described quite easily: its self-similarity map is given by $p \times p$ -matrices

$$t \mapsto \begin{pmatrix} 0 & \cdots & 0 & 1 \\ \vdots & \ddots & & 0 \\ \vdots & & \ddots & \\ 0 & \cdots & \cdots & 0 \end{pmatrix}, \quad u \mapsto \begin{pmatrix} t & 0 & \cdots & 0 \\ 0 & t & \ddots & \vdots \\ \vdots & \ddots & t & 0 \\ u & \cdots & 0 & t \end{pmatrix}, \quad v \mapsto \begin{pmatrix} u & 0 & \cdots & 0 \\ 0 & u & \ddots & \vdots \\ \vdots & \ddots & u & 0 \\ 0 & \cdots & 0 & u \end{pmatrix}.$$

References

- [1] Sergei I. Adyan, *The burnside problem and identities in groups*, Springer-Verlag, Berlin. Translated from the Russian by John Lennox and James Wiegold.
- [2] Sergei V. Alešin, *Finite automata and the burnside problem for periodic groups*, *Mat. Zametki* **11** (1972), 319–328.
- [3] Laurent Bartholdi, *Branch rings, thinned rings, tree enveloping rings*, *Israel J. Math.* **154** (2006), 93–139, available at [arXiv.org/math.RA/0410226](https://arxiv.org/math.RA/0410226). MR 2254535
- [4] ———, *The growth of grigorchuk's torsion group*, *Internat. Math. Res. Notices* (1998), no. 20, 1049–1054, available at [arXiv.org/abs/math/0012108](https://arxiv.org/abs/math/0012108). MR 1656258 (99i:20049)
- [5] ———, *Lower bounds on the growth of a group acting on the binary rooted tree*, *Internat. J. Algebra Comput.* **11** (2001), no. 1, 73–88, available at [arXiv.org/abs/math/9910068](https://arxiv.org/abs/math/9910068). MR 1818662 (2001m:20044)
- [6] Laurent Bartholdi and Rostislav I. Grigorchuk, *Lie methods in growth of groups and groups of finite width*, 2000, pp. 1–27. MR 1776763 (2001h:20046)
- [7] Laurent Bartholdi and Volodymyr V. Nekrashevych, *Thurston equivalence of topological polynomials*, *Acta Math.* **197** (2006), no. 1, 1–51, available at [arXiv.org/math.DS/0510082](https://arxiv.org/math.DS/0510082).
- [8] Laurent Bartholdi and Bálint Virág, *Amenability via random walks*, *Duke Math. J.* **130** (2005), no. 1, 39–56, available at [arXiv.org/abs/math/0305262](https://arxiv.org/abs/math/0305262). MR 2176547 (2006h:43001)
- [9] Jason P. Bell, *Examples in finite gelfand-kirillov dimension*, *J. Algebra* **263** (2003), no. 1, 159–175. MR1974084 (2004d:16042)

- [10] William Burnside, *On an unsettled question in the theory of discontinuous groups*, Quart. J. Pure Appl. Math. **33** (1902), 230–238.
- [11] Adrien Douady and John H. Hubbard, *A proof of thurston's topological characterization of rational functions*, Acta Math. **171** (1993), no. 2, 263–297. MR **1251582 (94j:58143)**
- [12] Étienne Ghys and Pierre de la Harpe, *Sur les groupes hyperboliques d'après mikhael gromov*, Progress in Mathematics, vol. 83, Birkhäuser Boston Inc., Boston, MA. Papers from the Swiss Seminar on Hyperbolic Groups held in Bern, 1988.
- [13] Evguenii S. Golod, *On nil-algebras and finitely approximable p -groups*, Izv. Akad. Nauk SSSR Ser. Mat. **28** (1964), 273–276.
- [14] Rostislav I. Grigorchuk, *On burnside's problem on periodic groups*, Функционал. Анал. и Приложен. **14** (1980), no. 1, 53–54. English translation: Functional Anal. Appl. **14** (1980), 41–43.
- [15] ———, *On the milnor problem of group growth*, Dokl. Akad. Nauk SSSR **271** (1983), no. 1, 30–33.
- [16] Mikhael L. Gromov, *Groups of polynomial growth and expanding maps*, Inst. Hautes Études Sci. Publ. Math. (1981), no. 53, 53–73.
- [17] Narain D. Gupta and Said N. Sidki, *On the burnside problem for periodic groups*, Math. Z. **182** (1983), 385–388.
- [18] Pierre de la Harpe, *Topics in geometric group theory*, University of Chicago Press, Chicago, IL.
- [19] Gunde Klaas, Charles R. Leedham-Green, and Wilhelm Plesken, *Linear pro- p -groups of finite width*, Lecture Notes in Mathematics, vol. 1674, Springer-Verlag, Berlin.
- [20] Hans Ulrich Krause, *Gruppenstruktur und gruppenbild*, Thesis, Eidgenössische Technische Hochschule, Zürich.
- [21] John W. Milnor, *Growth of finitely generated solvable groups*, J. Differential Geom. **2** (1968), 447–449.
- [22] ———, *Problem 5603*, Amer. Math. Monthly **75** (1968), 685–686.
- [23] Volodymyr V. Nekrashevych, *Self-similar groups*, Mathematical Surveys and Monographs, vol. 117, Amer. Math. Soc., Providence, RI, 2005.
- [24] Pet' r S. Novikov and Sergei I. Adyan, *Infinite periodic groups. i,ii,iii*, Izv. Akad. Nauk SSSR Ser. Mat. **32** (1968), 212–244,251–254,709–731.
- [25] V. M. Petrogradsky, *Examples of self-iterating lie algebras*, J. Algebra **302** (2006), no. 2, 881–886. MR **2293788 (2008b:17029)**
- [26] Alexander V. Rozhkov, *Lower central series of a group of tree automorphisms*, Mat. Zametki **60** (1996), no. 2, 225–237, 319.
- [27] I. P. Shestakov and E. Zelmanov, *Some examples of nil lie algebras*, J. Eur. Math. Soc. (JEMS) **10** (2008), no. 2, 391–398. MR 2390328
- [28] Said N. Sidki, *A primitive ring associated to a burnside 3-group*, J. London Math. Soc. (2) **55** (1997), no. 1, 55–64.
- [29] Albert S. Švarc, *A volume invariant of coverings*, Dokl. Akad. Nauk SSSR (1955), no. 105, 32–34 (Russian).
- [30] Jacques Tits, *Free subgroups in linear groups*, J. Algebra **20** (1972), 250–270.
- [31] Efim I. Zel'manov, *Lie ring methods in the theory of nilpotent groups*, Groups '93 galway/st. andrews, vol. 2, 1995, pp. 567–585.
- [32] ———, *Solution of the restricted burnside problem for 2-groups*, Mat. Sb. **182** (1991), no. 4, 568–592.
- [33] ———, *Solution of the restricted burnside problem for groups of odd exponent*, Izv. Akad. Nauk SSSR Ser. Mat. **54** (1990), no. 1, 42–59, 221.
- [34] ———, *Talk at the esf conference on algebra and discrete mathematics "group theory: from finite to infinite"*. Castelveccchio Pascoli.

***L*-FUNCTIONS, AUTOMORPHIC FORMS, AND ARITHMETIC**

Valentin Blomer

Department of Mathematics, University of Toronto
E-mail: vblomer@math.toronto.edu

Gergely Harcos

Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences
E-mail: gharcos@renyi.hu

Abstract. We give a short, informal survey on the role of automorphic L -functions in number theory. We present the strongest currently known subconvexity bounds for twisted L -functions over number fields due to the authors and give various arithmetic applications. This is based on a talk of the first author.

1. L -functions

Suppose you are given an interesting sequence $a(n)$, $n \in \mathbb{N}$, of complex numbers that you would like to investigate. The method of analytic number theory is to encode this sequence in a generating function. There are several choices, and if some multiplicativity is involved, one might consider the Dirichlet series

$$(1) \quad L(s) = \sum_{n=1}^{\infty} \frac{a(n)}{n^s}.$$

If we assume $a(n) \ll_{\varepsilon} n^{\varepsilon}$ for all $\varepsilon > 0$, or even only an average bound $\sum_{n \leq x} |a(n)| \ll_{\varepsilon} x^{1+\varepsilon}$, then (1) converges absolutely and uniformly on compacta in $\Re s > 1$, and thus defines a holomorphic function. We can hope that in this way we translate the arithmetic of the sequence $a(n)$ into analytic properties of the function $L(s)$, and

2000 Mathematics Subject Classification. Primary: 11-02, Secondary: 11F66, 11E25.

The first author was supported by an NSERC grant and the second author was supported by European Community grant MEIF-CT-2006-040371 under the Sixth Framework Programme and by OTKA grants K 72731 and PD 75126.

indeed there is very often a remarkable interplay between arithmetic and analysis. Let us look at a few examples (see also [21, 27, 32]):

1) Let $a(n) = 1$ for all n . While it is debatable if this is an interesting sequence, it gives no doubt an interesting object: the Riemann ζ -function

$$\zeta(s) := \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_p \left(1 - \frac{1}{p^s}\right)^{-1}, \quad \Re s > 1.$$

The Euler product shows that $\zeta(s) \neq 0$ in $\Re s > 1$, and it is a classical fact that the non-vanishing of ζ on the line $\Re s = 1$ is equivalent⁽¹⁾ to the prime number theorem

$$\pi(x) := \#\{p \leq x \mid p \text{ prime}\} \sim \frac{x}{\log x}, \quad x \rightarrow \infty.$$

The ζ -function can be extended meromorphically to all of \mathbb{C} , and one has more precisely an equivalence⁽²⁾

$$\begin{aligned} & \zeta(s) \neq 0 \text{ in } \Re s > 1 - \delta \text{ for some } 0 < \delta \leq 1/2 \\ \iff & \pi(x) = \int_2^x \frac{dt}{\log t} + O(x^{1-\delta+\varepsilon}) \text{ for some } 0 < \delta \leq 1/2 \text{ and all } \varepsilon > 0. \end{aligned}$$

This shows a very precise translation of an arithmetic statement (distribution of prime numbers) into an analytic statement (location of zeros).

2) Let K/\mathbb{Q} be a number field and let $a(n) := \#\{\text{integral ideals } \mathfrak{a} \mid \mathcal{N}\mathfrak{a} = n\}$. This gives the Dedekind ζ -function

$$\zeta_K(s) := \sum_{n=1}^{\infty} \frac{a(n)}{n^s} = \sum_{\mathfrak{a}} \frac{1}{(\mathcal{N}\mathfrak{a})^s},$$

which has a simple pole at $s = 1$. The analytic class number formula states

$$\operatorname{res}_{s=1} \zeta_K(s) = \frac{2^{r_1} (2\pi)^{r_2} R h}{w \sqrt{|D|}},$$

where as usual r_1, r_2 are the number of real resp. pairs of complex embeddings of K into \mathbb{C} , R is the regulator, h is the class number, w is the number of roots of unity in K and D is the discriminant. In other words, we find all algebraic invariants of K in the Laurent expansion of ζ_K at $s = 1$.

⁽¹⁾Of course, since both statements are true, they are in particular equivalent. But even without knowing the truth of either of these statements one can deduce one from the other.

⁽²⁾Here it is unknown if either of these statement holds for some $\delta > 0$.

3) Let χ be a primitive Dirichlet character to some large modulus q . This gives rise to a Dirichlet L -function

$$L(s, \chi) := \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s},$$

which again can be continued to an entire function. In practice, one often encounters character sums of the type $\sum_{n \leq x} \chi(n)$, and one would expect that there is a lot of cancellation in such a sum. For example, for $x = q$ one even has $\sum_{n \leq q} \chi(n) = 0$. Cancellation becomes a very delicate matter if the sum is short, i.e. if x is small compared to q . The Lindelöf hypothesis for $L(s, \chi)$ states that $L(1/2 + it, \chi) \ll_{\varepsilon} ((1 + |t|)q)^{\varepsilon}$ for all $\varepsilon > 0$. This is not known, but it would imply

$$\sum_{n \leq x} \chi(n) \ll_{\varepsilon} x^{1/2+\varepsilon}$$

for all $\varepsilon > 0$ and for all $x > 0$. Again there is an intimate connection between an arithmetic statement (equidistribution of character values to small arguments) and an analytic statement (growth on vertical lines).

4) If E/\mathbb{Q} is an elliptic curve, we know by Mordell's theorem that the set of rational points on E is a finitely generated Abelian group, $E(\mathbb{Q}) \cong \mathbb{Z}^r \oplus E_{\text{tor}}(\mathbb{Q})$. The rank r seems to be an elusive object; however, it is relatively simple to count points on (the reduction of) E over finite fields, and we can define

$$a_E(p) := \frac{p + 1 - \#E(\mathbb{F}_p)}{\sqrt{p}}$$

for a prime p . This can be extended in a more or less natural way to all integers, and yields an L -function $L_E(s) = \sum a_E(n)n^{-s}$. It is, in general, very hard to prove that this can be extended to an entire function, and is part of the seminal work of Wiles (and others) [9, 39, 43]. Given that $L_E(1/2)$ exists, the Birch and Swinnerton-Dyer conjecture states (among other things) that the rank can be recovered from the Laurent expansion at $1/2$, namely $\text{ord}_{s=1/2} L_E(s) = r$.

We observe that all four examples depend on an appropriate analytic continuation of the respective L -function and provide a connection between the arithmetic input and some analytic properties outside the region of absolute convergence.

Every decent L -function has a functional equation of the form

$$(2) \quad L(s)G(s) = \eta \overline{L(1-\bar{s})}G(1-s)$$

where $|\eta| = 1$ and

$$G(s) = N^{s/2} \prod_{j=1}^d \pi^{-s/2} \Gamma\left(\frac{s + \mu_j}{2}\right)$$

for some integer $N \in \mathbb{N}$ and some complex numbers μ_1, \dots, μ_d . The complexity of an L -function is measured by its *analytic conductor*

$$(3) \quad C := C(t) := N \prod_{j=1}^d (1 + |t + \mu_j|), \quad t = \Im s.$$

Since we are assuming that $L(s)$ converges absolutely in $\Re s > 1$, we have $L(s) \ll 1$ in $\Re s = 1 + \varepsilon$. The functional equation (2) and Stirling's formula translate this into⁽³⁾ $L(s) \ll C^{1/2+\varepsilon}$ on $\Re s = -\varepsilon$. If we assume in addition that L is of finite order (in the sense of complex analysis), which is always satisfied in applications, then a standard argument shows

$$(4) \quad L(1/2 + it) \ll C(t)^{1/4+\varepsilon}.$$

This is usually referred to as the *convexity bound*, and any exponent smaller than $1/4$ is called a *subconvexity bound*. If the generalized Riemann hypothesis holds for the L -function in question, then $1/4$ can be replaced with 0 .

2. Automorphic forms on GL_2

Let $G := \mathrm{PSL}_2(\mathbb{R})$. We have the Iwasawa decomposition $G = NAK$ where

$$N := \left\{ \begin{pmatrix} 1 & x \\ & 1 \end{pmatrix} \mid x \in \mathbb{R} \right\}, \quad A := \left\{ \begin{pmatrix} y^{1/2} & \\ & y^{-1/2} \end{pmatrix} \mid y > 0 \right\},$$

$$K := \left\{ \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \mid \theta \in [0, \pi) \right\}.$$

The group $K = \mathrm{PSO}_2(\mathbb{R})$ is a maximal compact subgroup of G . Let

$$\Gamma := \Gamma_0(N) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{PSL}_2(\mathbb{Z}) \mid c \equiv 0 \pmod{N} \right\}.$$

⁽³⁾Throughout this note, $\varepsilon > 0$ denotes an arbitrarily small constant, not necessarily the same on each occurrence.

Then G acts on $L^2(\Gamma \backslash G)$ by the right regular representation, $\rho(g)(\phi)(x) := \phi(xg)$ for $\phi \in L^2(\Gamma \backslash G)$, and we have a G -equivariant decomposition

$$(5) \quad L^2(\Gamma \backslash G) = \mathbb{C} \cdot \mathbf{1} \oplus \bigoplus_{\pi} V_{\pi} \oplus \sum_{\mathfrak{a}} \int_{\mathbb{R}} H_{\mathfrak{a}}(t) dt$$

into the constant functions, cuspidal irreducible representations (π, V_{π}) and Eisenstein series for the cusps \mathfrak{a} (that enter the picture because $\Gamma \backslash G$ is not compact). Each V_{π} decomposes further according to the characters of K :

$$V_{\pi} = \bigoplus_{q \in 2\mathbb{Z}} V_{\pi, q}$$

(in the Hilbert space sense), and it is known that $\dim V_{\pi, q} \leq 1$. The left and right G -invariant Laplace operator

$$\Delta := -y^2(\partial_x^2 + \partial_y^2) + y\partial_x\partial_{\theta}$$

acts (by a generalized version of Schur's lemma) on each V_{π} as a scalar $\lambda_{\pi} \in \mathbb{R}$. In algebraic terms, this is the Casimir element (up to normalization) of the universal enveloping algebra $U(\mathfrak{g})$. Sometimes we need a variant of the space $L^2(\Gamma \backslash G)$. For a character χ of modulus dividing N let $L^2(\Gamma \backslash G)$ denote the L^2 -space of functions $G \rightarrow \mathbb{C}$ that transform under Γ as $f\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} g\right) = \chi(d)f(g)$ for $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$.

Why is the space $L^2(\Gamma \backslash G)$ interesting? One reason is that it is equipped with additional structure. In general there is no left action of G on $L^2(\Gamma \backslash G)$: if f is Γ -invariant, then $f(g \cdot)$ is only $g^{-1}\Gamma g$ -invariant. However, if $\Gamma = \Gamma_0(N)$ and $g \in \mathrm{PSL}_2(\mathbb{Q})$, then $g^{-1}\Gamma g$ contains some finite index subgroup of $\Gamma_0(N)$, and using a suitable average, we can get back to $\Gamma_0(N)$. This is a special feature of groups like $\Gamma_0(N)$ (as opposed to arbitrary discrete subgroups of G) and yields a family of naturally defined operators $\{T_n \mid n \in \mathbb{N}\}$ that forms a commutative algebra, which also commutes with Δ , since Δ is left G -invariant. Mostly for technical reasons we consider only the subspace $L^2_{\mathrm{new}}(\Gamma \backslash G)$ whose irreducible representations are generated by so-called newforms, i.e. they do not come from subgroups with smaller index in $\mathrm{PSL}_2(\mathbb{Z})$. Then each operator T_n acts on each $V_{\pi} \subseteq L^2_{\mathrm{new}}(\Gamma \backslash G)$ as a scalar $\lambda_{\pi}(n)$, and a function $\phi \in V_{\pi, q}$ has a Fourier-Whittaker expansion

$$\begin{aligned} \phi \left(\begin{pmatrix} y^{1/2} & xy^{-1/2} \\ 0 & y^{-1/2} \end{pmatrix} \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \right) \\ = e^{iq\theta} \sum_{n \neq 0} \frac{\lambda_{\pi}(n)}{\sqrt{|n|}} W_{\mathrm{sgn}(n)q/2, \sqrt{1/4 - \lambda_{\pi}}}(4\pi|n|y) e(nx), \end{aligned}$$

where as usual $e(x)$ denotes the additive character $e^{2\pi i x}$, $W_{\alpha, \beta}$ is the Whittaker function⁽⁴⁾ [42, Chapter 16] and $\lambda(-n) = \eta \lambda(n)$ with $\eta \in \{-1, 0, 1\}$. This is relevant for us, because the Hecke eigenvalues $\lambda_\pi(n)$ carry often number theoretic information.

Examples. 1) Hecke [25] - Maaß [31]: Let N be the discriminant of the field $K = \mathbb{Q}(\sqrt{N})$ and χ the character of the extension K/\mathbb{Q} . Then there is a representation $(\pi, V_\pi) \subseteq L_{\text{new}}^2(\Gamma \backslash G, \chi)$ with $\lambda_\pi = 1/4$, such that $\lambda_\pi(n) = \#\{\text{ideals } \mathfrak{a} \subseteq \mathcal{O}_K \mid \mathcal{N} \mathfrak{a} = n\}$.

2) Wiles et al. [9, 39, 43]: Let E/\mathbb{Q} be an elliptic curve. Then there is a representation $(\pi, V_\pi) \subseteq L_{\text{new}}^2(\Gamma \backslash G)$ with $\lambda_\pi = 0$, such that $\lambda_\pi(p) = (p + 1 - \#E(\mathbb{F}_p))/\sqrt{p}$ for all primes p .

Langlands' philosophy suggests that "all interesting objects" arise in this way for suitable Γ and G . In any case, for each representation $(\pi, V_\pi) \subseteq L_{\text{new}}^2(\Gamma \backslash G)$ we can define an L -function

$$L(\pi, s) := \sum_{n=1}^{\infty} \frac{\lambda_\pi(n)}{n^s},$$

and we hope to learn more about Hecke eigenvalues by studying $L(\pi, s)$ from an analytic point of view.

If we work over a number field K/\mathbb{Q} with class number $h > 1$, the above setup is not appropriate. One could work with h copies of G modulo certain conjugates of Γ , but it is better to work adelicly. For each place v of K let K_v be the completion and \mathcal{O}_v the ring of integers (if $v \mid \infty$, then $\mathcal{O}_v = K_v$). Then the adèle ring is the restricted product

$$\mathbb{A} = \prod'_v K_v$$

with respect to the sets \mathcal{O}_v , with K embedded diagonally. There is a natural surjection from \mathbb{A}^\times to the group of non-zero fractional ideals of K , and we often do not distinguish between an idele and its image. Again $\text{GL}_2(\mathbb{A})$ acts by the right regular representation on $L^2(\mathbb{A}^\times \text{GL}_2(K) \backslash \text{GL}_2(\mathbb{A}))$ where \mathbb{A}^\times is identified with the center of $\text{GL}_2(\mathbb{A})$. This setting has no dependence on the level of the subgroup any more, since it treats simultaneously all subgroups $\Gamma_0(\mathfrak{c})$, with \mathfrak{c} an ideal in K . If we want to make the level explicit, we define

$$\mathcal{K}(\mathfrak{c}) := \prod_{\mathfrak{p} \text{ finite}} \mathcal{K}(\mathfrak{c}_{\mathfrak{p}}) \subseteq \text{GL}_2(\mathbb{A}_{\text{fin}})$$

⁽⁴⁾In the special case $q = 0$, this reduces essentially to a Bessel K -function.

for a nonzero ideal \mathfrak{c} , where

$$\mathcal{K}(\mathfrak{c}_p) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(K_p) \mid a, d \in \mathcal{O}_p, b \in \mathfrak{d}_p^{-1}, c \in \mathfrak{d}_p \mathfrak{c}, ad - bc \in \mathcal{O}_p^\times \right\}$$

with \mathfrak{d} the different of K . The conductor of an irreducible representation (π, V_π) contained in $L^2(\mathbb{A}^\times \mathrm{GL}_2(K) \backslash \mathrm{GL}_2(\mathbb{A}))$ is the smallest ideal such that V_π contains a right $K(\mathfrak{c})$ -invariant vector (which is automatically a “newvector”). The Fourier expansion in the adelic setting reads

$$\phi \left(\begin{pmatrix} y & x \\ & 1 \end{pmatrix} \right) = \sum_{r \in K^\times} \frac{\lambda_\pi(r y_{\mathrm{fin}})}{\sqrt{\mathcal{N}(r y_{\mathrm{fin}})}} W_\phi(r y_\infty) \psi(r x),$$

where $y = y_\infty \times y_{\mathrm{fin}} \in \mathbb{A}^\times$, $x \in \mathbb{A}$, ϕ a smooth vector in some cuspidal irreducible representation π , \mathcal{N} the norm, W_ϕ a product of Whittaker functions, ψ the standard additive character on \mathbb{A} , and $\lambda_\pi(r y_{\mathrm{fin}})$ depends only on the fractional ideal represented by $r y_{\mathrm{fin}}$ and is non-zero only if this ideal is integral.

How can we get new automorphic forms out of given ones? A typical way is twisting, and the simplest twist is by a character (that is, by an automorphic form on GL_1). Let $\chi : K^\times \backslash \mathbb{A}^\times \rightarrow S^1$ be a Hecke character of conductor \mathfrak{q} (that is, \mathfrak{q} is the largest ideal such that χ is trivial on finite ideles $\equiv 1 \pmod{\mathfrak{q}}$), and define the twist of a representation π on GL_2 with χ by

$$\pi \otimes \chi(g) := \chi(\det g) \pi(g).$$

This is another representation on GL_2 , and if the integral ideal \mathfrak{a} is coprime to the conductors of π and χ , then $\lambda_{\pi \otimes \chi}(\mathfrak{a}) = \lambda_\pi(\mathfrak{a}) \chi(\mathfrak{a})$. If χ has conductor \mathfrak{q} and π has conductor \mathfrak{c} coprime to \mathfrak{q} , then $\pi \otimes \chi$ has conductor $\mathfrak{c} \mathfrak{q}^2$, so the conductor of the character enters quadratically.

3. Subconvexity for automorphic L -functions

The following result is a combination of the results in [7, 6, 5]. We are interested in bounding a twisted automorphic L -functions in terms of the conductor of the twisting character, where the other parameters are essentially kept fixed.

Theorem 1. *Let K be a totally real number field of degree d , π an irreducible cuspidal representation on $\mathrm{GL}_2(\mathbb{A})$, χ a Hecke character of conductor \mathfrak{q} and $C = C(t, \pi)$ the analytic conductor of $L(s, \pi)$ in the sense of (3). Then*

$$L(\pi \otimes \chi, 1/2 + it) \ll C^A(\mathcal{N} \mathfrak{q})^{1/2 - \delta + \varepsilon}$$

where \mathcal{N} denotes the norm, A is some absolute constant, $\varepsilon > 0$ is arbitrarily small, and $\delta = \frac{1}{11}$ in general, and $\delta = \frac{1}{8}$ if $K = \mathbb{Q}$.

More precisely, the constant δ in the general case⁽⁵⁾ is $\frac{1}{8}(1 - 2\theta) > \frac{1}{11}$ where θ is the a bound towards the Ramanujan conjecture (currently $\theta \leq 1/9$ is known [29] and $\theta = 0$ is conjectured).

The convexity bound in this context is $L(\pi \otimes \chi, 1/2 + it) \ll_{t, \pi, \varepsilon} (\mathcal{N}q)^{1/2 + \varepsilon}$. The first subconvex bound in this direction was $\delta = 1/22$ for $K = \mathbb{Q}$ by Duke, Friedlander and Iwaniec [19] and an important contribution came from Bykovskii [12] that inspired both [16] and [6]. Our bound $\delta = 1/8$ matches the quality of Burgess' celebrated bound [11], where the case π an Eisenstein series is treated.

Over a number field K other than \mathbb{Q} a subconvexity bound was for a long time an open problem. In an unpublished manuscript [13] (see also [14]), Cogdell, Piatetskii-Shapiro and Sarnak obtained $\delta = 1/18$ for holomorphic Hilbert cusp forms using deep bounds for triple products [35]. As an application of an ingenious and very flexible geometric method, Venkatesh [40] (see also [34]) proved recently – among other things – Theorem 1 with $\delta = 1/24$. Our method is quite different from all of these works and will yield in particular as a by-product a solution of a problem of Selberg, see Theorem 2 below.

We will only sketch briefly the ideas that go into the proof; it rests on the following ingredients⁽⁶⁾:

- the amplification method [19] and an approximate functional equation [23] (this is essentially standard),
- the spectral decomposition of Dirichlet series associated with a shifted convolution sum [7] (solving a problem of Selberg) which makes good use of
- the Kirillov model and Sobolev norms [2].

Let us look at the first point. To start with, we have to find a way to work conveniently with the values $L(\pi \otimes \chi, 1/2 + it)$ since a priori they exist only by analytic continuation (or perhaps as a conditionally convergent series which is not useful in practice either). However, often a suitably truncated part of a divergent or conditionally convergent series gives a good approximation of the quantity that one

⁽⁵⁾At the time of writing, we need some technical assumptions that can most likely be removed with a little extra work.

⁽⁶⁾The proof for $K = \mathbb{Q}$ uses a somewhat different methods which avoid the dependence on Ramanujan bounds, see [6].

is interested in. For L -functions this can be made precise with an approximate functional equation, and in fact the first about $C^{1/2}$ terms of an L -function with analytic conductor C are a very good approximation on the critical line $\Re s = 1/2$. In other words, for all practical purposes

$$(6) \quad L(1/2, \pi \otimes \chi) \sim \sum_{\mathcal{N} \mathfrak{a} \sim \mathcal{N} \mathfrak{q}} \frac{\lambda_{\pi}(\mathfrak{a}) \chi(\mathfrak{a})}{(\mathcal{N} \mathfrak{a})^{1/2}}$$

and similarly for other points on the critical line, where \sim has to be understood in a very broad sense. Note by the way, that the trivial bound would recover the convexity estimate. Now that we have an explicit description of $L(s, \pi \otimes \chi)$ as a finite sum, let us try to exhibit cancellation in such sums. Let us first look at a simple example⁽⁷⁾. Suppose you want to prove that $|\sin x + \cos x| \leq \sqrt{2}$. There are certainly many ways of proving this. Here is one: Square the left hand side and add a “spectrally useful” nonnegative quantity:

$$|\sin x + \cos x|^2 + |\sin x - \cos x|^2 = 2.$$

Now drop the second term, and the proof is complete. In a similar way, it is useful to embed an L -function into a family. First let us cut the sum (6) into h pieces according to the ideal class of the ideal \mathfrak{a} . For simplicity we will only work with the principal class. We consider now the second moment

$$\sum_{\omega \in \Omega} |L(\pi \otimes \omega, 1/2 + it)|^2,$$

where Ω is a family of characters containing χ , for example the family of all characters of $(\mathcal{O}/\mathfrak{q})^{\times}$. These characters are in general not Hecke characters, because they may not be trivial on units, and so strictly speaking the expression $L(\pi \otimes \omega, 1/2 + it)$ does not make sense as a value of an automorphic L -function. It is typical in this context to consider such “fake-moments”; after all, we are free to add to our original quantity $L(\pi \otimes \chi, 1/2 + it)$ whatever we want as long as it is non-negative. Here the expression $L(\pi \otimes \omega, 1/2 + it)$ is just a notation for a suitable Dirichlet polynomial whose coefficients behave roughly like $\lambda_{\pi}(\alpha) \omega(\alpha)$ on principal ideals (α) . This family is of size about $\mathcal{N} \mathfrak{q}$; so even if we assume a sort of Generalized Lindelöf hypothesis in the sense $L(\pi \otimes \omega, 1/2 + it) \ll_{t, \pi} 1$, we can bound the above sum only by $\mathcal{N} \mathfrak{q}$ which after taking the square-root just recovers the convexity bound. The problem here is that our family, although very convenient to work with, is quite

⁽⁷⁾which can be viewed as an instance of arts and science in mathematics: it is art to find the second term, and it is science to prove the trigonometric identity

large. One could try to evaluate a fourth moment instead of a second moment, in which case one could take the fourth root at the end, but our current analytic techniques are not strong enough to estimate a fourth moment appropriately. The idea of Duke-Friedlander-Iwaniec [19] is to weight the sum in our favor so as to highlight the one term we are interested in, but not the other ones that we drop at the end anyway. Hence as a refinement, we consider

$$(7) \quad \sum_{\omega \in \Omega} |A(\omega)|^2 |L(\pi \otimes \omega, 1/2 + it)|^2,$$

where $A(\omega)$ is an ‘‘amplifier’’ that is large for $\omega = \chi$, and rather small otherwise. In practice, $A(\omega)$ will be a short Dirichlet polynomial, e.g.

$$A(\omega) = \sum_{\mathcal{N}(\alpha) \sim L} \chi(\alpha) \bar{\omega}(\alpha),$$

where L is a parameter that we can optimize later. Now we open the square and sum over ω . This shows that we have to bound nontrivially sums roughly of the form

$$(8) \quad \sum_{\substack{n_1, n_2 \in \mathcal{O}_K \cap \mathcal{B} \\ \alpha_1 n_1 \equiv \alpha_2 n_2 \pmod{\mathfrak{q}}} \lambda_\pi(n_1) \bar{\lambda}_\pi(n_2),$$

where α_1, α_2 are of norm about L (they come from the amplifier), and \mathcal{B} is a box in Minkowski space of the form

$$n^{\sigma_1}, \dots, n^{\sigma_d} \asymp (\sphericalangle \mathcal{N} \mathfrak{q})^{1/d}$$

with $\sigma_1, \dots, \sigma_d$ the embeddings of K into \mathbb{R} . We break this sum into pieces according to the value of

$$(9) \quad q := \alpha_1 n_1 - \alpha_2 n_2 \in \mathfrak{q}.$$

The term $q = 0$ is the diagonal term, and pretty straightforward to handle. Let us now assume $q \neq 0$. Expressions of the type (8) with a summation condition of type (9) are usually called shifted convolution sums. Selberg [37] considered in 1965 Dirichlet series (over \mathbb{Q}) of the type

$$D_q(s) := \sum_{n_1 - n_2 = q} \frac{\lambda_\pi(n_1) \bar{\lambda}_\pi(n_2)}{(n_1 + n_2)^s}$$

(which is not an automorphic L -function!) and proved in some cases an analytic continuation to some right half plane $\Re s > 1 - \delta$ with $\delta > 0$, however, without good control of the size in s and q (which is crucial for all known applications). Progress in this respect has been made by Good [22], Jutila [28], Sarnak [35] and Motohashi [33]. Our method, based on the Kirillov model and Sobolev norms gives not only the

analytic continuation with good growth estimates, but also the pleasing structural insight that the series $D_q(s)$ can be decomposed according to the decomposition (5). We present the result in the simplest case [7]. The general case is treated in [5].

Theorem 2. *Let $k > 60$ and $q > 0$ be any integers, and $\lambda(n)$ Hecke eigenvalues of any irreducible cuspidal representation on GL_2 of conductor 1. Then there exist holomorphic functions F_π in the strip $1/2 < \Re s < 3/2$ (depending on k) such that*

$$\sum_{m=n=q} \frac{\lambda(n)\overline{\lambda(m)}(nm)^{(k-1)/2}}{(n+m)^{s+k-1}} = q^{1/2-s} \int \lambda_\pi(q) F_\pi(s) d\pi, \quad \Re s > 1,$$

and

$$\int |F_\pi(s)| d\pi \ll_\varepsilon |s|^{22}, \quad \frac{1}{2} + \varepsilon \leq \Re s \leq \frac{3}{2},$$

where the integral is taken over the union of the discrete spectrum and the continuous spectrum.

Armed with Theorem 2 (or rather a slight generalization thereof) it is relatively straightforward to complete the proof of Theorem 1.

4. Applications

Although at first sight Theorem 1 may seem as some purely analytic trickery, it has, in accordance with the general philosophy of L -functions, interesting arithmetic applications. Perhaps the most appealing application of Theorem 1 is in combination with the formula of Waldspurger [41] and its extensions to number fields. More precisely, let $\tilde{\pi}$ be a cuspidal representation on the double cover $\widetilde{\mathrm{SL}}_2$, generated by a half-integral weight modular form satisfying some technical assumptions, and π the representation on GL_2 given by theta correspondence ("Shimura lift"). Then for squarefree m , Waldspurger's formula relates the square of the m -th Fourier coefficient of $\tilde{\pi}$ to $L(\pi \otimes \chi_m, 1/2)$ where χ_m is the quadratic character corresponding to the extension $K(\sqrt{m})/K$. In this way, Theorem 1 yields the currently best known bounds for Fourier coefficients of half-integral weight Hilbert modular forms.

One particular situation where such bounds are needed, are asymptotic formulae for the number of representations of totally positive integers by ternary quadratic forms, see [3] for an overview of this topic over \mathbb{Q} . Hilbert's eleventh problem asks more generally which integers are (integrally) represented by a given n -ary quadratic form Q over a number field K . If Q is a binary form, it corresponds to some element in the class group of a quadratic extension of K (see [17] for a nice account over \mathbb{Q}).

If Q is indefinite at some archimedean place, Siegel [38] for $n \geq 4$ and Kneser [30] and Hsia [26] for $n = 3$ proved a local-to-global principle, so Siegel's mass formula tells us exactly which integers are represented by Q . If Q is positive definite at every archimedean place and $n \geq 4$, again Siegel's mass formula and simple bounds for Fourier coefficients of Hilbert modular forms give a complete answer (some care has to be taken in the case $n = 4$). The only remaining case of Q positive definite and $n = 3$ was solved by Duke and Schulze-Pillot [20] for $K = \mathbb{Q}$. For arbitrary totally real K , the result was announced in [13] with a sketch of the proof being given in [14] in the class number one case. Combining the argument in [14] with Theorem 1 and Waldspurger's formula e.g. in the version of Baruch-Mao [1] one derives:

Theorem 3 (cf. [14, 13]). *Let K be a totally real number field and let Q be a positive integral ternary quadratic form over K . Then there is an ineffective constant $c > 0$ such that every totally positive squarefree integer $m \in \mathcal{O}_K$ with $\mathcal{N}m \geq c$ is represented integrally by Q if and only if it is integrally represented over every completion of K .*

The representation of non-squarefree integers is quite subtle, but in principle can again be characterized by more involved local considerations, cf. e.g. [36].

Theorem 3 can be refined in various ways and also made quantitative, which yields for example applications of the following type: Gauß proved in his *Disquisitiones* that a rational integer n can be written as a sum of three squares if and only if it is not of the form $4^k(8m+7)$, and if it is in addition not divisible by a very high power of 2, the number of such representations is about $L(1, \chi_n)\sqrt{n}$ which by Siegel's theorem is $n^{1/2+o(1)}$. Hence one may ask if all integers satisfying some natural congruence conditions can still be written as a sum of three squares of numbers with certain restrictions, e.g. sums of three squares of primes, or sums of three squares of squarefree numbers or sums of three squares of smooth numbers etc. Combining the previous results with a carefully designed sieve (the vector sieve as developed by Brüdern and Fouvry [10]) one can for example prove [4, 3]:

Theorem 4. *Let⁽⁸⁾ $n \equiv 3 \pmod{24}$, $5 \nmid n$, be sufficiently large, and let $\gamma = 1/567$. Then n is the sum of three squares of integers with all their prime factors greater than n^γ . The number of such representations exceeds $\gg n^{1/2-\varepsilon}$. In particular, every such n is the sum of three squares having at most 284 prime factors.*

⁽⁸⁾In [3, Proposition 3.1] the condition $n \equiv 3 \pmod{8}$ has to be replaced by $n \equiv 3 \pmod{24}$.

For similar results of this flavor see [3].

A different type of application of Theorem 1 can be found in [15] (cf. also [40, 44]) that generalizes work of Duke [18]: Under the assumption of a subconvex bound as above it is proved that a certain family of Heegner points and certain d -dimensional subvarieties are equidistributed on the Hilbert modular variety $\mathrm{PSL}_2(\mathcal{O}_K)\backslash\mathcal{H}^d$. For example, if $K = \mathbb{Q}$, and $-D$ is the discriminant of an imaginary quadratic field, then each ideal class $\mathfrak{a} = (a, \frac{1}{2}(b - \sqrt{-D}))$, say, in the class group of $\mathbb{Q}(\sqrt{-D})$ gives a Heegner point $z = (b - \sqrt{-D})/(2a)$ in $X := \mathrm{PSL}_2(\mathbb{Z})\backslash\mathcal{H}$. If $D \rightarrow \infty$, these points become denser and denser in X , and the above statement says that they become actually equidistributed (with an explicit rate of decay) with respect to the standard measure $y^{-2} dx dy$ on X . In order to prove this, one has to sum the values of a test function at these Heegner points, and by a spectral decomposition one can assume that the test function is an eigenform of the Laplacian. This leads to certain Weyl sums, which can be expressed as central values of twisted L -functions. Using bounds for the L -values as in Theorem 1, one derives an equidistribution statement.

Finally we note that the subconvex bound in Theorem 1 (in particular for $K = \mathbb{Q}$) is a crucial input for certain subconvex bounds of higher degree L -functions, which in turn have other arithmetic applications. We refer the reader to [24, 8] for more details.

References

- [1] E. M. BARUCH & Z. MAO – “Central value of automorphic L -functions”, *Geom. Funct. Anal.* **17** (2007), no. 2, p. 333–384.
- [2] J. BERNSTEIN & A. REZNIKOV – “Sobolev norms of automorphic functionals”, *Int. Math. Res. Not.* (2002), no. 40, p. 2155–2174.
- [3] V. BLOMER – “Ternary, quadratic forms, and sums of three squares with restricted variables”, *Anatomy of integers*, CRM Proc. Lecture Notes, vol. 46, Amer. Math. Soc., Providence, RI, 2008, p. 1–17.
- [4] V. BLOMER & J. BRÜDERN – “A three squares theorem with almost primes”, *Bull. London Math. Soc.* **37** (2005), no. 4, p. 507–513.
- [5] V. BLOMER & G. HARCOS – “Twisted L -functions over number fields, and Hilbert’s eleventh problem”, preprint.
- [6] ———, “Hybrid bounds for twisted L -functions”, *J. Reine Angew. Math.* **621** (2008), p. 53–79.
- [7] ———, “The spectral decomposition of shifted convolution sums”, *Duke Math. J.* **144** (2008), no. 2, p. 321–339.

- [8] V. BLOMER, G. HARCOS & P. MICHEL – “Bounds for modular L -functions in the level aspect”, *Ann. Sci. École Norm. Sup. (4)* **40** (2007), no. 5, p. 697–740.
- [9] C. BREUIL, B. CONRAD, F. DIAMOND & R. TAYLOR – “On the modularity of elliptic curves over \mathbf{Q} : wild 3-adic exercises”, *J. Amer. Math. Soc.* **14** (2001), no. 4, p. 843–939 (electronic).
- [10] J. BRÜDERN & É. FOUVRY – “Lagrange’s four squares theorem with almost prime variables”, *J. Reine Angew. Math.* **454** (1994), p. 59–96.
- [11] D. A. BURGESS – “On character sums and L -series”, *Proc. London Math. Soc. (3)* **12** (1962), p. 193–206.
- [12] V. A. BYKOVSKIĪ – “A trace formula for the scalar product of Hecke series and its applications”, *translated in J. Math. Sci. (New York)* **89** (1998), p. 915–932.
- [13] J. COGDPELL, I. PIATETSKII-SHAPIRO & P. SARNAK – “Estimates on the critical line for Hilbert modular L -functions and applications”, unpublished.
- [14] J. W. COGDPELL – “On sums of three squares”, *J. Théor. Nombres Bordeaux* **15** (2003), no. 1, p. 33–44, Les XXIIèmes Journées Arithmétiques (Lille, 2001).
- [15] P. B. COHEN – “Hyperbolic equidistribution problems on Siegel 3-folds and Hilbert modular varieties”, *Duke Math. J.* **129** (2005), no. 1, p. 87–127.
- [16] J. B. CONREY & H. IWANIEC – “The cubic moment of central values of automorphic L -functions”, *Ann. of Math. (2)* **151** (2000), no. 3, p. 1175–1216.
- [17] D. A. COX – *Primes of the form $x^2 + ny^2$* , A Wiley-Interscience Publication, John Wiley & Sons Inc., New York, 1989, Fermat, class field theory and complex multiplication.
- [18] W. DUKE – “Hyperbolic distribution problems and half-integral weight Maass forms”, *Invent. Math.* **92** (1988), no. 1, p. 73–90.
- [19] W. DUKE, J. FRIEDLANDER & H. IWANIEC – “Bounds for automorphic L -functions”, *Invent. Math.* **112** (1993), no. 1, p. 1–8.
- [20] W. DUKE & R. SCHULZE-PILLOT – “Representation of integers by positive ternary quadratic forms and equidistribution of lattice points on ellipsoids”, *Invent. Math.* **99** (1990), no. 1, p. 49–57.
- [21] J. B. FRIEDLANDER – “Bounds for L -functions”, *Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994)* (Basel), Birkhäuser, 1995, p. 363–373.
- [22] A. GOOD – “Beiträge zur Theorie der Dirichletreihen, die Spitzenformen zugeordnet sind”, *J. Number Theory* **13** (1981), no. 1, p. 18–65.
- [23] G. HARCOS – “Uniform approximate functional equation for principal L -functions”, *Int. Math. Res. Not.* (2002), no. 18, p. 923–932.
- [24] G. HARCOS & P. MICHEL – “The subconvexity problem for Rankin-Selberg L -functions and equidistribution of Heegner points. II”, *Invent. Math.* **163** (2006), no. 3, p. 581–655.
- [25] E. HECKE – “Zur Theorie der elliptischen Modulfunktionen”, *Math. Ann.* **97** (1927), no. 1, p. 210–242.
- [26] J. S. HSIA – “Representations by spinor genera”, *Pacific J. Math.* **63** (1976), no. 1, p. 147–152.
- [27] H. IWANIEC & P. SARNAK – “Perspectives on the analytic theory of L -functions”, *Geom. Funct. Anal.* (2000), no. Special Volume, Part II, p. 705–741, GAFA 2000 (Tel Aviv, 1999).

- [28] M. JUTILA – “The additive divisor problem and its analogs for Fourier coefficients of cusp forms. I”, *Math. Z.* **223** (1996), no. 3, p. 435–461.
- [29] H. H. KIM & F. SHAHIDI – “Cuspidality of symmetric powers with applications”, *Duke Math. J.* **112** (2002), no. 1, p. 177–197.
- [30] M. KNESER – “Darstellungsmasse indefiniter quadratischer Formen”, *Math. Z.* **77** (1961), p. 188–194.
- [31] H. MAASS – “Über eine neue Art von nichtanalytischen automorphen Funktionen und die Bestimmung Dirichletscher Reihen durch Funktionalgleichungen”, *Math. Ann.* **121** (1949), p. 141–183.
- [32] P. MICHEL – “Analytic number theory and families of automorphic L -functions”, Automorphic forms and applications, IAS/Park City Math. Ser., vol. 12, Amer. Math. Soc., Providence, RI, 2007, p. 181–295.
- [33] Y. MOTOHASHI – “A note on the mean value of the zeta and L -functions. XIV”, *Proc. Japan Acad. Ser. A Math. Sci.* **80** (2004), no. 4, p. 28–33.
- [34] A. V. P. MICHEL – “The subconvexity problem for GL_2 ”, preprint.
- [35] P. SARNAK – “Integrals of products of eigenfunctions”, *Internat. Math. Res. Notices* (1994), no. 6, p. 251 ff., approx. 10 pp. (electronic).
- [36] R. SCHULZE-PILLOT – “Darstellungsmaße von Spinorgeschlechtern ternärer quadratischer Formen”, *J. Reine Angew. Math.* **352** (1984), p. 114–132.
- [37] A. SELBERG – “On the estimation of Fourier coefficients of modular forms”, Proc. Sympos. Pure Math., Vol. VIII, Amer. Math. Soc., Providence, R.I., 1965, p. 1–15.
- [38] C. L. SIEGEL – “Indefinite quadratische Formen und Funktionentheorie. II”, *Math. Ann.* **124** (1952), p. 364–387.
- [39] R. TAYLOR & A. WILES – “Ring-theoretic properties of certain Hecke algebras”, *Ann. of Math. (2)* **141** (1995), no. 3, p. 553–572.
- [40] A. VENKATESH – “Sparse equidistribution problems, period bounds, and subconvexity”, to appear.
- [41] J.-L. WALDSPURGER – “Sur les coefficients de Fourier des formes modulaires de poids demi-entier”, *J. Math. Pures Appl. (9)* **60** (1981), no. 4, p. 375–484.
- [42] E. T. WHITTAKER & G. N. WATSON – *A course of modern analysis*, Cambridge University Press, Cambridge, 4th edition.
- [43] A. WILES – “Modular elliptic curves and Fermat’s last theorem”, *Ann. of Math. (2)* **141** (1995), no. 3, p. 443–551.
- [44] S.-W. ZHANG – “Equidistribution of CM-points on quaternion Shimura varieties”, *Int. Math. Res. Not.* (2005), no. 59, p. 3657–3689.

BOUNDED COHOMOLOGY OF THE p -ADIC UPPER HALF PLANE

Ehud de Shalit

Einstein Institute of Mathematics, Hebrew University, Jerusalem
E-mail: deshalit@math.huji.ac.il

Abstract. We consider the bounded cohomology of Drinfel'd's p -adic upper half plane, with values in a local system M . We show that it has a canonical splitting into the space of bounded one-forms with values in M . The paper is largely expository, and self-contained.

1. Introduction

Let K be a p -adic field, and \mathbb{C}_p the completion of its algebraic closure. Let \mathfrak{X} be the p -adic “upper half plane” over K introduced by Drinfel'd in [6]. The rigid analytic de-Rham cohomology of \mathfrak{X} is the space of rigid analytic \mathbb{C}_p -valued 1-forms on \mathfrak{X} , modulo exact ones,

$$(1.1) \quad H_{dR}^1(\mathfrak{X}) = \Omega / d\mathcal{O}.$$

In contrast to the complex upper half plane, this cohomology does not vanish. In fact, it is a classical result that it is isomorphic to the space of *harmonic* (\mathbb{C}_p -valued) 1-cochains on the Bruhat-Tits tree \mathcal{T} of the group $G = SL_2(K)$. (See [3] for background.) Recall that \mathcal{T} is a $q + 1$ regular tree (where q is the cardinality of the residue field of K), and a 1-cochain c on \mathcal{T} is harmonic if (i) $c(\bar{e}) = -c(e)$ if \bar{e} is the edge e with reversed orientation, and (ii) the sum of $c(e)$ for all the edges flowing into any given vertex, is zero. We denote the space of harmonic 1-cochains by C_{har}^1 . The isomorphism

$$(1.2) \quad H_{dR}^1(\mathfrak{X}) \simeq C_{har}^1$$

is given by the residue homomorphism. If $\omega \in \Omega$, we let

$$(1.3) \quad c_\omega(e) = \text{res}_e(\omega)$$

be the residue of ω along the oriented edge e (see [7] p.94). Then c_ω is a harmonic cochain, it vanishes if and only if ω is exact, and every harmonic cochain is of this type. Both Ω and C_{har}^1 are inverse limits of Banach spaces (in the case of C_{har}^1 these are finite dimensional Banach spaces), and carry a natural p -adic Fréchet topology in which they are reflexive (see [13]). The group G acts on the two spaces. The residue homomorphism respects both the topology and the G -action.

If M is a finite dimensional representation of G over K , we may tensor the above spaces with M to get the notions of M -valued rigid analytic functions or forms, cohomology with coefficients in M , and M -valued harmonic cochains. The isomorphism

$$(1.4) \quad H_{dR}^1(\mathfrak{X}; M) \simeq C_{har}^1(M)$$

(where the notation is self-explanatory) is obtained from the case of trivial coefficients by tensoring with M . One should only be aware that the G -action now involves both the geometric action on \mathfrak{X} , and the action on the coefficients.

An *integral structure* \mathbf{M} in M is the assignment, for every simplex σ of \mathcal{T} , of an \mathcal{O}_K -lattice $\mathbf{M}(\sigma)$ in M , such that $\mathbf{M}(\gamma\sigma) = \gamma\mathbf{M}(\sigma)$ for every $\gamma \in G$, and $\mathbf{M}(v) \subset \mathbf{M}(\varepsilon)$ if the vertex v belongs to the edge ε . Integral structures exist, and any two are commensurable. An M -valued harmonic cochain c is *bounded* if there exists an integral structure \mathbf{M} such that $c(e) \in \mathbf{M}(e)$ for every e . The space of bounded M -valued harmonic cochains is preserved by the action of G . We denote it by $C_{har}^1(M)^{bnd}$. It is not a-priori clear that bounded harmonic cochains exist, namely that the two conditions of harmonicity and boundedness can coexist. The simplest way to produce plenty of bounded harmonic cochains is to consider a discrete cocompact subgroup Γ in G , and note that $H^0(\Gamma, C_{har}^1(M)) \subset C_{har}^1(M)^{bnd}$. The spaces $H^0(\Gamma, C_{har}^1(M))$ have been studied in [4], among other places.

The cotangent space to \mathfrak{X} at z_0 has a canonical integral structure $\Omega|_{z_0}$ which is the \mathcal{O}_{C_p} -span of $dz/(z_0 - \zeta)$ for $\zeta \in K$ (this integral structure does not depend on the choice of the coordinate z and is G -equivariant). Let $r : \mathfrak{X} \rightarrow |\mathcal{T}|$ be the reduction map from \mathfrak{X} to the real realization of the tree (see [7] or [4]). An M -valued 1-form $\omega \in \Omega(M)$ is called *bounded* if there exists an integral structure \mathbf{M} such that

$$(1.5) \quad \omega|_z \in \Omega|_z \otimes \mathbf{M}(\sigma)$$

whenever $r(z) \in |\sigma|$. The space of bounded M -valued forms is denoted by $\Omega(M)^{bnd}$. It is G -stable. Once again, the simplest way to produce such forms is to produce Γ -invariant ones for some Γ .

We shall describe below a filtration, due to Schneider and Stuhler, of the module $\Omega(M)$, by certain coherent submodules $F^i\Omega(M)$ of decreasing ranks, which are direct summands (i.e. give rise to subvectorbundles), and which are G -stable. The last step in the filtration, $F^n\Omega(M)$, still maps surjectively onto $H_{dR}^1(\mathfrak{X}; M)$, so every cohomology class is represented by an M -valued form from that last piece.

We shall check that bounded M -valued forms get mapped, by the residue homomorphism, to bounded M -valued harmonic cochains.

Theorem 1.1. *The residue homomorphism induces an isomorphism*

$$(1.6) \quad F^n\Omega(M) \cap \Omega(M)^{bnd} \simeq C_{har}^1(M)^{bnd}.$$

In other words, every bounded M -valued harmonic cochain is of the form c_ω for an $\omega \in F^n\Omega(M) \cap \Omega(M)^{bnd}$, and if a bounded ω which lies in the last step of the filtration is exact, then it is zero.

The theorem is not new. For a survey see [3], Theorem 2.3.2, Corollary 2.3.4 and the references therein. However, our approach to Morita duality, to the notion of bounded differential forms, and to the injectivity statement in the theorem, is different than what may be found in the literature. Except for a certain technical verification of convergence at the very end, our exposition is self contained.

The surjectivity statement in the theorem asserts that the map from forms to cohomology splits G -equivariantly over the bounded cohomology. The existence of such a splitting has been known for some time and follows from the theorem of Amice-Velu-Vishik. Roughly speaking (we shall make everything precise below), to a bounded M -valued harmonic cochain c one attaches a (scalar valued) distribution λ_c on \mathbb{P}_K^1 with “growth conditions” of order n . This distribution extends in a natural way to a tempered distribution $\tilde{\lambda}_c$, i.e. to a linear functional on the space of the locally analytic functions on \mathbb{P}_K^1 (this is where one needs the estimates of Amice-Velu-Vishik). Teitelbaum’s Poisson integral

$$(1.7) \quad f(z) = \int \frac{d\tilde{\lambda}_c(\zeta)}{z-\zeta}$$

makes sense since the integrand, $1/(z-\zeta)$, is locally analytic in $\zeta \in \mathbb{P}_K^1$. As a function of $z \in \mathfrak{X}$, $f(z)$ is rigid analytic. The M -valued form

$$(1.8) \quad \omega = f(z)(u-zv)^n dz$$

(where we have realized M as the space of homogeneous polynomials of degree n in the two variables u and v) lies in the last step of the filtration, is bounded, and c_ω is our original c .

It is tempting to think about generalizations of the questions discussed in this paper to higher dimensions. Let us only mention that the notions of boundedness, either for differential forms, or for cohomology with coefficients in a rational representation M , can be defined in precisely the same way, and the higher dimensional residue homomorphism from k -forms to harmonic k -cochains (see [5]) carries bounded closed M -valued k -forms to bounded M -valued harmonic k -cochains. The paper [12] discusses an analogue of the filtration F studied here, but an analogue of the theorem mentioned above is not known to hold. Such an analogue would probably be non-void for forms of top degree only. For lower degree forms, and non-trivial coefficients, the notion of boundedness has to be modified.

For trivial coefficients, we expect that the bounded k -forms are nothing else but the Iovita-Spiess forms (see [9] and [1]). This will be the case if we can show that a bounded exact form must vanish, and it would be enough to check this last assertion for forms of top degree. For that see [8].

1.1. Notation. Let K be a p -adic field (a finite extension of \mathbb{Q}_p), V a two dimensional vector space over K , and $G = SL(V)$. Fix a basis, and identify V with column vectors, and G with $SL_2(K)$ acting from the left. Let u, v be the dual basis, so that $V^\vee = Ku + Kv$. For $n \geq 0$, let $M = M_n = \text{Symm}^n(V^\vee)$. Then M is the unique irreducible rational representation of G (viewed as an algebraic group over K) of dimension $n + 1$. In coordinates,

$$(1.9) \quad M = K[u, v]_{\text{hom. deg.} = n}$$

and the action of $\gamma \in G$ on $P \in M$ is

$$(1.10) \quad \gamma P = P \circ \gamma^{-1} = P(au + bv, cu + dv)$$

where

$$(1.11) \quad \gamma^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \gamma = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Let \mathbb{C}_p be the completion of a fixed algebraic closure of K . The p -adic upper half plane is the rigid analytic space $\mathfrak{X} \subset \mathbb{P}(V^\vee)$ which is the complement of the K -rational points:

$$(1.12) \quad \mathfrak{X} = \mathbb{P}(V^\vee) \setminus \mathbb{P}(V_K^\vee).$$

We do not distinguish between \mathfrak{X} (as a ringed space in a certain category) and the set of its \mathbb{C}_p -points. The latter is identified with the complement of K in \mathbb{C}_p by sending the line through $tu + sv$ to its *coordinate* $z = -st^{-1}$. With this convention the action of γ on z is the usual action

$$(1.13) \quad \gamma(z) = \frac{dz - b}{-cz + a}.$$

We denote by \mathcal{A} the set of K -rational lines in V with its p -adic topology. Thus $\mathcal{A} = \mathbb{P}(V_K)$. If $a \in \mathcal{A}$ the “hyperplane” H_a is the point $b \in \mathbb{P}(V_K^\vee)$ given by $\langle a, b \rangle = 0$, and the complement of \mathfrak{X} is the union of the H_a for $a \in \mathcal{A}$. In coordinates, if $a = (x : y)^t$ (column vector up to homothety) then H_a is the line through $-yu + xv$, so its coordinate is

$$(1.14) \quad z(H_a) = xy^{-1}$$

and we see that the standard (left) G action on a corresponds to the action on z , namely $z(H_{\gamma a}) = \gamma(z(H_a))$. This allows us to identify \mathcal{A} and $\mathbb{P}(V^\vee) \setminus \mathfrak{X}$ as G -sets. Of course, we could do it without introducing coordinates by observing that $\gamma(H_a) = H_{\gamma a}$. This discussion may seem superfluous, but we shall soon need to extend locally analytic functions on \mathcal{A} to a rigid analytic neighborhood of \mathcal{A} in \mathfrak{X} , so it is necessary to view \mathcal{A} and \mathfrak{X} in the same ambient space, despite the fact that a-priori they live in dual projective spaces. The correct generalization of this in higher dimensions requires some care.

We let \mathcal{T} denote the Bruhat-Tits tree of G . Then \mathcal{A} is identified canonically with the set of the *ends* of \mathcal{T} , and for an oriented edge e we let \mathcal{A}_e denote the subset of ends in the direction of e . Each \mathcal{A}_e is a disk in $\mathbb{P}(V_K)$ and \mathcal{A} is the disjoint union of \mathcal{A}_e and $\mathcal{A}_{\bar{e}}$. Let

$$(1.15) \quad \mathcal{O} = \underline{\mathcal{O}}(\mathfrak{X}), \quad \Omega = \underline{\Omega}(\mathfrak{X})$$

denote the \mathbb{C}_p -algebra of rigid analytic functions on \mathfrak{X} , and the \mathcal{O} -module of rigid analytic differential forms. The underlined symbols represent the corresponding sheaves in the rigid analytic topology. We let $\mathcal{O}(M)$ and $\Omega(M)$ stand for the same spaces tensored (over K) with M . These are the M -valued functions or forms. An element $\gamma \in G$ acts on $f \in \mathcal{O}(M)$ via the rule

$$(1.16) \quad (\gamma f)(z) = (1 \otimes \gamma)(f(\gamma^{-1}z)).$$

It acts on $\omega \in \Omega(M)$ via

$$(1.17) \quad (\gamma \omega)|_z = (1 \otimes \gamma)(\gamma^{-1})^*(\omega|_{\gamma^{-1}z}).$$

2. Differential forms and cohomology with coefficients in M

2.1. Filtration on $\mathcal{O}(M)$. We follow [11], pp.95-97. Let

$$(2.1) \quad \mathcal{O}(M) = F^0 \supset F^1 \supset \dots \supset F^n \supset \{0\} = F^{n+1}$$

where

$$(2.2) \quad F^k = \text{Span}_{\mathcal{O}} \left\{ (u-zv)^k u^{n-k-l} v^l; 0 \leq l \leq n-k \right\}.$$

Then:

(1) The filtration is G -stable. In fact, if γ^{-1} is as above, $f = (u-zv)^k u^{n-k-l} v^l$

$$(2.3) \quad (\gamma f)(z) = \gamma(f(\gamma^{-1}z)) = (cz+d)^{-k} (u-zv)^k (au+bv)^{n-k-l} (cu+dv)^l \in F^k.$$

(2) $F^k = \mathcal{O} \cdot (u-zv)^k v^{n-k} \oplus F^{k+1}$ (as \mathcal{O} -modules, but not as G -modules).

(3) $\mathcal{O}(M)$ is free over \mathcal{O} with basis $(u-zv)^k v^{n-k}$ ($0 \leq k \leq n$).

(4) For any $m \in \mathbb{Z}$ let $\mathcal{O}(m)$ be the ring \mathcal{O} with the *twisted* G -action

$$(2.4) \quad \gamma \cdot f = (cz+d)^m (f \circ \gamma^{-1}),$$

where (c, d) is the bottom row of γ^{-1} . Then $f \mapsto f(z)(u-zv)^k v^{n-k} \bmod F^{k+1}$ is an isomorphism

$$(2.5) \quad \Theta_k : \mathcal{O}(n-2k) \simeq F^k / F^{k+1}.$$

(5) Similarly $F \cdot \Omega = \Omega \otimes_{\mathcal{O}} F$ is a G -stable filtration on $\Omega(M)$ and

$$(2.6) \quad f \mapsto f(z)(u-zv)^k v^{n-k} dz \bmod \Omega \otimes F^{k+1}$$

is an isomorphism

$$(2.7) \quad \Theta'_k : \mathcal{O}(n-2k-2) \simeq \Omega \otimes F^k / \Omega \otimes F^{k+1}.$$

(6) For $1 \leq k \leq n$, the G -homomorphism $d : \mathcal{O}(M) \rightarrow \Omega(M)$ maps F^k to $\Omega \otimes F^{k-1}$ (*Griffiths transversality*) and induces a commutative diagram

$$(2.8) \quad \begin{array}{ccc} \mathcal{O}(n-2k) & \xrightarrow{(-k) \cdot} & \mathcal{O}(n-2k) \\ \downarrow \Theta_k & & \downarrow \Theta'_{k-1} \\ F^k / F^{k+1} & \xrightarrow{d} & \Omega \otimes F^{k-1} / \Omega \otimes F^k \end{array} .$$

(7) There is a decomposition as a direct sum of abelian groups with G -action

$$(2.9) \quad \Omega(M) = d(F^1) \oplus (\Omega \otimes F^n).$$

Indeed, (6) shows that

$$\Omega(M) = d(F^1) + (\Omega \otimes F^1) = d(F^1) + (\Omega \otimes F^2) = \dots = d(F^1) + (\Omega \otimes F^n).$$

For every $1 \leq k \leq n$, if $f \in F^k - F^{k+1}$, then $df \in \Omega \otimes F^{k-1} - \Omega \otimes F^k$. Since any non-zero $f \in F^1$ falls in some $F^k - F^{k+1}$, $df \notin \Omega \otimes F^k$, hence clearly $df \notin \Omega \otimes F^n$ and the sum is direct.

Note that $d(F^1)$ is *not* an \mathcal{O} -submodule of $\Omega(M)$, so this decomposition is not a splitting of the last step in the filtration of $\Omega(M)$. For example, if $n = 1$ the decomposition is

$$(2.10) \quad \{\alpha u dz\} \oplus \{\beta v dz\} = \{-f v dz + f'(u - zv) dz\} \oplus \{g(u - zv) dz\}$$

where $(f, g) = (-\beta - z\alpha, 2\alpha + \beta' + z\alpha')$ and $(\alpha, \beta) = (g + f', -f - zf' - zg)$.

(8) Let $n \geq 0$. There is a commutative diagram of additive (but not \mathcal{O} -linear) G -homomorphisms

$$(2.11) \quad \begin{array}{ccc} \mathcal{O}(n) & \xrightarrow{\frac{1}{n!} \left(\frac{d}{dz} \right)^{n+1}} & \mathcal{O}(-n-2) \\ \downarrow \Theta_0 & & \downarrow \Theta'_n \\ \mathcal{O}(M)/F^1 & \xrightarrow{pr_2 \circ d} & \Omega \otimes F^n \end{array},$$

where pr_2 is the projection on the second factor in the decomposition (7).

Indeed, for $1 \leq k \leq n$

$$(2.12) \quad d \circ \Theta_k(f) = \left\{ f'(z)(u - zv)^k v^{n-k} - kf(z)(u - zv)^{k-1} v^{n-k+1} \right\} dz \bmod dF^{k+1}$$

implies

$$(2.13) \quad f'(z)(u - zv)^k v^{n-k} dz \equiv kf(z)(u - zv)^{k-1} v^{n-k+1} dz \bmod dF^1$$

hence, iterating,

$$(2.14) \quad f(z)v^n dz \equiv \frac{1}{n!} f^{(n)}(z)(u - zv)^n dz \bmod dF^1.$$

From this we get

$$(2.15) \quad d(g(z)v^n) \equiv \frac{1}{n!} g^{(n+1)}(z)(u - zv)^n dz \bmod dF^1,$$

or

$$(2.16) \quad pr_2 \circ d \circ \Theta_0(g) = \Theta'_n \left(\frac{1}{n!} g^{(n+1)} \right).$$

Note that the fact that the top arrow is a G -homomorphism is not easy to establish via a direct computation, since differentiation does not commute with the action of G . It is rather a consequence of the fact that the other three arrows commute with G . The surjectivity of the horizontal arrows has the following consequence.

Corollary 2.1. For $n \geq 0$, let $P_n(n)$ be the space of polynomials of degree at most n , with G -action induced from (4). Then there is an exact sequence of G -representations

$$(2.17) \quad 0 \rightarrow \mathcal{O}(n)/P_n(n) \xrightarrow{\frac{1}{n!} \left(\frac{d}{dz}\right)^{n+1}} \mathcal{O}(-n-2) \rightarrow H_{dR}^1(\mathfrak{X}; M_n) \rightarrow 0.$$

Proof. De-Rham cohomology with values in M is given by

$$(2.18) \quad \begin{aligned} \Omega(M)/d\mathcal{O}(M) &= \Omega \otimes F^n / (d\mathcal{O}(M) \cap \Omega \otimes F^n) \\ &= \Omega \otimes F^n / pr_2 \circ d(\mathcal{O}(M)/F^1) \end{aligned}$$

so we conclude by the commutative diagram of step (8), and by the fact that the kernel of $n + 1$ -fold differentiation is P_n . □

2.2. Morita duality (trivial coefficients). Locally analytic and meromorphic functions. Let \mathcal{C}^{an} be the space of locally-analytic, \mathbb{C}_p -valued functions on \mathcal{A} . Since \mathcal{A} is compact, each $\varphi \in \mathcal{C}^{an}$ admits a decomposition of \mathcal{A} into a finite union of closed disks, such that φ is rigid analytic on each of them. Let $\mathcal{C} = \mathcal{C}^{mer}$ be the space of locally meromorphic \mathbb{C}_p -valued functions on \mathcal{A} , and \mathcal{R} the subspace of \mathbb{C}_p -valued rational functions with poles in \mathcal{A} . Note that \mathcal{R} is a subring, but not a field. Of course, $\mathcal{R} \cap \mathcal{C}^{an} = \mathbb{C}_p$, the constants, and $\mathcal{R} + \mathcal{C}^{an} = \mathcal{C}$ (the theorem on principal parts).

Topologies. The space \mathcal{O} (hence $\Omega = \mathcal{O}dz$) is a locally convex topological vector space (for *nonarchimedean functional analysis* consult [13]). The topology is given by the family of norms $|\cdot|_{\mathfrak{X}_n}$ where \mathfrak{X}_n is an increasing sequence of affinoids exhausting \mathfrak{X} . In other words,

$$(2.19) \quad \Omega = \varprojlim \Omega(\mathfrak{X}_n).$$

The space \mathcal{C}^{an} on the other hand is topologized as an inductive limit (union)

$$(2.20) \quad \mathcal{C}^{an} = \varinjlim \prod_{U \in \mathcal{U}} \mathcal{O}(U),$$

where the limit is over all the finite coverings \mathcal{U} of \mathcal{A} by disjoint unions of closed disks, and for a closed disk U , $\mathcal{O}(U)$ is given the usual Banach (sup norm) topology.

Pairing. We give \mathcal{C}/\mathcal{R} the natural topology arising from the identification

$$(2.21) \quad \mathcal{C}/\mathcal{R} = \mathcal{C}^{an}/\mathbb{C}_p.$$

If \mathcal{T}' is a finite subtree of \mathcal{T} we denote by $\text{Ends}(\mathcal{T}')$ the collection of oriented edges (u, v) such that $u \in \mathcal{T}'$ but $v \notin \mathcal{T}'$. If $\varphi \in \mathcal{C}$ and $\omega \in \Omega$ we define a pairing

$$(2.22) \quad \langle \varphi, \omega \rangle = \lim_{\mathcal{T}'} \sum_{e \in \text{Ends}(\mathcal{T}')} \text{res}_e(\varphi \omega).$$

The limit means that we take \mathcal{T}' large enough. The sum then makes sense and is independent of \mathcal{T}' .

Theorem 2.2. (Morita duality) *The above pairing induces a perfect pairing of topological vector spaces*

$$(2.23) \quad \mathcal{C}/\mathcal{R} \times \Omega \rightarrow \mathbb{C}_p.$$

In other words, this pairing identifies each of the two spaces (algebraically) as the space of all continuous linear functionals on the other.

In fact, more is true. Since \mathcal{A} is a compact set, \mathcal{C}^{an} , hence \mathcal{C}/\mathcal{R} , is a topological vector space of compact type (an inductive limit of Banach spaces under inclusion maps which are compact). It is complete and reflexive, and its strong dual

$$(2.24) \quad (\mathcal{C}^{an})'_b = \lim_{\leftarrow} \left(\prod_{U \in \mathcal{U}} \mathcal{O}(U) \right)'$$

is a Fréchet space. The above pairing identifies Ω with $(\mathcal{C}/\mathcal{R})'_b$ and \mathcal{C}/\mathcal{R} with Ω'_b .

Theorem 2.3. *The annihilator of the exact forms $d\mathcal{O}$ under the above pairing is the closed space $\mathcal{C}^\infty/\mathbb{C}_p$ of locally constant functions on \mathcal{A} modulo constants, so one obtains a duality*

$$(2.25) \quad \mathcal{C}^\infty/\mathbb{C}_p \times H_{dR}^1(\mathfrak{X}) \rightarrow \mathbb{C}_p.$$

In fact, if φ is locally constant we have

$$(2.26) \quad \begin{aligned} \langle \varphi, \omega \rangle &= \lim_{\mathcal{T}'} \sum_{e \in \text{Ends}(\mathcal{T}')} \varphi|_{\mathcal{A}_e} \text{res}_e(\omega) \\ &= \int_{\mathcal{A}} \varphi d\mu_\omega. \end{aligned}$$

Here μ_ω is the distribution defined by

$$(2.27) \quad \mu_\omega(\mathcal{A}_e) = c_\omega(e) = \text{res}_e(\omega).$$

2.3. Morita duality with values in M , and the filtrations. Now let $M = M_n$ as before. We introduce the filtration $\Phi^k = \Phi^{k\mathcal{C}}$ on $\mathcal{C}(M) = \mathcal{C} \otimes_K M$, as follows:

$$(2.28) \quad \mathcal{C}(M) = \Phi^0 \supset \Phi^1 \supset \dots \supset \Phi^n \supset \{0\} = \Phi^{n+1}$$

where

$$(2.29) \quad \Phi^k = \text{Span}_{\mathcal{C}} \left\{ (u - \zeta v)^k u^{n-k-l} v^l; 0 \leq l \leq n - k \right\}.$$

Then:

(1) The filtration is G -stable. In fact, if

$$(2.30) \quad \gamma^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

and $\varphi = (u - \zeta v)^k u^{n-k-l} v^l$ then

$$(2.31) \quad (\gamma\varphi)(\zeta) = \gamma(\varphi(\gamma^{-1}\zeta)) = (c\zeta + d)^{-k} (u - \zeta v)^k (au + bv)^{n-k-l} (cu + dv)^l \in \Phi^k.$$

Notice that we could not make this definition if we let \mathcal{C} stand for \mathcal{C}^{an} because, contrary to the situation with $\mathcal{O}(\mathcal{X})$, $c\zeta + d$ may vanish on \mathcal{A} .

(2) $\Phi^k = \mathcal{C} \cdot (u - \zeta v)^k v^{n-k} \oplus \Phi^{k+1}$ (as \mathcal{C} -modules, but not as G -modules).

(3) $\mathcal{C}(M)$ is free over \mathcal{C} with basis $(u - \zeta v)^k v^{n-k}$ ($0 \leq k \leq n$).

(4) For any $m \in \mathbb{Z}$ let $\mathcal{C}(m)$ be the ring \mathcal{C} with the twisted G -action

$$(2.32) \quad \gamma \cdot \varphi = (c\zeta + d)^m (\varphi \circ \gamma^{-1}),$$

where (c, d) is the bottom row of γ^{-1} . Then $\varphi \mapsto \varphi(\zeta)(u - \zeta v)^k v^{n-k} \pmod{\Phi^{k+1}}$ is an isomorphism

$$(2.33) \quad \Psi_k : \mathcal{C}(n - 2k) \simeq \Phi^k / \Phi^{k+1}.$$

In a completely analogous way we introduce the filtration $\Phi^k \mathcal{R}$

$$(2.34) \quad \Phi^k \mathcal{R} = \mathcal{R} \cap \Phi^k \mathcal{C}$$

and deduce that

$$(2.35) \quad \Phi^k(\mathcal{C}/\mathcal{R}) = \Phi^k \mathcal{C} / \Phi^k \mathcal{R}$$

is a filtration on \mathcal{C}/\mathcal{R} .

Morita duality with values in M .

Consider the perfect pairing

$$(2.36) \quad (\cdot, \cdot) : M_n \times M_n \rightarrow K$$

given by $\text{Sym}^n(\det)$. Explicitly, fix an identification \det of $\wedge^2 V^\vee$ with K . For $x_i, y_i \in V^\vee$ the pairing

$$(2.37) \quad (x_1 \otimes \cdots \otimes x_n, y_1 \otimes \cdots \otimes y_n) = \frac{1}{n!} \sum_{\sigma \in S_n} \prod_{i=1}^n \det(x_{\sigma(i)} \wedge y_i)$$

factors through the symmetric power. In particular if $\det(u \wedge v) = 1$ then

$$(2.38) \quad (u^i v^{n-i}, u^{n-i} v^i) = \frac{i!(n-i)!}{n!} (-1)^{n-i}$$

and $(u^i v^{n-i}, u^{i'} v^{n-i'}) = 0$ if $i' \neq n - i$. Note that this pairing is symmetric for even n and alternating for odd n .

If $\varphi \in \mathcal{C}(M)$ and $\omega \in \Omega(M)$ we let (φ, ω) be the \mathbb{C}_p -valued function obtained from the above pairing of M with itself (defined in the complement of some large enough affinoid in \mathfrak{X}), and

$$(2.39) \quad \langle \varphi, \omega \rangle = \lim_{\mathcal{F}'} \sum_{e \in \text{Ends}(\mathcal{F}')} \text{res}_e((\varphi, \omega)).$$

The usual Morita duality implies that this is a perfect pairing between $\mathcal{C}(M)/\mathcal{R}(M)$ and $\Omega(M)$ with values in \mathbb{C}_p . Of course,

$$(2.40) \quad \mathcal{C}(M)/\mathcal{R}(M) \simeq \mathcal{C}^{an}(M)/\mathbb{C}_p \otimes_K M.$$

We now study the filtrations in this pairing.

Lemma 2.4. Φ^k and $\Omega \otimes F^{n+1-k}$ are orthogonal to each other under the pairing of $\mathcal{C}/\mathcal{R}(M)$ and $\Omega(M)$.

Proof. This is a computation that reduces to the fact that for $0 \leq d \leq k-1$

$$(2.41) \quad \sum_{m=0}^k \frac{m^d}{m!(k-m)!} (-1)^m = 0.$$

To prove this well-known identity, differentiate d times the function $(1 - e^x)^k$ and evaluate at $x = 0$. \square

Corollary 2.5. The isomorphisms Θ'_{n-k} and Ψ_k induce a duality

$$(2.42) \quad (\mathcal{C}/\mathcal{R})(n-2k) \times \Omega(2k-n) \rightarrow \mathbb{C}_p$$

identifying each side with the strong dual of the other side. This is nothing else than the pairing

$$(2.43) \quad \Phi^k(\mathcal{C}/\mathcal{R})/\Phi^{k+1}(\mathcal{C}/\mathcal{R}) \times F^{n-k}\Omega/F^{n-k+1}\Omega \rightarrow \mathbb{C}_p$$

induced from the lemma (where we have written $F^{n-k}\Omega$ for $\Omega \otimes F^{n-k}$).

Corollary 2.6. $\Phi^k(\mathcal{C}/\mathcal{R})$ and $F^{n+1-k}\Omega$ are exact annihilators of each other.

Another interpretation. It is often desirable to have an interpretation for $\mathcal{C}/\mathcal{R}(m)$ as we had when $m = 0$:

$$(2.44) \quad \mathcal{C}/\mathcal{R} = \mathcal{C}^{an}/\mathbb{C}_p.$$

Note that ($m \geq 0$)

$$(2.45) \quad \mathcal{C}/\mathcal{R} = \mathcal{C}^{an}[m\infty]/P_m$$

where $\mathcal{C}^{an}[m\infty]$ denotes the functions that are locally analytic, *except* near ∞ where

$$(2.46) \quad \text{ord}_\infty \varphi + m \geq 0,$$

and

$$(2.47) \quad P_m = \mathcal{C}^{an}[m\infty] \cap \mathcal{R}.$$

This means that at ∞ the functions are allowed to have a polynomial part of degree at most m , and P_m is of course the polynomials of degree at most m . If $m < 0$, then $P_m = 0$, and $\mathcal{C}^{an}[m\infty]$ are functions vanishing to order $|m|$ at infinity. The point is that these spaces are invariant under the m th twisted action of G : if φ is of this sort,

$$(2.48) \quad (c\zeta + d)^m \varphi \left(\frac{a\zeta + b}{c\zeta + d} \right)$$

is also there: you only have to check zeroes and poles at $\zeta = \infty$ and $\zeta = -d/c$.

In particular we get, when $k = 0$ and $n \geq 0$

$$(2.49) \quad \Omega(-n)'_b \simeq (\mathcal{C}^{an}[n\infty]/P_n)(n)$$

(the twist reminding us of the action of G). This is [3], Theorem 2.2.1 (their k is our $n + 2$, and their $\mathcal{O}(k)$ is our $\mathcal{O}(-k)$).

The subspace $d\mathcal{O}(M)$ is closed in $\Omega(M)$. We have seen that its annihilator is the space $\mathcal{C}^\infty(M)/\mathbb{C}_p \otimes_K M$, which is therefore dual to $H^1_{dR}(\mathcal{X}; M)$, or to $C^1_{har}(M)$. Under projection of $\mathcal{C}/\mathcal{R} \otimes M$ modulo Φ^1 onto $\mathcal{C}/\mathcal{R}(n) = (\mathcal{C}^{an}[n\infty]/P_n)(n)$, the space $\mathcal{C}^\infty(M)/\mathbb{C}_p \otimes_K M$ gets mapped isomorphically onto

$$(2.50) \quad \mathcal{C}^\infty(M)/\mathbb{C}_p \otimes_K M \simeq (\mathcal{C}^{pol,n}[n\infty]/P_n)(n).$$

Here $\mathcal{C}^{pol,n}$ is the space of locally-polynomial-of-degree- n functions on \mathcal{A} and the $[n\infty]$ is there only to remind us that they are locally analytic everywhere except at ∞ where they have a pole of order n . The notation (n) refers to the G action. The map can be computed. It sends $\chi u^i v^{n-i}$ for a locally constant function χ to $\chi \zeta^i$.

The (strong) dual of the exact sequence

$$(2.51) \quad 0 \rightarrow \mathcal{O}(n)/P_n(n) \rightarrow F^n \Omega \rightarrow H_{dR}^1(\mathfrak{X}; M) \rightarrow 0$$

in which the first arrow takes f to $\frac{1}{n!} f^{(n+1)}(z)(u-zv)^n dz$, is the exact sequence

$$(2.52) \quad 0 \leftarrow \mathcal{C}^{an}[(-n-2)\infty](-n-2) \leftarrow (\mathcal{C}^{an}[n\infty]/P_n)(n) \leftarrow (\mathcal{C}^{pol,n}[n\infty]/P_n)(n) \leftarrow 0$$

where the first (backward) arrow is $\frac{(-1)^n}{n!} \left(\frac{d}{dz}\right)^{n+1}$. Note that this operator kills the principal (polynomial) part at ∞ and increases the order of the zero at infinity to $n+2$.

Summary. On \mathfrak{X} we have

$$(2.53) \quad \begin{array}{ccccccc} 0 & \rightarrow & \mathcal{O}(n)/P_n(n) & \xrightarrow{\frac{1}{n!} \left(\frac{d}{dz}\right)^{n+1}} & \mathcal{O}(-n-2) = F^n \Omega & \rightarrow & H_{dR}^1(\mathfrak{X}; M_n) \rightarrow 0 \\ & & \cap & & \cap & & \parallel \\ 0 & \rightarrow & d\mathcal{O}(M_n) & \rightarrow & \Omega(M_n) & \rightarrow & C_{har}^1(M_n) \rightarrow 0 \end{array}$$

The strong dual of this diagram is

$$(2.54) \quad \begin{array}{ccccc} \mathcal{C}^{an}[(-n-2)\infty](-n-2) & \leftarrow & (\mathcal{C}^{an}[n\infty]/P_n)(n) & \leftarrow & (\mathcal{C}^{pol,n}[n\infty]/P_n)(n) \\ \uparrow & & \uparrow \text{ mod } \Phi^1 & & \parallel \text{ mod } \Phi^1 \\ \mathcal{C}^{an}(M_n)/\mathcal{C}^\infty(M_n) & \leftarrow & \mathcal{C}/\mathcal{R}(M_n) & \leftarrow & \mathcal{C}^\infty(M_n)/\mathbb{C}_p \otimes_K M_n \end{array}$$

The maps I and J (= the Poisson kernel) in [3] are the duality isomorphisms

$$(2.55) \quad \begin{array}{l} (I_k) \quad I \quad : \quad \mathcal{O}(-n-2)'_b \simeq (\mathcal{C}^{an}[n\infty]/P_n)(n) \\ \lambda \quad \mapsto \quad I(\lambda)(\zeta) = \lambda\left(\frac{1}{z-\zeta}\right) \end{array}$$

and

$$(2.56) \quad \begin{array}{l} ({}^t I_k \quad = \quad J_k) \quad J : (\mathcal{C}^{an}[n\infty]/P_n)(n)'_b \simeq \mathcal{O}(-n-2) \\ \mu \quad \mapsto \quad \int_{\mathbb{P}^1(K)} \frac{1}{z-\zeta} d\mu(\zeta). \end{array}$$

3. Bounded forms and bounded cohomology

3.1. Bounded differential forms. Let \mathbf{M} be an integral structure on M (as defined in the introduction).

Definition 3.1. $f \in \mathcal{O}(M)$ is \mathbf{M} -integral if for any $\sigma \in \mathcal{F}$ (vertex or an open edge), $f(z) \in \mathbf{M}(\sigma) \otimes_{\mathcal{O}_K} \mathcal{O}_{\mathbb{C}_p}$ for all $z \in \mathfrak{X}_\sigma = r^{-1}(|\sigma|)$, where $r : \mathfrak{X} \rightarrow |\mathcal{F}|$ is the reduction map.

Definition 3.2. (i) The canonical integral structure Ω on $\underline{\Omega}$ is the $\mathcal{O}_{\mathbb{C}_p}$ -subsheaf of all the sections ω such that the image of ω in the fiber at any $z_0 \in \mathfrak{X}$, denoted $\omega|_{z_0}$, belongs to the $\mathcal{O}_{\mathbb{C}_p}$ -span of $d \log(z - \zeta)|_{z_0}$, for all $\zeta \in K$. (Note that this submodule is independent of the coordinate z .)

(ii) Call $\omega \in \Omega(M)$ \mathbf{M} -integral if for any σ and any $z_0 \in \mathfrak{X}_\sigma$

$$(3.1) \quad \omega|_{z_0} \in \Omega|_{z_0} \otimes_{\mathcal{O}_K} \mathbf{M}(\sigma).$$

(iii) As before, bounded means integral w.r.t. some integral structure.

Proposition 3.1. The residue map $\omega \mapsto c_\omega$ maps a bounded M -valued differential form to a bounded M -valued harmonic cochain.

$$(3.2) \quad \Omega(M)^{bnd} \rightarrow C_{har}^1(M)^{bnd}.$$

Proof. Let \mathbf{M} be an integral structure such that ω is \mathbf{M} -integral. Fix a vertex v and let $\mathfrak{X}(v)$ be the preimage, under the reduction map r , of the vertex v and the $q + 1$ edges e_0, \dots, e_q starting at v . We label them so that e_0 is the unbounded edge (i.e. $\infty \in \mathcal{A}_{e_0}$). Let

$$(3.3) \quad \rho = \text{dist}(z, \mathbb{P}_K^1)$$

be the distance from any $z \in \mathfrak{X}_v = r^{-1}(v)$ to the boundary of \mathfrak{X} . Let $\zeta_i \in \mathcal{A}_{e_i}$ and $\zeta_0 \in \mathcal{A}_{e_0}$. Then the Mittag-Leffler decomposition of ω in $\mathfrak{X}(v)$ is

$$(3.4) \quad \omega = \sum_{i=0}^q f_{e_i}(z) dz$$

where for $1 \leq i \leq q$

$$(3.5) \quad f_{e_i} = \sum_{v=1}^{\infty} a_{i,v}(z - \zeta_i)^{-v}$$

and

$$(3.6) \quad f_{e_0} = \sum_{v=0}^{\infty} a_{0,v}(z - \zeta_0)^v.$$

For $z \in \mathfrak{X}_v$ and $1 \leq j \leq q$, $|z - \zeta_j| = \rho$ so $(z - \zeta_j)^{-1} dz$ is a basis of the fiber $\Omega|_z$ over $\mathcal{O}_{\mathbb{C}_p}$.

Let l be any linear functional $\mathbf{M}(v) \rightarrow \mathcal{O}_K$. For any $z \in \mathfrak{X}_v$ $l(\omega) \in \Omega|_z$ and therefore

$$(3.7) \quad \left| \sum_{i=0}^q l(f_{e_i}(z)) \right| \leq \rho^{-1}.$$

By the Mittag-Leffler theorem each $|l(f_{e_i}(z))| \leq \rho^{-1}$ throughout \mathfrak{X}_v , and by the equality of the Gauss norm and the sup norm, for every $1 \leq i \leq q$, $v \geq 1$

$$|l(a_{i,v})| \leq \rho^{v-1}$$

and in particular $|l(a_{i,1})| \leq 1$ for every l , so $a_{i,1} \in \mathcal{O}_{\mathbb{C}_p} \otimes_{\mathcal{O}_K} \mathbf{M}(v)$. But these are the residues of ω along e_i , and the residue along e_0 is, up to a sign, the sum of them. It follows that $c_\omega(e_i) \in \mathbf{M}(v) \subset \mathbf{M}(e_i)$ for every i . Since v was arbitrary, c_ω is \mathbf{M} -integral. \square

Theorem 3.2. *The residue homomorphism induces an isomorphism*

$$(3.8) \quad \Omega(M_n)^{bnd} \cap F^n \Omega \simeq C_{har}^1(M_n)^{bnd}.$$

We begin by checking the injectivity. Let

$$(3.9) \quad \omega = f(z)(u - zv)^n dz$$

be a form in $F^n \Omega$. To proceed we need to introduce some notation. For every vertex v let ρ_v be the diameter of the set \mathfrak{X}_v , which is also the distance from it to the boundary of \mathfrak{X} . Let $\mathcal{E}(v)$ be the set of the $q+1$ oriented edges starting at v . If e is an edge and $\mathfrak{X}_e = r^{-1}(|e|)$ is the corresponding annulus, then its complement consists of two closed disks, and we denote by ρ_e the diameter of the disk which is bounded, and ζ_e a K -rational point in it, chosen at random. If $e \in \mathcal{E}(v)$ then $\rho_e = |\pi| \rho_v$ if e is bounded, and $\rho_e = \rho_v$ if e is unbounded.

Lemma 3.3. *There exist functions f_e , one for each oriented edge e , with the following properties:*

(i) *If e is bounded, f_e is defined in the complement of $\{z \mid |z - \zeta_e| \leq \rho_e\}$ by a convergent Laurent series without constant term*

$$(3.10) \quad f_e = \sum_{v=1}^{\infty} a_{e,v} (z - \zeta_e)^{-v}.$$

If e is unbounded, f_e is defined in $\{z \mid |z - \zeta_e| < |\pi|^{-1} \rho_e\}$ by a convergent Taylor series

$$(3.11) \quad f_e = \sum_{v=0}^{\infty} a_{e,v} (z - \zeta_e)^v.$$

(ii) *If \mathcal{T}' is a finite connected subtree of \mathcal{T} and $\mathcal{E}(\mathcal{T}')$ is the set of edges not in \mathcal{T}' but adjacent to it, oriented away from \mathcal{T}' , then if $\mathfrak{X}(\mathcal{T}') = r^{-1}(|\mathcal{T}'|)$ is the affinoid reducing to \mathcal{T}' , we have the Mittag-Leffler decomposition*

$$(3.12) \quad f|_{\mathfrak{X}(\mathcal{T}')} = \sum_{e \in \mathcal{E}(\mathcal{T}')} f_e.$$

(iii) For every bounded edge e

$$(3.13) \quad \text{res}_e \omega = \text{res}_e f_e(z)(u - zv)^n dz.$$

Proof. For a given finite subtree \mathcal{T}' , the existence of a collection of functions as in (i), indexed by $e \in \mathcal{E}(\mathcal{T}')$, satisfying (ii), is guaranteed by the Mittag-Leffler decomposition for the affinoid subdomain $\mathfrak{X}(\mathcal{T}')$. Let us employ temporarily the notation $f_{e, \mathcal{T}'}$ ($e \in \mathcal{E}(\mathcal{T}')$) for this collection, to emphasize the dependence on \mathcal{T}' . We claim that the f_e do not depend on \mathcal{T}' . It is enough to consider the case where \mathcal{T}'' is obtained from \mathcal{T}' by adding a vertex u and an edge $e_0 = (v, u)$ for some $v \in \mathcal{T}'$. Let e_1, \dots, e_q be the edges starting at u different from \bar{e}_0 . Define

$$(3.14) \quad f_{e_0, \mathcal{T}''} = \sum_{i=1}^q f_{e_i, \mathcal{T}''}$$

and $f_{e, \mathcal{T}'} = f_{e, \mathcal{T}''}$ for every $e \in \mathcal{E}(\mathcal{T}') \cap \mathcal{E}(\mathcal{T}'')$. If e_0 is bounded, then so are all the e_i ($1 \leq i \leq q$), and it is readily checked that $f_{e_0, \mathcal{T}''}$ is holomorphic in the complement of the bounded disk defined by e_0 , and vanishes at infinity, so is given by a Laurent series without constant term. If e_0 is unbounded, then precisely one of the e_i is unbounded, and again it follows that $f_{e_0, \mathcal{T}''}$ is holomorphic in the bounded disk defined by e_0 and given by a convergent Taylor series there. Finally the fact that $f|_{\mathfrak{X}(\mathcal{T}')} = \sum_{e \in \mathcal{E}(\mathcal{T}')} f_{e, \mathcal{T}'}$ follows from the analogous decomposition for \mathcal{T}'' by restricting the domain.

To prove (iii), look at the Mittag-Leffler decomposition in $\mathfrak{X}(v)$ where v is the origin of e , and note that all the other $f_{e'}$ appearing in that decomposition extend holomorphically across $\{z \mid |z - \zeta_e| \leq \rho_e\}$ so do not contribute to the residue. \square

Lemma 3.4. *Let $\|f_e\|$ denote the sup norm of f_e in $\{z \mid |z - \zeta_e| \geq |\pi^{-1}| \rho_e\}$ if e is bounded, and in $\{z \mid |z - \zeta_e| \leq \rho_e\}$ if e is unbounded. Note that*

$$(3.15) \quad \|f_e\| = \|f_e\|_{\mathfrak{X}_v}$$

if v is the origin of e . Assume that ω is bounded and that all its residues vanish. Then

$$(3.16) \quad \|f_e\| \rho_e^{\frac{n+2}{2}}$$

are bounded as e runs over the oriented edges of the tree.

Proof. The vanishing of the residues of ω is equivalent to the vanishing of the residues along every e of $z^i f(z) dz$ for $0 \leq i \leq n$. If e is bounded this simply means that the Laurent series giving f_e starts in degree $-n - 2$, and has no terms in higher degrees, or equivalently that it vanishes at infinity to order which is at least $n + 2$.

Let v_0 be the standard vertex of \mathcal{T} and pick an arbitrary vertex $v = \gamma(v_0)$. Since ω is bounded there exists an \mathcal{O}_K -lattice M_0 in M independent of v or γ such that for $w \in \mathfrak{X}_v$

$$(3.17) \quad \omega|_w \in \Omega|_w \otimes_{\mathcal{O}_K} \gamma(M_0).$$

Write $w = \gamma(z)$ for $z \in \mathfrak{X}_{v_0}$ and note that $\Omega|_w = (\gamma^{-1})^* \Omega|_z$. It follows that

$$(3.18) \quad \begin{aligned} (\gamma^{-1}\omega)|_z &= (1 \otimes \gamma^{-1})\gamma^*(\omega|_w) \\ &\in \Omega|_z \otimes_{\mathcal{O}_K} M_0. \end{aligned}$$

Let us invoke the Mittag-Leffler decomposition of ω at v :

$$(3.19) \quad \omega|_{\mathfrak{X}_v} = \sum_{e \in \mathcal{E}(v)} f_e(z)(u - zv)^n dz.$$

Then

$$(3.20) \quad \gamma^{-1}\omega|_{\mathfrak{X}_{v_0}} = \sum_{e \in \mathcal{E}(v_0)} f_{\gamma e}(\gamma z)(cz + d)^{-n-2}(u - zv)^n dz$$

where (c, d) is the bottom row of γ . The key to the lemma is the observation that since all the residues of $z^i f(z) dz$ for $0 \leq i \leq n$ vanish, this last sum is the Mittag Leffler decomposition of $\gamma^{-1}\omega$ at v_0 . Indeed, let us distinguish three types of $e \in \mathcal{E}(v_0)$. If γe and e are both bounded, then

$$(3.21) \quad g_e(z) = f_{\gamma e}(\gamma z)(cz + d)^{-n-2}$$

is holomorphic everywhere in $\{z \mid |z - \zeta_e| > \rho_e\}$, including the questionable point $-d/c$, since $f_{\gamma e}$ vanishes at infinity to order $\geq n+2$. It also clearly vanishes at infinity (in fact, to order $\geq n+2$). If γe is bounded but e is unbounded, the same analysis applies, noting that γ carries the bounded disk $\{z \mid |z - \zeta_e| \leq \rho_e\}$ onto the unbounded disk $\{w \mid |w - \zeta_{\gamma e}| \geq |\pi^{-1}| \rho_{\gamma e}\}$. If γe is unbounded then $-d/c$ does not belong to the domain of definition of g_e (whether bounded or not) so again g_e is holomorphic there. It follows that g_e satisfy the conditions characterizing the Mittag-Leffler decomposition of $\gamma^{-1}\omega$ at v_0 , so by the uniqueness of the decomposition,

$$(3.22) \quad \gamma^{-1}\omega = \sum_{e \in \mathcal{E}(v_0)} g_e(z)(u - zv)^n dz$$

is the Mittag-Leffler decomposition of $\gamma^{-1}\omega$ at v_0 .

Now we have seen above that the boundedness of ω is translated to the uniform boundedness of all the $\gamma^{-1}\omega$ (in the usual sense) on \mathfrak{X}_{v_0} . This means that $\|g_e\| = \|g_e\|_{\mathfrak{X}_{v_0}}$ are uniformly bounded (for every γ). However, if ζ is a point in one of

the bounded disks in the complement of \mathfrak{X}_{v_0} , for which $\gamma(\zeta)$ also lies in one of the bounded disks in the complement of \mathfrak{X}_v ($v = \gamma(v_0)$), then

$$(3.23) \quad \gamma(z) - \gamma(\zeta) = \frac{z - \zeta}{(cz + d)(c\zeta + d)}$$

and

$$(3.24) \quad |c\zeta + d| = |c||\zeta - (-d/c)| = |c||z - (-d/c)| = |cz + d|$$

(note that ζ and $-d/c$ do not lie in the same disk, as the latter is mapped to ∞ by γ and the first to a bounded disk). This computation shows that

$$(3.25) \quad |cz + d|^2 = \rho_{v_0} / \rho_v.$$

It follows that

$$(3.26) \quad \|g_e\| = \|f_{\gamma e}\| (\rho_v / \rho_{v_0})^{\frac{n+2}{2}}$$

and the lemma follows because ρ_v and $\rho_{\gamma e}$ at most differ by a factor of $|\pi|$. □

Proposition 3.5. *A bounded form in $F^n\Omega$, all of whose residues vanish, is identically 0. In other words, the map*

$$(3.27) \quad \Omega(M_n)^{bnd} \cap F^n\Omega \rightarrow C_{har}^1(M_n)^{bnd}$$

is injective.

Proof. (compare [4], (3.9.5) where the same proof was applied to Γ -invariant forms for a discrete cocompact subgroup Γ). Let ω be as in the proposition. From the last lemma we know that there exists a constant R such that

$$(3.28) \quad \|f_e\| (|\pi^{-1}| \rho_e)^{\frac{n+2}{2}} \leq R$$

for every e . Fix an affinoid $\mathcal{K} = \mathfrak{X}(\mathcal{T}')$ for a finite connected subtree \mathcal{T}' and let $0 < \delta < 1$ be small enough so that if $e \in \mathcal{E}(\mathcal{T}')$ is bounded, $\delta < \rho_e$ and if e is unbounded, $\rho_e < \delta^{-1}$. Fix a second affinoid $\mathcal{K}' = \mathfrak{X}(\mathcal{T}'')$ containing \mathcal{K} so that its ends satisfy the same estimates on ρ_e with δ replaced by $|\pi|\delta^4$. Let $z \in \mathcal{K}$ and consider the Mittag-Leffler decomposition which corresponds to \mathcal{K}' :

$$(3.29) \quad \omega = \sum_{e \in \mathcal{E}(\mathcal{T}'')} f_e(z)(u - zv)^n dz.$$

If e is bounded, $|z - \zeta_e| > \delta$ since $z \in \mathcal{K}$, so

$$(3.30) \quad \begin{aligned} |a_{e,v}(z - \zeta_e)^{-v}| &\leq \delta^{-v} \|f_e\| (|\pi^{-1}| \rho_e)^v \leq \delta^{-v} R (|\pi^{-1}| \rho_e)^{v - \frac{n+2}{2}} \\ &\leq R \delta^{3v - 2(n+2)} \leq R \delta^{n+2}. \end{aligned}$$

In the last step we used the fact that only $v \geq n + 2$ count. On the other hand, if e is the unbounded end, and $v \geq 0$,

$$(3.31) \quad \begin{aligned} |a_{e,v}(z - \zeta_e)^v| &\leq \|f_e\| \leq R(|\pi^{-1}|\rho_e|)^{-\frac{n+2}{2}} \\ &\leq R|\pi|^{n+2}\delta^{2(n+2)}. \end{aligned}$$

In any case, letting $\delta \rightarrow 0$, we see that f vanishes on \mathcal{X} . Since \mathcal{X} was arbitrary, $\omega = 0$. \square

3.2. The theorem of Amice-Velu-Vishik. Next we deal with surjectivity.

Proposition 3.6. *Every $c \in C_{har}^1(M_n)^{bnd}$ is obtained as c_ω for an $\omega \in \Omega(M_n)^{bnd} \cap F^n\Omega$.*

We follow the sketch of [3], 2.3.2, fill in some details, but omit verifications of convergence which may be found at [10] and [2].

Let $c \in C_{har}^1(M_n)^{bnd}$ and let

$$(3.32) \quad \lambda_c \in \left(C^{pol,n}[n\infty]/P_n \right) (n)'$$

be the corresponding linear functional defined by

$$(3.33) \quad \lambda_c(\zeta^i \chi_{U(\varepsilon)}) = \left(c(\varepsilon), u^i v^{n-i} \right)$$

for $0 \leq i \leq n$, ε an oriented edge of \mathcal{T} and $U(\varepsilon)$ the subset of $\mathbb{P}^1(K) = \text{Ends}(\mathcal{T})$ to which ε “is pointing” (denoted \mathcal{A}_ε before). Here χ_U is the characteristic function of U . Let

$$(3.34) \quad g = \begin{pmatrix} \pi^{-m} & -\pi^{-m}a \\ 0 & \pi^m \end{pmatrix} \in G = SL_2(K)$$

so that $g : D(a, |\pi^{2m}|) \simeq D(0, 1)$. Then in the G action on $(C^{pol,n}[n\infty]/P_n) (n)$

$$(3.35) \quad g^{-1} \left(\zeta^j \chi_{D(0,1)} \right) = \pi^{mn} g(\zeta)^j \chi_{D(a, |\pi^{2m}|)}$$

so

$$(3.36) \quad \begin{aligned} &\lambda_c \left((\zeta - a)^j \chi_{D(a, |\pi^{2m}|)} \right) \\ &= \lambda_c \left(\pi^{(2j-n)m} g^{-1} \left(\zeta^j \chi_{D(0,1)} \right) \right) \\ &= \pi^{(2j-n)m} g(\lambda_c) \left(\zeta^j \chi_{D(0,1)} \right) \\ &= \pi^{(2j-n)m} \left\langle c(g^{-1}\varepsilon_0), g^{-1}(u^j v^{n-j}) \right\rangle \end{aligned}$$

where ε_0 is the edge corresponding to the standard annulus such that $U(\varepsilon_0) = D(0, 1) \cap K$. Now by boundedness, and by the G -equivariance of the pairing on M_n , the expression in the brackets belongs to some lattice which is independent of g . We conclude that as long as $2j > n$ the whole thing tends to 0 as $m \rightarrow \infty$.

A similar computation holds in the unbounded disk: take

$$(3.37) \quad g = \begin{pmatrix} 0 & -\pi^{-m} \\ \pi^m & 0 \end{pmatrix}$$

so that $g : D(\infty, |\pi^{-2m}|) \simeq D(0, 1)$ and

$$(3.38) \quad g^{-1} \left(\zeta^j \chi_{D(0,1)} \right) = (-1)^j \pi^{(n-2j)m} \zeta^{n-j} \chi_{D(\infty, |\pi^{-2m}|)}$$

and

$$(3.39) \quad \begin{aligned} & \lambda_c \left(\zeta^{n-j} \chi_{D(\infty, |\pi^{-2m}|)} \right) \\ &= (-1)^j \pi^{(2j-n)m} \left\langle c(g^{-1} \varepsilon_0), g^{-1}(u^j v^{n-j}) \right\rangle \end{aligned}$$

and the same conclusion holds. Note that this time it is ζ^{n-j} for $2j > n$ whose integrals over smaller and smaller neighborhoods of ∞ tend to 0.

As a corollary of the computations above we can make the following definition. Let $\varphi \in (C^{an}[n\infty]/P_n)(n)$ and c as above. Extend λ_c to φ by taking a large subtree \mathcal{T}' , finding for any $\varepsilon \in \text{Ends}(\mathcal{T}')$ a Taylor (resp. MacLaurin) expansion of φ in $U(\varepsilon)$ centered at some $\xi_\varepsilon \in U(\varepsilon)$ (resp. at ∞ if $U(\varepsilon)$ is an unbounded disk), letting $T_{\xi_\varepsilon}^n(\varphi)$ be the *truncation* of this expansion at degrees $\leq n$ (resp. degrees ≥ 0) and setting

$$(3.40) \quad \tilde{\lambda}_c(\varphi) = \lim_{\mathcal{T}'} \sum_{\varepsilon} \left\langle c(\varepsilon), T_{\xi_\varepsilon}^n(\varphi) \right\rangle$$

where you identify $P_n(n)$ with M_n via $\zeta^i \mapsto u^i v^{n-i}$.

Theorem 3.7. *The limit exists and is independent of the choices involved. Furthermore if we put*

$$(3.41) \quad \omega_c = \left(\int_{\mathbb{P}^1(K)} \frac{d\tilde{\lambda}_c(\zeta)}{z - \zeta} \right) (u - zv)^n dz$$

then the Poisson integral (which is defined for every z pointwise since the integrand is locally analytic in ζ) is a holomorphic function of z so $\omega_c \in F^n \Omega$ and it represents the functional $\tilde{\lambda}_c$ under Morita duality. Moreover, ω_c is bounded.

For the verification, based on the growth estimates on λ_c derived above, see [10], Section 11 (where the same estimates are used in the construction of p -adic L functions), or [2], Theorem 2.5 (where it is shown that λ_c extends to a distribution "of order n ", and not just to locally analytic functions).

References

- [1] G. ALON & E. DE SHALIT – “On the cohomology of Drinfel’d’s p -adic symmetric domain”, *Israel J. Math.* **129** (2002), p. 1–20.
- [2] P. CLOMEZ – “Fonctions d’une variable p -adique”, preprint, 2005.
- [3] S. DASGUPTA & J. TEITELBAUM – “The p -adic upper half plane”, preprint.
- [4] E. DE SHALIT – “Eichler cohomology and periods of modular forms on p -adic Schottky groups”, *J. Reine Angew. Math.* **400** (1989), p. 3–31.
- [5] ———, “Residues on buildings and de Rham cohomology of p -adic symmetric domains”, *Duke Math. J.* **106** (2001), no. 1, p. 123–191.
- [6] V. G. DRINFEL’D – “Coverings of p -adic symmetric domains”, *Funkcional. Anal. i Priložen.* **10** (1976), no. 2, p. 29–40.
- [7] L. GERRITZEN & M. VAN DER PUT – *Schottky groups and Mumford curves*, Lecture Notes in Mathematics, vol. 817, Springer, Berlin, 1980.
- [8] E. GROSSE-KLÖNNE – “Sheaves of bounded p -adic logarithmic differential forms”, *Annales Scientifiques de l’école Normale Supérieure* **40** (2007), no. 3, p. 351 – 386.
- [9] A. IOVITA & M. SPIESS – “Logarithmic differential forms on p -adic symmetric spaces”, *Duke Math. J.* **110** (2001), no. 2, p. 253–278.
- [10] B. MAZUR, J. TATE & J. TEITELBAUM – “On p -adic analogues of the conjectures of Birch and Swinnerton-Dyer”, *Invent. Math.* **84** (1986), no. 1, p. 1–48.
- [11] P. SCHNEIDER & U. STUHLER – “The cohomology of p -adic symmetric spaces”, *Invent. Math.* **105** (1991), no. 1, p. 47–122.
- [12] P. SCHNEIDER – “The cohomology of local systems on p -adically uniformized varieties”, *Math. Ann.* **293** (1992), no. 4, p. 623–650.
- [13] ———, *Nonarchimedean functional analysis*, Springer Monographs in Mathematics, Springer-Verlag, Berlin, 2002.

ON ARAKELOV VECTOR BUNDLES OVER ARITHMETIC CURVES

Norbert Hoffmann

Mathematisches Institut, Georg-August Universität Göttingen
E-mail: hoffmann@uni-math.gwdg.de

Abstract. Here ‘arithmetic curve’ means the set of all places X of a number field K . Under the classical analogy between number fields and function fields, Arakelov vector bundles over X correspond to algebraic vector bundles over a smooth projective curve C . The text discusses analogues over X of three results on such bundles over C : Riemann-Roch, the Narasimhan-Seshadri theorem, and Faltings’ criterion for semistability.

Introduction

This text consists of the author’s talk at the conference on Symmetries in Algebra and Number Theory in Göttingen, October 2008. It deals with the following question: What corresponds, under the classical analogy between function fields and number fields, to the theory of vector bundles on algebraic curves?

After recalling the analogue ‘arithmetic curve’ of an algebraic curve in section 1, and the analogue ‘Arakelov vector bundle’ of a vector bundle on it in section 2, each of the remaining three sections considers one important result about vector bundles on algebraic curves and discusses arithmetic analogues.

Section 3 is about Riemann-Roch; its arithmetic analogue turns out to be much harder, and in particular to contain the problem of lattice sphere packings.

October 2008.

2000 Mathematics Subject Classification. 14G40.

Section 4 deals with the Narasimhan-Seshadri theorem [9] about unitary structures on stable holomorphic vector bundles, or rather its corollary that tensor products preserve semistability (in characteristic zero). Its arithmetic analogue is an interesting open problem posed by J.-B. Bost.

Finally, section 5 recalls Faltings' criterion [5, 6] for semistability in terms of tensor products, and reports on the (weaker) arithmetic analogue [8] of it.

1. Arithmetic curves

Let K be a number field. Its ring of integers $\mathcal{O}_K \subset K$ is a Dedekind domain, so its spectrum

$$\text{Spec}(\mathcal{O}_K) := \{\mathfrak{p} \subset \mathcal{O}_K \text{ prime ideal}\}$$

behaves like an affine algebraic curve. The elements of $\text{Spec}(\mathcal{O}_K)$, except for the generic point $\eta = (0)$, are called the *finite places* of K . The *infinite places* of K are the embeddings of K into \mathbb{C} , up to complex conjugation. Their set

$$X_\infty := \{\iota : K \hookrightarrow \mathbb{C} \text{ embedding}\} / \iota \sim \bar{\iota}$$

is finite. The *arithmetic curve* associated to K is the disjoint union of sets

$$X := \text{Spec}(\mathcal{O}_K) \cup X_\infty.$$

In other words, X consists of the generic point η and of all places of K . The set $X \setminus \{\eta\}$ of all places of K is also in canonical bijection to the set

$$\{\nu : K \longrightarrow \mathbb{R}_{\geq 0} \text{ valuation}\} / \sim$$

of all valuations of the field K , up to equivalence of valuations. Here finite places correspond to non-archimedean (p -adic) valuations; infinite places correspond to archimedean valuations obtained from the absolute value on \mathbb{C} .

One justification for calling $X = \text{Spec}(\mathcal{O}_K) \cup X_\infty$ a 'curve' comes from the classical analogy between number fields and function fields, which can be traced back at least to [4]. To state it, suppose that X is now a compact Riemann surface (or a smooth projective algebraic curve), and let $\mathbb{C}(X)$ denote the field of meromorphic functions on X . Then X is in canonical bijection to the set

$$\{\nu : \mathbb{C}(X) \longrightarrow \mathbb{R}_{\geq 0} \text{ valuation}\} / \sim$$

of all valuations of the field $\mathbb{C}(X)$, up to equivalence of valuations. Here a point $p \in X$ corresponds to the valuation $\nu_p : \mathbb{C}(X) \rightarrow \mathbb{R}_{\geq 0}$ defined by $\nu_p(f) := c^{-\text{ord}_p(f)}$ for some constant $c > 1$, where $\text{ord}_p(f) \in \mathbb{Z}$ denotes the order of p as a zero of the

meromorphic function f ; in particular, $v_p(f) \leq 1$ means that f is holomorphic at p . Thus the underlying set of the Riemann surface X can be reconstructed from the function field $\mathbb{C}(X)$; applying the same construction to the number field K instead yields the arithmetic curve $X = \text{Spec}(\mathcal{O}_K) \cup X_\infty$.

One can think of X as a ‘compactification’ of the ‘affine curve’ $\text{Spec}(\mathcal{O}_K)$ by adding the finite set of points X_∞ ‘at infinity’. For a Riemann surface X , compactness is equivalent to the relation

$$(1) \quad \sum_{p \in X} \text{ord}_p(f) = 0 \quad \text{for all } f \in \mathbb{C}(X) \setminus \{0\}.$$

An analogue of (1) for $X = \text{Spec}(\mathcal{O}_K) \cup X_\infty$ an arithmetic curve is the well-known *product formula*

$$(2) \quad \prod_{v: K \rightarrow \mathbb{R}_{\geq 0}} v(a) = 1 \quad \text{for all } a \in K \setminus \{0\}$$

where the product is taken over all suitably normalized valuations v . This formula justifies to think of $X = \text{Spec}(\mathcal{O}_K) \cup X_\infty$ as some sort of ‘compact curve’.

2. Arakelov vector bundles on arithmetic curves

Definition 1. An Arakelov vector bundle E over the above arithmetic curve $X = \text{Spec}(\mathcal{O}_K) \cup X_\infty$ consists of:

- a locally free \mathcal{O}_K -module $E_{\mathcal{O}_K}$ of finite rank, and
- for each infinite place $v \in X_\infty$, a scalar product $\langle _, _ \rangle_{E, v}$ on the K_v -vector space $E_{K_v} := E_{\mathcal{O}_K} \otimes K_v$, where $K_v \cong \mathbb{R}$ or \mathbb{C} denotes the completion of K .

Definition 2. Let E be an Arakelov vector bundle over the arithmetic curve $X = \text{Spec}(\mathcal{O}_K) \cup X_\infty$. The elements of the finite pointed set

$$\Gamma(X, E) := \{s \in E_{\mathcal{O}_K} : \langle s, s \rangle_v \leq 1 \text{ for all } v \in X_\infty\}$$

are called *global sections of E* .

Standard constructions from linear algebra make sense for Arakelov vector bundles as well. For example, one has a direct sum $E \oplus E'$ and a tensor product $E \otimes E'$ of Arakelov vector bundles E and E' over X , as well as exterior powers $\wedge^n E$ and in particular the determinant line bundle $\det(E) := \wedge^{\text{rank}(E)} E$.

A morphism of Arakelov vector bundles $\phi : E \rightarrow E'$ is an \mathcal{O}_K -linear map $\phi_{\mathcal{O}_K} : E_{\mathcal{O}_K} \rightarrow E'_{\mathcal{O}_K}$ such that for each infinite place $v \in X_\infty$, the induced map $\phi_{K_v} : E_{K_v} \rightarrow$

E'_{K_v} satisfies $\langle \phi_{K_v}(s), \phi_{K_v}(s) \rangle_{E',v} \leq \langle s, s \rangle_{E,v}$ for all $s \in E_{K_v}$. The resulting category of Arakelov vector bundles on X is *not* additive.

Definition 3. Given an Arakelov line bundle L over $X = \text{Spec}(\mathcal{O}_K) \cup X_\infty$, choose $0 \neq s \in L_{\mathcal{O}_K}$. The degree of L is

$$\deg(L) := \log(\#L_{\mathcal{O}_K}/\mathcal{O}_K \cdot s) - \sum_{v \in X_\infty} \epsilon_v \log(\langle s, s \rangle_v^{1/2}) \in \mathbb{R}$$

where $\epsilon_v := 1$ for $K_v \cong \mathbb{R}$ and $\epsilon_v := 2$ for $K_v \cong \mathbb{C}$. Note that this degree does not depend on the choice of s according to the product formula (2).

Given an Arakelov vector bundle E over $X = \text{Spec}(\mathcal{O}_K) \cup X_\infty$, the degree of E is

$$\deg(E) := \deg(\det E) \in \mathbb{R}.$$

If the rank of E is nonzero, then the slope of E is

$$\mu(E) := \deg(E)/\text{rank}(E).$$

3. Riemann-Roch

Roughly speaking, the Riemann-Roch problem asks for information on $\Gamma(X, E)$ in terms of the numerical invariants $\text{rank}(E)$ and $\deg(E)$. The purpose of this section is to discuss the following slightly more specific question:

Assume $\Gamma(X, E) = (0)$. How large can the slope $\mu(E)$ then be?

We can consider this question in different situations:

- X is an algebraic curve of genus g .

Here $\mu(E) \leq g - 1$ by Riemann-Roch, and an easy moduli count shows that for any given rank, there are in fact vector bundles E of slope $g - 1$ with $\Gamma(X, E) = (0)$. Thus the largest possible value for $\mu(E)$ is precisely $g - 1$.

- $X = \text{Spec}(\mathbb{Z}) \cup \{\infty\}$, the 'easiest' arithmetic curve (coming from $K = \mathbb{Q}$).

Here an Arakelov vector bundle E over X is the same thing as a euclidean vector space $E_{\mathbb{R}}$ together with a lattice $E_{\mathbb{Z}} \subset E_{\mathbb{R}}$, and its degree is

$$\deg(E) = -\log \text{vol}(E_{\mathbb{R}}/E_{\mathbb{Z}}).$$

The assumption $\Gamma(X, E) = (0)$ means that the unit ball in $E_{\mathbb{R}}$ contains no nonzero lattice point, or equivalently that $2E_{\mathbb{Z}} \subset E_{\mathbb{R}}$ is a lattice sphere packing (in the sense that the closed unit balls centered at points in $2E_{\mathbb{Z}}$ are disjoint). Hence finding the largest possible value for $\mu(E)$ is equivalent to finding the largest possible density of lattice sphere packings – still an open problem today.

A good reference for such sphere packings is [3]. Here we just mention results of Minkowski which imply the following:

The largest possible value of $\mu(E)$ is $\frac{1}{2} \log \text{rank}(E) + O(1)$.

More precisely, the famous Minkowski-Hlawka existence theorem for lattice sphere packings provides a constant C_1 such that for any given rank, there are Arakelov vector bundles E of slope $\mu(E) = \frac{1}{2} \log \text{rank}(E) + C_1$ with $\Gamma(X, E) = (0)$. On the other hand, Minkowski's theorem on lattice points in convex sets provides a constant C_2 such that $\Gamma(X, E) \neq (0)$ for any Arakelov vector bundle E of slope $\mu(E) = \frac{1}{2} \log \text{rank}(E) + C_2$.

(For $\text{rank}(E) \rightarrow \infty$, according to [3] nobody has managed to improve the constant C_1 coming from the Minkowski-Hlawka theorem, whereas several people have improved the other constant C_2 .)

- $X = \text{Spec}(\mathcal{O}_K) \cup X_\infty$, an arbitrary arithmetic curve.

Here adelic versions of Minkowski's arguments show the following:

The largest possible value of $\mu(E)$ is $\frac{d}{2} \log \text{rank}(E) + \frac{1}{2} \log \Delta_{K/\mathbb{Q}} + O(d)$, where $d = [K : \mathbb{Q}]$ denotes the degree and $\Delta_{K/\mathbb{Q}}$ denotes the discriminant of K over \mathbb{Q} (cf. [7, Section 3.4] for details). The existence of Arakelov vector bundles with such a slope and no nonzero global sections is also a special case of Theorem 8 to be explained below.

4. The Narasimhan-Seshadri theorem

Let X be an algebraic or arithmetic curve.

Definition 4. *A(n Arakelov) vector bundle E over X is stable (resp. semistable) if $\mu(E') < \mu(E)$ (resp. \leq) holds for all nontrivial subbundles $0 \neq E' \subsetneq E$.*

Theorem 5 (Narasimhan-Seshadri 1965). *Let E be a holomorphic vector bundle over a compact Riemann surface X . If E is stable of degree zero, then E comes from a (unique) irreducible unitary representation $\rho : \pi_1(X) \rightarrow \text{U}(n)$.*

In particular, every stable holomorphic vector bundle of degree zero has a canonical (flat) metric. This formulation admits a generalization to nonzero degree, and also to higher-dimensional X (Kobayashi-Hitchin correspondence).

Conversely, every holomorphic vector bundle E over the compact Riemann surface X coming from a unitary representation $\rho : \pi_1(X) \rightarrow \text{U}(n)$ is semistable of degree zero. If ρ is moreover irreducible, then E is stable.

Corollary 6. *If E and F are semistable holomorphic vector bundles over a compact Riemann surface X , then $E \otimes F$ is again semistable.*

The corresponding statement for an algebraic curve X over a field of characteristic $p > 0$ is known to be false. In 1997, J.-B. Bost asked whether the corresponding statement holds for an arithmetic curve X . Today (in October 2008), this question is still open, but there are recent partial results:

- Over $X = \text{Spec}(\mathbb{Z}) \cup \{\infty\}$, de Shalit and Parzanchevski [10] proved that $E \otimes F$ is semistable if E and F are semistable with $\text{rank}(E) + \text{rank}(F) \leq 5$.
- In the general case $X = \text{Spec}(\mathcal{O}_K) \cup X_\infty$, Bost and Künnemann [1] proved an upper bound for the slope of subbundles of $E \otimes F$, for semistable E and F . This bound has subsequently been improved by Chen [2].

5. Faltings' criterion for semistability

Let E and F be vector bundles over an algebraic curve of genus g . If $E \otimes F$ satisfies $\Gamma(X, E \otimes F) = (0)$ and has the largest possible slope $\mu(E \otimes F) = g - 1$, then it is easy to see that E and F are automatically semistable. Indeed, for every nontrivial subbundle $0 \neq E' \subsetneq E$ we have $\Gamma(X, E' \otimes F) = (0)$ and hence $\mu(E' \otimes F) \leq g - 1$, which implies $\mu(E') \leq \mu(E)$. Here is a converse:

Theorem 7 (Faltings 1993). *Let E be a semistable vector bundles over an algebraic curve X of genus g . Then there exists another vector bundle F over X with $\Gamma(X, E \otimes F) = (0)$ and $\mu(E \otimes F) = g - 1$.*

As long as we don't even know the maximal slope of Arakelov vector bundles without global sections, we cannot expect an equally precise analogue of this for arithmetic curves X . However, there is some weaker statement in that direction:

Theorem 8. *Let E be a semistable Arakelov vector bundle over the arithmetic curve $X = \text{Spec}(\mathcal{O}_K) \cup X_\infty$. Then there exists another Arakelov vector bundle F over X with $\Gamma(X, E \otimes F) = (0)$ and*

$$\mu(E \otimes F) = \frac{d}{2} \log \text{rank}(F) + \frac{1}{2} \log \Delta_{K/\mathbb{Q}} - \frac{1 + \log 2\pi}{2} d.$$

This theorem is proved in [7] and in [8]. In the special case where E is trivial of rank one, it states the existence of Arakelov vector bundles with a certain slope and no nonzero global sections, as mentioned above. In particular, it contains as a

special case some form of the Minkowski-Hlawka theorem. The strategy of the proof is the same as in this special case; we give an outline below.

Sketch of the proof. Fix $n \gg 0$ and an Arakelov line bundle L over X with

$$(3) \quad \mu(E) + \deg(L) = \frac{d}{2} \log(n) + \frac{1}{2} \log \Delta_{K/\mathbb{Q}} - \frac{1 + \log 2\pi}{2} d.$$

We consider all Arakelov vector bundles F over X of rank n and determinant L . These can be parameterized by the coset space $\mathrm{SL}_n(\mathbb{A})/\mathrm{SL}_n(K)$, where \mathbb{A} denotes the ring of adeles. The idea is to show that the ‘average’ number of sections

$$(4) \quad \int_{\mathrm{SL}_n(\mathbb{A})/\mathrm{SL}_n(K)} \#\Gamma(X, E \otimes F) d\tau(F)$$

is less than two. Here τ is the Tamagawa measure on $\mathrm{SL}_n(\mathbb{A})/\mathrm{SL}_n(K)$, which is the unique $\mathrm{SL}_n(\mathbb{A})$ -invariant probability measure on this homogeneous space.

The main tool for computing the integral (4) is an adelic version of Siegel’s mean value formula. It states, for every $l \leq n$ and every nonnegative measurable function f on the space $\mathrm{Mat}_{n \times l}(\mathbb{A})$ of $n \times l$ adèle matrices, that

$$(5) \quad \int_{\mathrm{SL}_n(\mathbb{A})/\mathrm{SL}_n(K)} \sum_{\substack{M \in \mathrm{Mat}_{n \times l}(K) \\ \mathrm{rank}(M)=l}} f(g \cdot M) d\tau(g) = \int_{\mathrm{Mat}_{n \times l}(\mathbb{A})} f d\tau$$

where τ again denotes Tamagawa measures. The formula (5) follows from the theory of these measures, as presented in Weil’s book [11].

Let $B_{\mathbb{R}}^N \subset \mathbb{R}^N$ be the unit ball for the standard scalar product. Its volume is

$$(6) \quad V(N) := \mathrm{vol}(B_{\mathbb{R}}^N \subset \mathbb{R}^N) = \frac{\pi^{N/2}}{(N/2)!}.$$

Similarly, let $B_{\mathbb{C}}^N \subset \mathbb{C}^N$ be the unit ball for the standard (hermitian) scalar product. For finite places $v \in X$, put $B_{K_v}^N := \mathcal{O}_v^N \subset K_v^N$. Let the adelic unit ball $B_{\mathbb{A}}^N \subset \mathbb{A}^N$ be the product, over all places $v \in X$, of these local unit balls $B_{K_v}^N \subset K_v^N$. Due to the standard normalizations, its Tamagawa measure is

$$(7) \quad \tau(B_{\mathbb{A}}^N) = \Delta_{K/\mathbb{Q}}^{-N/2} \cdot V(N)^{r_1} \cdot (2^N V(2N))^{r_2}$$

where r_1 and r_2 are the number of real and complex places in X_{∞} .

For every global section $s \in \Gamma(X, E \otimes F)$, there is a unique minimal subbundle $E' \subseteq E$ with $s \in \Gamma(X, E' \otimes F)$. Then $l := \mathrm{rank}(E') \leq \mathrm{rank}(F) = n$, and if we choose bases of the K -vector spaces E_{η} and F_{η} , then $s_{\eta} \in E_{\eta} \otimes F_{\eta}$ is given by an $n \times l$ matrix of full rank l . For fixed l and E' , the average number of such sections s can thus be

computed using the formula (5), taking the function f to be 1 on $B_{\mathbb{A}}^{nl} \subset \mathbb{A}^{nl}$ and 0 elsewhere. The result is

$$\exp(l \deg(L) + n \deg(E')) \cdot \tau(B_{\mathbb{A}}^{nl})$$

where the factor $\exp(l \deg(L) + n \deg(E'))$ comes from the fact that we work with Arakelov vector bundles $E' \otimes F$ whose determinant is fixed, but not trivial; $l \deg(L) + n \deg(E')$ is just the common degree of all these bundles $E' \otimes F$.

Summing over all l and all E' , we get that the integral (4) equals

$$(8) \quad 1 + \sum_{l=1}^{\text{rank}(E)} \sum_{\substack{E' \subseteq E \\ \text{rank}(E')=l}} \exp(l \deg(L) + n \deg(E')) \cdot \tau(B_{\mathbb{A}}^{nl})$$

where the summand 1 comes from $l = 0$, i. e. from the trivial section $s = 0$.

For fixed l and $n \gg 0$, the inner sum is dominated by the summands with $\deg(E')$ maximal; we can estimate them using $\mu(E') \leq \mu(E)$, since we have assumed that E is semistable. More precisely, one can show that there is a constant C_E such that for all $n \gg 0$ and all l , we have

$$\sum_{\substack{E' \subseteq E \\ \text{rank}(E')=l}} \exp(l \deg(L) + n \deg(E')) \leq C_E \cdot \exp(nl(\mu(E) + \deg(L))).$$

Using this estimate, the assumption (3) on the common slope $\mu(E) + \deg(L)$ of the bundles $E \otimes F$, the formula (7) for the Tamagawa measure $\tau(B_{\mathbb{A}}^{nl})$, and Stirling's formula for the factorials in (6), the average number (8) of global sections of $\Gamma(X, E \otimes F)$ turns out to be less than $1 + 1$ for $n \gg 0$. Hence there is an Arakelov vector bundle F of rank n and determinant L with $\Gamma(X, E \otimes F) = (0)$. \square

References

- [1] J.-B. BOST & K. KÜNNEMANN – “Hermitian vector bundles and extension groups on arithmetic schemes. I. Geometry of numbers.”, Preprint math/0701343, available at <http://www.arXiv.org>.
- [2] H. CHEN – “Maximal slope of tensor product of Hermitian vector bundles”, Preprint arXiv:0706.0690, available at <http://www.arXiv.org>.
- [3] J. H. CONWAY & N. J. A. SLOANE – *Sphere packings, lattices and groups*, Springer-Verlag, New York, 1988.
- [4] R. DEDEKIND & H. WEBER – “Theorie der algebraischen Functionen einer Veränderlichen.”, *J. Reine Angew. Math.* **92** (1882), p. 181–290 (German).
- [5] G. FALTINGS – “Stable G -bundles and projective connections”, *J. Algebraic Geom.* **2** (1993), no. 3, p. 507–568.

- [6] _____, “Moduli-stacks for bundles on semistable curves”, *Math. Ann.* **304** (1996), no. 3, p. 489–515.
- [7] N. HOFFMANN – *On vector bundles over algebraic and arithmetic curves*, Bonner Mathematische Schriften, 351, Ph. D. thesis, Bonn, 2002.
- [8] _____, “Stability of Arakelov bundles and tensor products without global sections”, *Doc. Math.* **8** (2003), p. 115–123 (electronic).
- [9] M. S. NARASIMHAN & C. S. SESHADRI – “Stable and unitary vector bundles on a compact Riemann surface”, *Ann. of Math. (2)* **82** (1965), p. 540–567.
- [10] E. DE SHALIT & O. PARZANCHEVSKI – “On tensor products of semistable lattices”, Preprint, 2006.
- [11] A. WEIL – *Adeles and algebraic groups*, Birkhäuser Boston, Mass., 1982.

A CONJECTURAL NON-COMMUTATIVE GENERALIZATION OF A VOLUME FORMULA OF MCMULLEN-SCHNEIDER

Tobias Finis

Erez Lapid

Institute of Mathematics, Hebrew University of Jerusalem
E-mail: erezla@math.huji.ac.il

Abstract. The Fourier transform of the characteristic function of a convex polytope is given by a localization formula due to Brion [2]. On the other hand, an argument by McMullen-Schneider expresses the mixed volume of d polytopes in a d -dimensional space in terms of volumes of parallelotopes formed by d -tuples of edges, one from each polytope [4]. When we specialize these formulas we get two seemingly different expressions for the volume of a polytope. We propose a non-commutative conjectural generalization of this identity. In the case of a root zonotope, such an identity is known and used in the study of the spectral side of Arthur's trace formula [3].

1. Preliminaries

Let V be a real vector space of dimension d and V^* its dual space. By a *cone* in V^* we will always mean a closed polyhedral cone σ with apex 0 such that $\sigma \cap -\sigma = \{0\}$. Let Σ be a *fan* in V^* , i.e., a collection of cones such that

1. if $\sigma \in \Sigma$ then any face of σ belongs to Σ ,
2. if $\sigma_1, \sigma_2 \in \Sigma$ then $\sigma_1 \cap \sigma_2$ is a face in both.

October 2008.

2000 Mathematics Subject Classification. Primary: 52A38, 52A39, 52B11, 52B45; Secondary: 19L47, 55N91.

Authors partially supported by a grant from the German-Israeli Foundation.
Talk given by second named author.

We will assume that Σ is *complete*, that is $\cup \Sigma = V^*$. The elements of Σ are called *faces*. We denote by $\Sigma(i)$ the set of i -dimensional faces of Σ . In particular, $\Sigma(d)$, $\Sigma(d-1)$ and $\Sigma(1)$ are the sets of *chambers*, *walls* and *rays* of Σ respectively. Two chambers are *adjacent* if they intersect in a wall. Any wall is contained in exactly two chambers (which are adjacent). We will write $\sigma \overset{\tau}{\leftarrow} \tilde{\sigma}$ if σ and $\tilde{\sigma}$ are adjacent with common wall $\tau = \sigma \cap \tilde{\sigma}$. If we want to distinguish σ we will write $\sigma \overset{\tau}{\rightarrow} \tilde{\sigma}$ and speak of a *directed wall* ω emerging from σ . We denote by $\tilde{\omega}$ the opposite directed wall $\tilde{\sigma} \overset{\tau}{\rightarrow} \sigma$. **Henceforth, we will assume that Σ is simplicial**, that is, each cone in Σ is simplicial. Equivalently, any chamber σ has precisely d directed walls emerging from it.

For any cone σ we denote by $\mathcal{V}(\sigma)$ its linear span and by σ^\perp its annihilator in V . A d -tuple (τ_1, \dots, τ_d) of walls is called *transversal* if $\sum_{i=1}^d \tau_i^\perp = V$, i.e. if $\cap_{i=1}^d \mathcal{V}(\tau_i) = 0$.

A basic example of a simplicial fan is the normal fan Σ_P of a simple convex polytope P in V , whose affine hull is V . It is given by $\Sigma_P = \{\tau(F) : F \in \mathcal{F}(P)\}$ where $\mathcal{F}(P)$ denotes the lattice of faces of P and

$$\tau(F) = \{\lambda \in V^* : \lambda|_P \text{ attains its maximum on } F\}.$$

Note that $\tau : \mathcal{F}(P) \rightarrow \Sigma_P$ is an inclusion reversing bijection and $\dim F + \dim \tau(F) = d$. (We recall that not every fan is the normal fan of a polytope.)

Given $\sigma \in \Sigma$ we say that $v \in V$ is positive with respect to σ if $\langle \lambda, v \rangle > 0$ for any $\lambda \in \text{relint } \sigma$. Given a directed wall $\omega : \sigma \overset{\tau}{\rightarrow} \tilde{\sigma}$, a *directed normal* for ω , or an ω -directed normal, is an element of τ^\perp which is positive with respect to σ . Such a vector is uniquely determined up to multiplication by a positive scalar. If v is a directed normal for ω then $-v$ is a directed normal for the opposite wall $\tilde{\omega}$

2. Piecewise polynomial functions

Let $S = \text{Sym}(V)$ be the ring of polynomial functions on V^* . For any face $\sigma \in \Sigma$ we denote by \mathcal{J}_σ the ideal of S generated by the subspace σ^\perp of V .

A *piecewise polynomial* with respect to Σ is a function on V^* whose restriction to any chamber (hence, to any face) is a polynomial. We denote by $\mathcal{A} = \mathcal{A}_\Sigma$ the graded algebra of piecewise polynomials with respect to Σ . It is known that \mathcal{A} is a free S -module, and is generated as an algebra by its degree 1 elements (the piecewise linear functionals). Moreover, the dimension of \mathcal{A}_1 is the number of rays in Σ .

We can view an element of \mathcal{A} as a collection X_σ of elements of S , one for each chamber σ , such that $X_{\sigma_1} - X_{\sigma_2} \in \mathcal{J}_{\sigma_1 \cap \sigma_2}$ for any two chambers σ_1, σ_2 . (It is enough

to check this condition for σ_1, σ_2 adjacent.) For any directed wall $\omega : \sigma \xrightarrow{\tau} \tilde{\sigma}$ we write $X_\omega = X_\sigma - X_{\tilde{\sigma}} \in \mathcal{J}_\tau$.

More generally if M is an S -module we define $M_\Sigma := M \otimes_S \mathcal{A}_\Sigma$ to be the \mathcal{A}_Σ -module of Σ -piecewise elements of M . If M is flat over S then an element of M_Σ can be described as a collection $m_\sigma \in M$, $\sigma \in \Sigma(d)$ such that $m_{\sigma_1} - m_{\sigma_2} \in \mathcal{J}_{\sigma_1 \cap \sigma_2} M$ for any chambers σ_1, σ_2 .

Suppose that P is a polytope in V whose normal fan is Σ . Then the vertices of P are indexed by the chambers of Σ and give rise to a piecewise linear form L_P on V^* with respect to Σ . We say that L_P is the piecewise linear form defined by P . These piecewise linear forms are characterized by the property that X_ω is positive with respect to σ for any directed wall $\omega : \sigma \xrightarrow{\tau} \tilde{\sigma}$.

Fix $0 \neq \beta = \beta_\Sigma \in (\wedge^d V)^*$.

Let σ be a chamber and let $\omega_i : \sigma \xrightarrow{\tau_i} \sigma_i$, $i = 1, \dots, d$ be the directed walls emerging from σ . Set

$$\theta_\sigma = \frac{v_1 \cdots v_d}{|\beta(v_1 \wedge \cdots \wedge v_d)|} \in S$$

where v_i is a directed normal of ω_i . As the notation suggests, θ_σ depends only on σ and not on the choice of the v_i 's or the order of the ω_i 's.

It is well-known (e.g. [2]) that we have an S -linear map $\delta_\Sigma : \mathcal{A}_\Sigma \rightarrow S$ defined by

$$(X_\sigma)_{\sigma \in \Sigma(d)} \mapsto \sum_{\sigma \in \Sigma(d)} \frac{X_\sigma}{\theta_\sigma}.$$

Extending scalars, we get for any S -module M an S -linear map

$$\delta = \delta_{\Sigma; M} : M_\Sigma \rightarrow M.$$

3. The setup

Let $\mathbb{C}[[V]]$ denote the algebra of formal power series in V , i.e. the completion of S at the origin, which is also the dual space of the vector space $\mathbb{C}[V]$ of polynomials on V . We denote by $[\cdot, \cdot] : \mathbb{C}[V] \times \mathbb{C}[[V]] \rightarrow \mathbb{C}$ the pairing and by $A \mapsto A(0) = [1, A]$ the evaluation map $\mathbb{C}[[V]] \rightarrow \mathbb{C}$.

For any vector space U we set $U[[V]] = \mathbb{C}[[V]] \otimes U$. We write $[\cdot, \cdot] : \mathbb{C}[V] \times U[[V]] \rightarrow U$ for the bilinear pairing. Given a bilinear map $\circ : U_1 \times U_2 \rightarrow U_3$ we will continue to denote by \circ the $\mathbb{C}[[V]]$ -bilinear map $\circ : U_1[[V]] \times U_2[[V]] \rightarrow U_3[[V]]$ obtained by extending the scalars.

Henceforth, whenever V' is a subspace of V we will identify $\mathbb{C}[[V']]$ with a subalgebra of $\mathbb{C}[[V]]$.

Definition 3.1. An intertwining family with respect to Σ consists of the following data

1. For each chamber σ , a finite-dimensional vector space W_σ ,
2. For any pair of chambers $\sigma_1, \sigma_2 \in \Sigma(d)$, an element $A_{\sigma_2|\sigma_1} \in \text{Hom}(W_{\sigma_1}, W_{\sigma_2})[[\sigma_1 \cap \sigma_2]^\perp]$,

with the following properties

1. $A_{\sigma|\sigma} = \text{id}_{W_\sigma} \otimes 1$ for all chambers σ ,
2. For any triple of chambers $\sigma_1, \sigma_2, \sigma_3$ we have

$$A_{\sigma_3|\sigma_1} = A_{\sigma_3|\sigma_2} \circ A_{\sigma_2|\sigma_1}$$

(an equality in $\text{Hom}(W_{\sigma_1}, W_{\sigma_3})[[V]]$) where \circ denotes composition.

Note that the data is determined by $A_{\sigma_2|\sigma_1}$ where σ_1, σ_2 are adjacent. Given a directed wall $\omega : \sigma \xrightarrow{\tau} \bar{\sigma}$ we write $A_\omega = A_{\bar{\sigma}|\sigma} \in \text{End}(W_\sigma, W_{\text{tilde{id}\sigma}})[[\tau^\perp]]$.

We fix a chamber σ_0 and consider for any chamber σ the element $m_\sigma := A_{\sigma_0|\sigma}(0)A_{\sigma|\sigma_0}$ of $M = \text{End}(W_{\sigma_0})[[V]]$. It is easy to see that this defines a Σ -piecewise element of M . Let $\mathcal{D}_{\sigma_0} A := \delta_{\Sigma; M}((m_\sigma)_{\sigma \in \Sigma(d)}) \in M$. Observe that $\mathcal{D}_{\bar{\sigma}_0} A = A_{\bar{\sigma}_0|\sigma_0}(0) \circ \mathcal{D}_{\sigma_0} A \circ A_{\sigma_0|\bar{\sigma}_0}$ for any other chamber $\bar{\sigma}_0$ in Σ .

Let $\omega : \sigma \xrightarrow{\tau} \bar{\sigma}$ be a directed wall. For any $f \in \mathbb{C}[\tau^\perp]$ we set

$$f_\omega^{\sigma_0}(A) = A_{\sigma_0|\sigma}(0) \circ A_\omega(0)^{-1} \circ [f, A_\omega] \circ A_{\sigma|\sigma_0}(0) \in \text{End}(W_{\sigma_0}).$$

Note that if $\nu^* \in (\tau^\perp)^* \subseteq \mathbb{C}[\tau^\perp]$ then

$$(\nu^*)_\omega^{\sigma_0}(A) = (-\nu^*)_\omega^{\sigma_0}(A).$$

Also,

$$f_\omega^{\bar{\sigma}_0}(A) = A_{\bar{\sigma}_0|\sigma_0}(0) \circ f_\omega^{\sigma_0}(A) \circ A_{\sigma_0|\bar{\sigma}_0}(0)$$

for any chamber $\bar{\sigma}_0$.

Let τ_1, \dots, τ_d be walls. For each $i = 1, \dots, d$ choose a directed wall $\omega_i : \sigma_i \xrightarrow{\tau_i} \bar{\sigma}_i$ and a directed normal ν_i for ω_i . Let $\nu_i^* \in (\tau_i^\perp)^* \subseteq \mathbb{C}[\tau_i^\perp]$ be such that $\langle \nu_i^*, \nu_i \rangle = 1$. We set

$$\partial_{\tau_1, \dots, \tau_d}^{\sigma_0} A = |\beta(\nu_1 \wedge \dots \wedge \nu_d)| (\nu_1^*)_{\omega_1}^{\sigma_0}(A) \circ \dots \circ (\nu_d^*)_{\omega_d}^{\sigma_0}(A).$$

Note that this expression depends only on (τ_1, \dots, τ_d) and not on the choice of the ω_i 's or the ν_i 's. Also note that

$$\partial_{\tau_1, \dots, \tau_d}^{\bar{\sigma}_0} A = A_{\bar{\sigma}_0|\sigma_0}(0) \circ \partial_{\tau_1, \dots, \tau_d}^{\sigma_0} A \circ A_{\sigma_0|\bar{\sigma}_0}(0)$$

for any chamber $\bar{\sigma}_0$.

4. The conjectural formula

Let A be an intertwining family with respect to Σ .

Conjecture 4.1. *For any choice of $\vec{\lambda} = (\lambda_1, \dots, \lambda_d) \in (V^*)^d$ in general position with respect to Σ we have*

$$[\mathcal{D}^{\sigma_0} A](0) = \frac{(-1)^d}{d!} \sum_{(\tau_1, \dots, \tau_d) \in \mathcal{X}_{\vec{\lambda}}} \partial_{\tau_1, \dots, \tau_d}^{\sigma_0} A$$

where $\mathcal{X}_{\vec{\lambda}}$ is the set of d -tuples (τ_1, \dots, τ_d) of transversal walls such that the translates $\lambda_i + \tau_i$, $i = 1, \dots, d$ intersect (necessarily in a point).

The conjecture is trivially true for $d = 1$. It can be also proved for $d = 2$ by direct computation. A special case of the conjecture for Coxeter fans (corresponding to root hyperplane arrangements) was established in [3]. It played a role in the analysis of the spectral side of Arthur’s trace formula.

5. Remarks

Suppose that P is a polytope in V and Σ is its normal fan in V^* . Let L_σ , $\sigma \in \Sigma(d)$ denote the vertices of P . We can form the intertwining family with $W_\sigma = \mathbb{C}$ for all chambers and $A_{\sigma_2|\sigma_1} = e^{L_{\sigma_2} - L_{\sigma_1}}$. Then $\mathcal{D}^{\sigma_0} A$ is the Fourier transform of the translate of P by σ_0 ([2]). Thus, conjecture 4.1 reduces in this case to an argument by McMullen-Schneider ([5], p. 200-1) expressing $\text{vol}(P)$ as $1/d!$ times the sum of the volume of the parallelotope formed by the vectors $\vec{e}_1, \dots, \vec{e}_d$ as (e_1, \dots, e_d) range over the d -tuples of edges of P for which there exists $\mu \in V^*$ such that $\max(\mu + \lambda_i)|_P$ is attained on e_i , $i = 1, \dots, d$. (Here, $\vec{e}_i \in V$ denotes the vector corresponding to e_i .) In fact, the argument in [loc. cit.] applies more generally for mixed volumes.

Next, we comment about the dependence on $\vec{\lambda}$. Given a fan Σ in U and a linear surjective map $p: U \rightarrow U'$ the *quotient fan* on U' is defined as the common refinement of $p(\sigma)$, $\sigma \in \Sigma$ (cf. [4], [1]). In the case where $U = V^*$ and Σ is the normal fan of a polytope P in V , the quotient fan is the normal fan of the *fiber polytope* of P , in the sense of Billera-Sturmfels, with respect to the projection $V \rightarrow V/(\text{Ker } p)^\perp = (\text{Ker } p)^*$ ([4, Proposition 2.3]).

In particular, consider V^* embedded diagonally in $(V^*)^d$ and let $p: (V^*)^d \rightarrow (V^*)^d/V^*$ be the canonical projection. Let $\Sigma^d = \underbrace{\Sigma \times \dots \times \Sigma}_{d \text{ times}}$ (a fan in $(V^*)^d$), and let $\bar{\Sigma}$ be the quotient fan in $(V^*)^d/V^*$. The precise condition on $\vec{\lambda}$ to be in general

position with respect to Σ is that it lies outside the walls of $\overline{\Sigma}$, i.e. it lies in the interior of a chamber of $\overline{\Sigma}$. Moreover, the set $\mathcal{X}_{\vec{\lambda}}$ depends only on the chamber to which $\vec{\lambda}$ belongs.

References

- [1] L. J. BILLERA & B. STURMFELS – “Iterated fiber polytopes”, *Mathematika* **41** (1994), no. 2, p. 348–363.
- [2] M. BRION – “The structure of the polytope algebra”, *Tohoku Math. J. (2)* **49** (1997), no. 1, p. 1–32.
- [3] T. FINIS, E. LAPID & W. MÜLLER – “On the spectral side of Arthur’s trace formula II”, preprint.
- [4] M. M. KAPRANOV, B. STURMFELS & A. V. ZELEVINSKY – “Quotients of toric varieties”, *Math. Ann.* **290** (1991), no. 4, p. 643–655.
- [5] P. MCMULLEN & R. SCHNEIDER – “Valuations on convex bodies”, *Convexity and its applications*, Birkhäuser, Basel, 1983, p. 170–247.

THE j -INVARIANT OF A PLANE TROPICAL CUBIC

Hannah Markwig

Mathematisches Institut, Georg-August-Universität Göttingen
E-mail: hannah@uni-math.gwdg.de

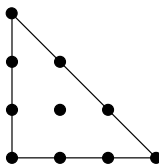
Abstract. In tropical geometry, algebraic varieties are replaced by certain degenerations called tropical varieties. Tropical varieties are piece-wise linear objects that can be studied using combinatorics and linear algebra methods. We can use tropical geometry to prove theorems about algebraic geometry by means of these new methods. To reach this aim, we have to understand the connection between algebraic geometry and tropical geometry. In this talk, we want to understand the tropical analogue of the j -invariant of a smooth elliptic curve.

1. The algebraic setting

A smooth elliptic curve can be embedded into \mathbb{P}^2 as a cubic:

$$f = a_0 + a_1x + a_2y + a_3x^2 + a_4xy + a_5y^2 + a_6x^3 + a_7x^2y + a_8xy^2 + a_9y^3.$$

Define the Newtonpolygon of a polynomial to be the convex hull of exponent vectors. For our polynomial f above this is:



The j -invariant is an invariant of the complex structure of the elliptic curve. Two elliptic curves are isomorphic if and only if their j -invariants coincide. The j -invariant can be computed as a rational function in the coefficients of the polynomial f :

$$j(f) = \frac{A(a_i)}{\Delta(a_i)}, \quad A, \Delta \in \mathbb{Q}[a_0, \dots, a_9].$$

Note that the two polynomials A and Δ are very big, they have between a thousand and two thousand terms.

2. The tropical setting

First we want to show in what sense tropical curves arise as degenerations of algebraic curves. Let K be the field of Puiseux-series $K = \{\alpha(t) = a_1 t^{q_1} + a_2 t^{q_2} + \dots\}$ where $a_i \in \mathbb{C}$ and $q_1 < q_2 < \dots \in \mathbb{Q}$ share a common denominator. There is a valuation: $\text{val} : K^* \rightarrow \mathbb{R} : \alpha(t) \mapsto q_1$. For $V \subset (K^*)^2$ define

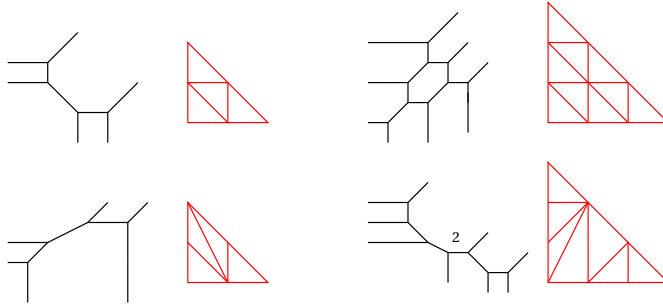
$$\text{Val}(V) := \overline{\{(-\text{val } x, -\text{val } y), (x, y) \in V\}} \subset \mathbb{R}^2,$$

the Tropicalization of V . Next we shall see why tropical curves are piece-wise linear. For $f = \sum a_{ij} x^i y^j \in K[x, y]$ and $w \in \mathbb{R}^2$ define $\text{in}_w(f)$ to be the terms $a_{ij} x^i y^j$ of f for which $-\text{val}(a_{ij}) + i w_1 + j w_2$ is maximal. Define

$$\begin{aligned} \text{Trop}(f) &:= \{w \in \mathbb{R}^2, \text{in}_w(f) \text{ is not a monomial}\} \\ &= \{w \in \mathbb{R}^2, \max\{-\text{val}(a_{ij}) + i w_1 + j w_2\} \\ &\quad \text{is attained at least twice}\}. \end{aligned}$$

Theorem 1 (Kapranov). $\text{Val}(V(f)) = \text{Trop}(f)$.

Tropical curves are combinatorial objects, because they are dual to a subdivision of the Newton polygon, in the sense that vertices of the tropical curve correspond to polygons in the subdivision and edges correspond to orthogonal edges. The following picture shows some tropical curves and their dual Newton subdivision.



An elliptic tropical curve has a cycle if and only if the interior lattice point of the Newton polygon is a vertex in the dual Newton subdivision.

3. The result

We proved the following theorem ([2]):

Theorem 2. *Given an elliptic cubic curve $C = V(f)$ over K satisfying some genericity condition and such that $\text{Trop}(f)$ has a cycle, then*

$$-\text{val}(j(C)) = \text{cycle length of } \text{Trop}(f).$$

It is an immediate corollary that if $\text{val}(j(C)) > 0$ the tropical curve cannot have a cycle. This is similar to bad reduction of semistable elliptic curves over discrete valuation rings. There is work in progress to study the connection between tropical curves and semistable elliptic curves over discrete valuation rings.

There are some older results that gave evidence already that the cycle length is an analogue of the j -invariant. In [4], it is shown that the isomorphism class of a tropical elliptic curve is determined by its cycle length. In [3], numbers of elliptic curves with fixed j -invariant and satisfying incidence conditions are proved to be equal to numbers of tropical curves with fixed cycle length and satisfying incidence conditions.

Here are some ideas of the proof of Theorem 2. Let us consider $-\text{val}(j)$ first. Since j can be expressed as the rational function $\frac{A}{\Delta}$, we have $-\text{val}(j) = \text{val}(\Delta) - \text{val}(A)$. Assume $\Delta = \sum_u b_u a_0^{u_0} \dots a_9^{u_9}$, $b_u \in \mathbb{Q}$ and $a_0 = c_0 t^{w_0} + \dots$, \dots , $a_9 = c_9 t^{w_9} + \dots$. Assume further $\text{in}_{-w}(\Delta)$ is a monomial, say $\text{in}_{-w}(\Delta) = b_v a_0^{v_0} \dots a_9^{v_9}$. Then $\text{val}(\Delta(a_0, \dots, a_9)) =$

$w_0 v_0 + \dots + w_9 v_9$, since

$$\begin{aligned}\Delta(a_0, \dots, a_9) &= b_v (c_0 t^{w_0} + \dots)^{v_0} \cdot \dots \cdot (c_9 t^{w_9} + \dots)^{v_9} + \dots \\ &= b_v c_0^{v_0} \dots c_9^{v_9} t^{w_0 v_0 + \dots + w_9 v_9} + \dots,\end{aligned}$$

where the dots represent higher order terms in t . Recall that

$$\begin{aligned}\text{in}_{-w}(\Delta) &= \text{terms where } -\text{val}(b_u) - w_0 u_0 - \dots - w_9 u_9 \text{ is maximal} \\ &= \text{terms where } w_0 u_0 + \dots + w_9 u_9 \text{ is minimal.}\end{aligned}$$

Thus knowing the initial form $\text{in}_{-w}(\Delta)$ (and also $\text{in}_{-w}(A)$) helps us to determine $-\text{val}(\Delta)$, $-\text{val}(A)$ and hence $-\text{val}(j)$. In particular, we can see that $-\text{val}(j)$ is a piecewise linear function in the w_i , under the genericity assumption that the initial forms are monomials. To go on, we can compute the Newton polygon of Δ and A using POLYMAKE ([1]). The dual fan of this polytope is the Gröbner fan, which groups all w to one cone for which $\text{in}_{-w}(\Delta)$ coincides. The cones satisfy inequalities in the w_i , and those inequalities give us information about the dual Newton subdivision of the tropical curve $\text{Trop}(f(a_0, \dots, a_9))$, which is the tropical curve corresponding to $\max\{w_0, w_1 x, \dots, w_9 + 3y\}$. We can observe that there is only one cone of the Gröbner fan of A such that the corresponding tropical curves have a cycle. So if we assume that the tropical curves have a cycle, then we know already that w has to be in this cone, and we thus know $\text{in}_{-w}(A)$ and hence $\text{val}(A(a))$. The Gröbner fan of Δ on the other hand has 79 cones that correspond to tropical curves which have a cycle. Again, in each cone, the inequalities imposed on the w_i are enough to determine important parts of the dual Newton subdivision, which suffice to express the cycle length of those tropical curves as a linear function in the w_i . To show the desired equality, it is hence enough to compare the two linear functions $-\text{val}(j)$ and cycle length for each of those 79 cones. That can be done with a computer, but also, we can use symmetries in the dual Newton subdivisions and thus show it only for a smaller number of cases by hand.

References

- [1] E. GAWRIŁOW & M. JOSWIG – “POLYMAKE 2.3”, Tech. report, TU Berlin and TU Darmstadt, <http://www.math.tu-berlin.de/polymake/>.
- [2] E. KATZ, H. MARKWIG & T. MARKWIG – “The j -invariant of a plane tropical cubic”, *Journal of Algebra* **320** (2008), p. 3832–3848, arXiv:0709.3785.
- [3] M. KERBER & H. MARKWIG – “Counting tropical elliptic plane curves with fixed j -invariant”, *Commentarii Mathematici Helvetici* (to appear). arXiv:math.AG/0608472, 2006.
- [4] G. MIKHALKIN – “Tropical geometry and its applications”, *Invited lectures v. II, Proceedings of the ICM Madrid* (M. S.-S. et al., ed.), 2006, arXiv:math.AG/0601041, p. 827–852.

LEOPOLDT'S CONJECTURE FOR SOME GALOIS EXTENSIONS

Preda Mihailescu

Mathematisches Institut der Universität Göttingen
E-mail: `preda@uni-math.gwdg.de`

Abstract. In this paper we use Baker theory for proving some special cases of Leopoldt's conjecture. Thus we show that if the conjecture is true for some field \mathbb{K} , then it is true for solvable extensions thereof. Also, if it holds for arbitrary extensions with simple groups, then it holds for any number field. The proofs are essentially based on the talk given at the SANT Conference.

*He will always carry on
Some things are lost, some things are found,
They will keep on speaking his name
Some things are changed, some still the same.⁽¹⁾*

To Alan Baker

March 2009.

2000 Mathematics Subject Classification. 11R23 Iwasawa Theory, 11R27 Units.

**The readers are invited to check for updates on the arXiv version
<http://arxiv.org/abs/0905.1274> and/or contact the author by mail.**

⁽¹⁾From a Hymn of *Pretenders*

1. Introduction

Let \mathbb{K}/\mathbb{Q} be a finite galois extension and p be a rational prime. It was conjectured by Leopoldt in [4] that the p -adic regulator of \mathbb{K} does not vanish. Some equivalent statements are explained below. The conjecture was proved for abelian extensions in 1967 by Brumer [2], using a local version of Baker's linear forms in logarithms: the result is known as the Baker-Brumer theorem. A theorem proved by Ax in [1] allows to relate the Leopoldt conjecture for abelian extensions to transcendence theory. In his paper, Ax mentions that he could expect his method to work also for non-abelian extensions. This was attempted by Emsalem and Kissilewski, who obtained in [3] results for some particular, non abelian extensions.

The main result of this paper is

Theorem 1. *Let \mathbb{L}/\mathbb{K} be a finite solvable extension of number fields and p a rational prime. If Leopoldt's conjecture holds for \mathbb{K} then it holds for \mathbb{L} .*

This implies in particular the Leopoldt conjecture for absolute solvable extensions. We state from [2] the central theorem on p -adic forms in logarithms, which we shall use here:

Theorem 2 (Baker and Brumer). *Let $\overline{\mathbb{Q}}_p$ be an algebraic closure of \mathbb{Q}_p and $\mathbb{U} \subset \overline{\mathbb{Q}}_p$ be the units. Let $\alpha_1, \alpha_2, \dots, \alpha_n$ be elements of \mathbb{U} which are algebraic over \mathbb{Q} and whose p -adic logarithms exist and are independent over \mathbb{Q} . These logarithms are then independent over \mathbb{Q}' , the algebraic closure of \mathbb{Q} in $\overline{\mathbb{Q}}_p$.*

2. Baker theory and Leopoldt's conjecture

Let \mathbb{K}/\mathbb{Q} be an arbitrary galois field with group G , let p be a rational prime and $P = \{\wp \subset \mathcal{O}(\mathbb{K}) : (p) \subset \wp\}$ be the set of conjugate prime ideals above p in \mathbb{K} .

We shall prove in this section two important consequences of the Theorem 2, one for absolute and one for relative galois extensions.

The algebra $\mathbb{K}_p = \mathbb{K} \otimes_{\mathbb{Q}} \mathbb{Q}_p$ is the product of all completions of \mathbb{K} at the places in the set P :

$$\mathbb{K}_p = \prod_{\wp \in P} \mathbb{K}_{\wp}.$$

The global field \mathbb{K} is dense in \mathbb{K}_{\wp} in the product topology and G acts on this completion faithfully, so for any $x \in \mathbb{K}_p$, $x = \lim_n x_n$, $x_n \in \mathbb{K}$ and for all $g \in G$ we have

$g(x) = \lim_n g(x_n)$. The units $U \subset \mathbb{K}_p$ are products of the units in $U_\wp \subset \mathbb{K}_\wp$ and E embeds diagonally to $\bar{E} \subset U$. The conjecture of Leopoldt says that

$$\mathbb{Z}_p - \text{rank}(\bar{E}) = \mathbb{Z} - \text{rank}(E).$$

Let $\delta \in E$ be a Minkowski unit with $\delta \equiv 1 \pmod{p^2}$. Then the p -adic logarithms of δ^g exist in all completions \mathbb{K}_\wp and for all $g \in G$. If $A \subset \mathbb{K}_p$ is a multiplicative group, we write the action of G exponentially, so $a^g = g(a)$. If G is not commutative and $g, h \in G$ we have

$$(1) \quad a^{gh} = (a^g)^h = h \circ g(a),$$

and the definition of a contravariant multiplication $G \times G \rightarrow G$ with $g \cdot h = h \circ g$ makes A into a right $\mathbb{Z}_p[G]$ -module, and likewise for $\mathbb{Z}[G]$ -modules. In particular, U, \bar{E} are $\mathbb{Z}_p[G]$ -modules and Minkowski units generate submodules of maximal $\mathbb{Z}_p - \text{rank}$: since \mathbb{K} is dense in \mathbb{K}_p , it follows that $\mathbb{Z}_p - \text{rank}(\bar{E}) = \mathbb{Z}_p - \text{rank}(\delta^{\mathbb{Z}_p[G]})$. With this structure we also define

$$\delta^\top = \{x \in \mathbb{Z}[G] : \delta^x = 1\}, \quad \delta_p^\top = \{x \in \mathbb{Z}_p[G] : \delta^x = 1\},$$

the \mathbb{Z} - and \mathbb{Z}_p -annihilators of δ . Then Leopoldt's conjecture is also equivalent to

$$(2) \quad \delta_p^\top = \delta^\top \otimes_{\mathbb{Z}} \mathbb{Z}_p.$$

In the context of this conjecture we are interested in ranks and not in torsion of modules over rings. It is thus a useful simplification to tensor these modules with fields, so we introduce the following

Definition 1. Let G be a finite group and A, B a \mathbb{Z} , respectively a \mathbb{Z}_p -module, which are torsion free. Let $a \in A, b \in B$. We denote

$$\begin{aligned} \hat{A} &= A \otimes_{\mathbb{Z}} \mathbb{Q}, & \hat{a} &= a \otimes 1, \\ \tilde{B} &= B \otimes_{\mathbb{Z}_p} \mathbb{Q}_p, & \tilde{b} &= a \otimes 1, \end{aligned}$$

Note that $\mathbb{Z} - \text{rank}(A) = \mathbb{Q} - \text{rank}(\hat{A})$ and $\mathbb{Z}_p - \text{rank}(B) = \mathbb{Q}_p - \text{rank}(\tilde{B})$. We shall simply write $\text{rank}(X)$ for the rank of a module when the ring of definition is clear (being one of \mathbb{Z}, \mathbb{Z}_p or \mathbb{Q}, \mathbb{Q}_p .)

For instance, $\tilde{E} = \bar{E} \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$. The definition of \hat{E} is not important for absolute extensions, but relevant in relative extensions \mathbb{L}/\mathbb{K} , when $\mathbf{N}_{\mathbb{L}/\mathbb{K}}(E(\mathbb{L})) \subsetneq E(\mathbb{K})$.

We start with the case of an absolute extension \mathbb{K}/\mathbb{Q} , as introduced above. Let $r = r_1 + r_2 - 1 = \mathbb{Z} - \text{rank}(E)$ and $H = \{g_1, g_2, \dots, g_r\} \subset G \setminus \{1\}$ be a maximal set of

automorphisms, such that δ^{g_i} are \mathbb{Z} - independent. In particular, there is a \mathbb{Z} - linear map $e : \mathbb{Z}[G] \rightarrow \mathbb{Z}[H]$ such that

$$(3) \quad \delta^\sigma = \delta^{e(\sigma)}$$

for each $\sigma \in G$. The map is the identity on H and extends to G due to the Minkowski property, which implies that $\delta^{\mathbb{Z}[H]} = \delta^{\mathbb{Z}[G]}$.

We have the following consequence of Theorem 2

Lemma 1. *Let the notations be like above and $\mathbb{Z}' = \mathbb{Q}' \cap \mathbb{Z}_p$ be the integers in the algebraic closure $\mathbb{Q}' \subset \mathbb{Q}_p$ of \mathbb{Q} . Then*

$$\delta_p^\top \cap \mathbb{Z}'[G] = \delta^\top.$$

In particular, if $\delta_p^\top = \alpha \mathbb{Z}_p[G]$ with $\alpha \in \mathbb{Z}'[G]$, then Leopoldt's conjecture holds for \mathbb{K} .

Proof. Let $\varphi \in P$ be fixed and $\delta_\tau = \iota_\varphi(\delta^\tau)$; then $\delta_\tau \in \mathbb{Z}'$. Since $\{\delta^\tau : \tau \in H\}$ are \mathbb{Z} - independent, $\{\delta_\tau : \tau \in H\}$ are a fortiori \mathbb{Z} - independent. Indeed, if $t \in \mathbb{Z}[H]$ was a linear dependence for δ_τ , such that $\iota_\varphi(\delta^t) = 1$, then $d = \delta^t \in E$ verifies $\iota_\varphi(d) = 1$. But in the diagonal embedding of E , a projection is 1 if and only if the unit itself is 1, thus $d = 1$: a contradiction of the independence of $\delta^\tau, \tau \in H$.

Let $\theta_0 \in \delta_p^\top \cap \mathbb{Z}'[G]$; in view of (3), $\theta = e(\theta_0) \in \delta_p^\top \cap \mathbb{Z}'[H]$ is also an annihilator. Let $\theta = \sum_{\tau \in H} c_\tau \tau, c_\tau \in \mathbb{Z}'$. We show that Theorem 2 implies $\theta = 0$, so $\theta_0 \in e^{-1}(0) \subset \mathbb{Z}[G]$ for all $\theta_0 \in \delta_p^\top \cap \mathbb{Z}'[G]$, which is the claim.

We have $\iota_\varphi(\delta^\theta) = \prod_{\tau \in H} \delta_\tau^{c_\tau} = 1 \in \mathbb{K}_\varphi$, and taking the p - adic logarithm we find the vanishing linear form in logarithms

$$\sum_{\tau \in H} c_\tau \log_p(\delta_\tau) = 0.$$

Since $c_\tau, \delta_\tau \in \mathbb{Z}'$ and $\{\delta_\tau : \tau \in H\}$ are \mathbb{Z} - independent, the Theorem of Baker and Brumer implies that $\theta = 0$.

If $\delta_p^\top = \theta \mathbb{Z}_p[G]$ and $\theta \in \mathbb{Z}'[G]$, then the proof above shows that $\theta \in \mathbb{Z}[G]$, which implies (2) and confirms Leopoldt's conjecture. □

We now treat the case of relative galois extensions:

Lemma 2. *Let \mathbb{L}/\mathbb{K} be a relative galois extension of number fields with abelian group G , let $N = \sum_{\sigma \in G} \sigma$ and let $\delta \in E(\mathbb{L})$ be a unit such that*

$$\delta_{\mathbb{L}/\mathbb{K}}^\top := \{x \in \mathbb{Z}[G] : \delta^x \in \mathbb{K}\} = N\mathbb{Z}[G].$$

Then

$$\delta_{p, \mathbb{L}/\mathbb{K}}^\top := \{x \in \mathbb{Z}_p[G] : \delta^x \in \mathbb{K}\} = N\mathbb{Z}_p[G].$$

Proof. Since G is a abelian, the extension \mathbb{L}/\mathbb{K} arises from a succession of cyclic extensions of prime degree, so it suffices to assume this case. Let $G = \langle \sigma \rangle$ with $|G| = [\mathbb{L} : \mathbb{K}] = q$, for a prime q which is not necessarily different from p . The group $\mathbb{Q}_p[G]$ decomposes as $\mathbb{Q}_p[G] = e_1\mathbb{Q}_p[G] \oplus (1 - e_1)\mathbb{Q}_p[G]$, where e_1 is the idempotent $\frac{N}{q}$. Suppose thus that $\tilde{\delta}_{p,\mathbb{L}/\mathbb{K}}^\top = (ae_1 + be_\chi)\mathbb{Q}_p[G]$, where e_χ is a (non trivial) sum of central idempotents for the augmentation part $\mathbb{Q}_p[I_{\mathbb{L}/\mathbb{K}}]$ and $a, b \in \{0, 1\}$. We shall show that $a = 1$ and $b = 0$.

From the definition of $\tilde{\delta}_{p,\mathbb{L}/\mathbb{K}}^\top$ we have

$$\tilde{\delta}^{ae_1+be_2} = N(\tilde{\delta}) \cdot \tilde{\delta}^{be_2} \in \tilde{E}(\mathbb{K}).$$

Since $N(\tilde{\delta}) \in \tilde{E}(\mathbb{K})$ we also have $d := \tilde{\delta}^{be_2} \in \tilde{E}(\mathbb{K})$. The group G is cyclic and e_2 is in the augmentation, so $e_2N = 0$. Taking the norm in the definition of d and using the fact that $d^\sigma = d$ and thus $N(d) = d^q$, we find that

$$\tilde{\delta}^{be_2N} = d^q = \tilde{\delta}^{be_2q} = 1.$$

But $e_2q \in \mathbb{Z}_p[G]$ and thus $\delta^{be_2q} = 1$ and we may apply the Theorem 2 like in the previous proof, concluding that $e_2 = 0$, since by hypothesis there is no rational dependence for δ in the augmentation. \square

This lemma suggests a relative version of Leopoldt's conjecture, which may hold in the abelian case.

Conjecture 1 (Leopoldt for relative extensions). *Let \mathbb{L}/\mathbb{K} be a galois extension. If Leopoldt's conjecture holds for \mathbb{K} then it holds for \mathbb{L} .*

We prove the following special case:

Theorem 3. *Let $\mathbb{L}/\mathbb{K}/\mathbb{Q}$ be a tower of real number fields, such that \mathbb{L}/\mathbb{Q} and \mathbb{K}/\mathbb{Q} are galois with groups G, G' and \mathbb{L}/\mathbb{K} is abelian with group H . If Leopoldt's conjecture holds for \mathbb{K} then it holds for \mathbb{L} .*

Proof. Let \mathbb{M}/\mathbb{L} be the maximal product of \mathbb{Z}_p - extensions that intersects \mathbb{L}_∞ in \mathbb{L}_n , $n \geq 0$ and $\Delta = \text{Gal}(\mathbb{M}/\mathbb{L})$ be the galois group, a $\mathbb{Z}_p[G]$ - module. We shall relate annihilators of some generators $v \in \Delta$ in order to derive annihilators of global Minkowski units. For this we use class field theory and some of the results of [5]. In order to apply Baker theory, we need to start with a global Minkowski unit δ , which is thus algebraic over \mathbb{Q} . There is a local Minkowski unit ξ such that $\xi^\alpha = \delta$, where α is an idempotent that generates $\xi^\top = \{x \in \mathbb{Z}_p[G] : \xi^x \in \bar{E}\}$.

Let $v = \varphi(\xi) \in \Delta$, a generator of the group. By Proposition 2 of [5], \tilde{v}^\top is also generated by α . We use tensoring with \mathbb{Q}_p to circumvent finite indices. The hypothesis that Leopoldt's conjecture holds for \mathbb{K} implies that $\Delta^N = \{1\}$, where $N = \sum_{\sigma \in H} \sigma$. In particular $\tilde{v}^{\tau N} = 1$ for all $\tau \in G$. Since H is abelian, there is a central idempotent $e_\tau \in \mathbb{Q}_p[H]$ which generates \tilde{v}^\top . Likewise, for $a \in \mathbb{Z}[G]$ we define $e_a \in \mathbb{Q}_p[H]$ to be the central idempotent generating $(\tilde{v}^a)^\top$. There is a finite number of irreducible idempotents in $\mathbb{Q}_p[H]$ – this comes from the fact that H is abelian. We construct a $\mu \in v^{(\mathbb{Z}[H])^\times}$ which has the annihilating idempotent $e_\mu = \sum_\tau e_\tau$. Let

$$\begin{aligned} X_\tau &= \{\psi \in \mathbb{Q}_p[G] : e_\psi \cdot e_\tau = e_\psi\} \\ X &= \cup_{\tau \in G} X_\tau, \quad \text{and in general} \\ X_a &= \{\psi \in \mathbb{Q}_p[G] : e_\psi \cdot e_a = e_\psi\}, \quad \forall a \in \mathbb{Q}[G]. \end{aligned}$$

be lists of all irreducible idempotents which occur in e_τ respectively in at least one of the e_τ . Since e_τ is itself an idempotent, $e_\tau = \sum_{\psi \in X_\tau} e_\psi$. Let $\mu_0 = v$; if $X_1 \neq X$, let $\tau \in G$ such that $X_\tau \not\subset X_1$. We construct a $\mu_1 = v^{1+c\tau}$ with $c \in \mathbb{Z}$ and such that $X_{\mu_1} = X_1 \cup X_\tau$. We need to ascertain that the multiplication does not cancel some idempotents $e \in X_1 \cap X_\tau$. Let $c_1 \neq c_2 \in \mathbb{Z}$; if $e \cdot (\tilde{v}^{1+c_1\tau})^\top = 0$ for $k = 1, 2$, then e is not in the annihilator of $\tilde{v}^{(1+c_2\tau) - (1+c_1\tau)} = \tilde{v}^{(c_2-c_1)\tau}$. But to this automorphism corresponds X_τ and we chose $e \in X_1 \cap X_\tau$. Therefore, at the exception of at most finitely many values of c , $X_{1+c\tau} = X_1 \cup X_\tau$. By iterating the process, we see that for all but finitely many $a \in \mathbb{Z}[G]$ we have $X_a = \cup_{\tau \in G} X_\tau$. In particular, we may choose $a \in (\mathbb{Z}[G])^\times$ with the given property, so $\mu = v^a$ also generates Δ . But then e_a annihilates v^τ for all $\tau \in G$ and thus it annihilates Δ . It need however not generate the annihilator $\tilde{\mu}^\top \subset \mathbb{Q}_p[G]$.

Suppose that e_a is central in $\mathbb{Q}_p[G]$ – such as for instance the norm $\mathbf{N}_{\mathbb{L}/\mathbb{K}}$; then we renormalize finding that $e'_a = e_a/|G'|$ is a central idempotent of $\mathbb{Q}_p[G]$. Since it annihilates Δ , we have $e'_a \in \tilde{\Delta}^\top$, which has also a central generator. But then the minimality of e_a shows that $(e'_a) = \tilde{\Delta}^\top$. Taking the inverse Artin map, let

$$\xi' = \phi^{-1}(\mu) = \phi^{-1}(v^a) = \xi^a.$$

Since (e_a) is the annihilator ideal of $\tilde{\mu}$, applying the inverse of the Artin map $\phi^{-1} : \Delta \rightarrow U$, we see that it also is the annihilator of ξ^a in \bar{E} , so $\xi^{ae_a} = \delta'$ is a global Minkowski unit. But $ae_a = e_a a$ since e_a is central; $a \in \mathbb{Z}[G]^*$ being a unit we find that $\xi^{e_a} \in \bar{E}$ is also a Minkowski unit and we may assume that $\alpha = e_a$. Let $b = |H| - e_a \in \mathbb{Z}'[G]$; if $e_a \neq |H|$, then $b \neq 0$ and $\delta^b = 1$. We have an annihilator of δ in $\mathbb{Z}'[G]$ and the Lemma 1 shows that it must be the absolute norm $\mathbf{N}_{\mathbb{L}/\mathbb{Q}}$, thus $e_a = 1$ and $\mathbb{M} = \mathbb{L}$. This shows that Leopoldt's conjecture holds for \mathbb{L} .

We still have to show that e_a must be central. Let $e_\psi \in \mathbb{Q}_p[H]$ be an irreducible idempotent. The group $H \subset G$ is a normal divisor, so for $\tau \in G/H$ a set of coset representatives – or lifts of G' to G – there is a unique irreducible idempotent $\psi(\tau)$ such that $\tau e_\psi = e_{\psi(\tau)} \cdot \tau$. Let $E_\psi = \sum_{\tau \in C} e_{\psi(\tau)}$ be the orbit of ψ under C . Obviously E_ψ is central in $\mathbb{Q}_p[G]$. We claim that if $\psi \in X$ then $\psi(\tau) \in X$ for all $\tau \in C$. Indeed, assuming without restriction of generality that $\psi \in X_1$, then one verifies that $\psi(\tau) \in X_{\tau^{-1}}$ and since $X = \cup_{\tau \in C} X_\tau$, the claim follows. This shows that $e_a = \sum_{\psi \in X} e_\psi$ is central in $\mathbb{Q}_p[G]$ and completes the proof. \square

3. The solvable case and beyond

The above result can be applied to solvable extensions:

Proposition 1. *Let \mathbb{K}/\mathbf{k} be a solvable extension. Then there is an $\mathbb{L} \supseteq \mathbb{K}$ such that:*

1. *There is a chain of extensions $\mathbf{k} \subset \mathbb{K}_0 \subset \mathbb{K}_1 \subset \dots \subset \mathbb{K}_m = \mathbb{L}$ with $\mathbb{L} \not\subseteq \mathbb{K}_{m-1}$.*
2. *Let $M = \text{lcm}(\exp(\mathbb{K}_{i+1}/\mathbb{K}_i))$. Then the M -th root of unity $\zeta_M \in \mathbb{K}_0$.*
3. *The extensions \mathbb{K}_i/\mathbf{k} are galois with groups G_i and $\mathbb{K}_i/\mathbb{K}_{i-1}$ is abelian with group H_i , which is a $\mathbb{Z}[G_{i-1}]$ -module, for all $i > 0$.*

Proof. Let \mathbb{K}/\mathbf{k} be solvable. Let $n = [\mathbb{K} : \mathbf{k}]$ and $\mathbb{K}_0 = \mathbf{k}[\zeta_n]$ contain the n -th roots of unity. Then \mathbb{K}_0/\mathbf{k} is abelian with group G_0 . Let now $x_1 \in \mathbb{K}_0$ be such that $x_1^{1/n} \in \mathbb{K} \setminus \mathbb{K}_0$ – note that x_1 can in general be a power in \mathbb{K}_0 , such that the radical has fixed order n . We declare

$$\mathbb{K}_1 = \prod_{\tau \in G_0} \mathbb{K}_0[x_1^{\tau/n}].$$

Then \mathbb{K}_1/\mathbf{k} is galois with group G_1 and $\mathbb{K}_1/\mathbb{K}_0$ is abelian with group H_1 . Also, H_1 is by construction a cyclic $\mathbb{Z}[G_0]$ -module of fixed exponent. Furthermore, since $x_1^{1/n} \notin \mathbb{K}_0$ we have the inductive decrease of degrees

$$[\mathbb{K} : \mathbb{K}_1] < [\mathbb{K} : \mathbb{K}_0].$$

The procedure can be repeated inductively, taking at each step an $x_i \in \mathbb{K}_i$ such that $x_i^{1/n} \in \mathbb{K} \setminus \mathbb{K}_i$ and building \mathbb{K}_{i+1} . Since the extension \mathbb{K}/\mathbf{k} is constructed by radicals, the procedure eventually stops with $\mathbb{K} \subset \mathbb{K}_m$. This completes the proof. \square

As a consequence we have:

Theorem 4. *Let \mathbb{K}/\mathbf{k} be a finite solvable extension of number fields. If Leopoldt's conjecture holds for \mathbf{k} , then it holds for \mathbb{K} . In particular, the conjecture holds unconditionally for real solvable extensions of fields in which p is totally split.*

Proof. Since \mathbb{K}/\mathbf{k} is solvable, it can be embedded in a galois extension \mathbb{L}/\mathbf{k} which arises as described in Proposition 1. Since we assumed that Leopoldt's conjecture holds for \mathbf{k} , the Theorem 3 implies successively, that it holds for \mathbb{K}_i for $i = 0, 1, \dots, m$. Thus it holds for \mathbb{L} which is a galois extension of \mathbb{K} and the reduction theorem from [5] shows that it also holds for \mathbb{K} . We have shown that the conjecture holds for totally real fields in which p is totally split. If \mathbf{k} is such an extension and \mathbb{K}/\mathbf{k} is real and solvable, then it follows that the conjecture holds for \mathbb{K} . The result is unconditional. In particular, the conjecture holds for solvable extensions of \mathbb{Q} . \square

The next result reduces the Leopoldt conjecture to the simple case (which is not simple at all).

Theorem 5. *Suppose that p is a prime and Leopoldt's conjecture for p holds for all simple extensions of \mathbb{Q} . Then it holds for p .*

Proof. We use induction. Let \mathbb{K}/\mathbb{Q} be a galois extension with group G which is not simple. If there is only one prime above p , then \mathbb{K} is solvable and we have seen that the conjecture holds unconditionally for \mathbb{K} . We assume by induction that it holds for all galois extensions in which p splits in at most $g - 1$ distinct primes, so let $D(\varphi) \subset G$ be the decomposition group of some prime above p : this is a solvable group. We let $C = D(\varphi) \backslash G$ be a set of coset representatives, assume that $|C| = g$ and use induction on the size $F := ef = |D(\varphi)|$. If $F = 1$, then p is totally split, and the result follows from the Theorem 2 in [5]. Suppose thus that the claim holds for all $k < F$. Since G is not simple, there is a subfield $\mathbb{K}' \subset \mathbb{K}$ which is galois over \mathbb{Q} . Thus $[\mathbb{K}' : \mathbb{Q}] = e(\mathbb{K}') \cdot f(\mathbb{K}') \cdot g(\mathbb{K}') < [\mathbb{K} : \mathbb{Q}]$, with the obvious meaning for $e(\mathbb{K}')$, $f(\mathbb{K}')$, $g(\mathbb{K}')$, so either $g(\mathbb{K}') < g$ or $e(\mathbb{K}') \cdot F(\mathbb{K}') < F$. The claim follows by induction. \square

Acknowledgments: Much of the material presented here was completed after a two day visit of intensive work at the Laboratoire de Mathématique Nicolas Oresme of the University of Caen. I am most grateful to Bruno Anglès and David Vauclair for the helpful and stimulating discussions which had an important contribution for clarifying the central ideas of these two papers. I thank all the colleagues in Caen for participation at the work seminars and the questions raised there: most of them received their answer in this paper.

References

- [1] J. AX – “On the units of an algebraic number field”, *Illinois Journal of Mathematics* **9** (1965), p. 584–589.
- [2] A. BRUMER – “On the units of algebraic number fields”, *Mathematika* **14** (1967), p. 121–124.

- [3] M. EMSALEM, H. KISILEWKY & D. WALES – “Indépendance linéaire sur $\overline{\mathbb{Q}}$ de logarithmes p - adiques de nombres algébriques et rang p - adique du groupe des unités d'un corps de nombres”, *Journal of Number Theory* **19** (1984), p. 384–391.
- [4] H. -W. LEOPOLDT – “Zur Arithmetik in abelschen Zahlkörpern”, *J. Reine Angew. Mathematik* **209** (1962), p. 54–71.
- [5] P. MIHĂILESCU – “On Leopoldt's conjecture and a special case of Greenberg's conjecture”, *Proceedings SANT*, Universitätsverlag Göttingen, 2009.

ON LEOPOLDT'S CONJECTURE AND A SPECIAL CASE OF GREENBERG'S CONJECTURE

Preda Mihăilescu

Mathematisches Institut der Universität Göttingen
E-mail : `preda@uni-math.gwdg.de`

Abstract. The conjecture of Leopoldt states that the p -adic regulator of a number field does not vanish. It was proved for the abelian case in 1967 by Brumer, using Baker theory. We show that when \mathbb{K} is a totally real extension in which the prime p is totally split, then Leopoldt's conjecture is equivalent to a special case of Greenberg's conjecture, which we prove. This case allows thus a proof which does not use transcendence theory. At the end of the paper we give a new approach for the general case of Leopoldt's conjecture, which is currently under peer review. The two papers in these Proceedings gather various approaches for attacking the conjecture of Leopoldt.

Par la mère apprenant que son fils est guéri,
par l'oiseau rappelant l'oiseau tombé du nid,
par l'herbe qui a soif et recueille l'ondée,
par le baiser perdu par l'amour redonné,
et par le mendiant retrouvant sa monnaie:

Je vous salue, Marie⁽¹⁾

To Theres and Seraina

March 2009.

2000 Mathematics Subject Classification. 11R23 Iwasawa Theory, 11R27 Units.

The readers are invited to check for updates on the arXiv version
<http://arxiv.org/abs/0905.1274> and/or contact the author by mail.

⁽¹⁾Francis James: *Prière*. Music by Georges Brassens

1. Introduction

Let \mathbb{K}/\mathbb{Q} be a finite galois extension with group G . Dirichlet's unit theorem states that, up to torsion made up by the roots of unity $W(\mathbb{K}) \subset \mathbb{K}^\times$, the units $E = \mathcal{O}(\mathbb{K})^\times$ are a free \mathbb{Z} -module of \mathbb{Z} -rank $r_1 + r_2 - 1$. As usual, r_1 and r_2 are the numbers of real, resp. pairs of complex conjugate embeddings $\mathbb{K} \hookrightarrow \mathbb{C}$. Let p be a rational prime. We consider the set $P = \{\mathfrak{p} \subset \mathcal{O}(\mathbb{K}) : (\mathfrak{p}) \subset p\}$ of distinct prime ideals above p and let

$$\mathbb{K}_p = \prod_{\mathfrak{p} \in P} \mathbb{K}_{\mathfrak{p}} = \mathbb{K} \otimes_{\mathbb{Q}} \mathbb{Q}_p$$

be the product of all completions of \mathbb{K} at primes above p . Let $\iota : \mathbb{K} \hookrightarrow \mathbb{K}_p$ be the diagonal embedding. We write $\iota_{\mathfrak{p}}(x)$ for the projection of $\iota(x)$ in the completion at $\mathfrak{p} \in P$. If $y \in \mathbb{K}_p$, then $\iota_{\mathfrak{p}}(y)$ is simply the component of y in $\mathbb{K}_{\mathfrak{p}}$.

If $U \subset \mathbb{K}_p^\times$ are the units, thus the product of local units at the same completions, then E embeds diagonally via $\iota : E \hookrightarrow U$. Furthermore one can use ι for inducing a galois structure on \mathbb{K}_p (see §2.1).

Let $\bar{E} = \overline{\iota(E)} \subset U$ be the closure of $\iota(E)$; this is a \mathbb{Z}_p -module with \mathbb{Z}_p -rank $\text{rank}(\bar{E}) \leq \mathbb{Z}$ -rank $(E) = r_1 + r_2 - 1$. The difference

$$\mathcal{D}(\mathbb{K}) = (\mathbb{Z} - \text{rank}(E)) - (\mathbb{Z}_p - \text{rank}(\bar{E}))$$

is called the *Leopoldt defect*. The defect is positive if relations between the units arise in the local closure, which are not present in the global case. Equivalently, if the p -adic regulator of \mathbb{K} vanishes.

It was conjectured by Leopoldt that $\mathcal{D} = 0$ for all number fields. The conjecture of Leopoldt was proved in 1967 for abelian extensions by Brumer [3], using a local version of Baker's linear forms in logarithms. It is still open for arbitrary non abelian extensions.

It is easy to show that if \mathbb{K}'/\mathbb{Q} is an extension such that Leopoldt's conjecture holds for some galois extension \mathbb{K}/\mathbb{Q} which contains \mathbb{K}' , then it holds for \mathbb{K}' . See for instance [5], the final remark on p. 108. We may thus concentrate on galois extensions of \mathbb{Q} .

1.1. Connection to Iwasawa theory. We shall take here an approach using class field and Iwasawa theory. Let \mathbb{B}/\mathbb{Q} be the \mathbb{Z}_p -cyclotomic extension of \mathbb{Q} ; we assume that $\mathbb{K} \cap \mathbb{B} = \mathbb{Q}$ and let $\mathbb{K}_\infty = \mathbb{B} \cdot \mathbb{K}$, the \mathbb{Z}_p -cyclotomic extension of \mathbb{K} . The intermediate fields will be denoted by $\mathbb{B}_n, \mathbb{K}_n$, respectively. As usual, we let τ be a topological generator of $\Gamma = \text{Gal}(\mathbb{K}_\infty/\mathbb{K})$ and $T = \tau - 1$, $\Lambda = \mathbb{Z}_p[[T]]$.

For all $n \geq 0$ we let $\Gamma'_m = \text{Gal}(\mathbb{K}_m/\mathbb{K})$, A_n the p -Sylow subgroups of \mathbb{K}_n and A the projective limit, a Λ -module. Then $A'_T = \{a \in A : a^T = 1\}$, a Λ -submodule. If \mathbb{K} is totally real, an important conjecture of Greenberg states that A is finite. In particular, A'_T must be finite. Our first result is

Theorem 1. *Let \mathbb{K} be a totally real extension of \mathbb{Q} . Then A'_T is finite.*

The proof uses galois and Kummer theory together with class field theory. We relate this result to Leopoldt's conjecture by considering \mathbb{M}/\mathbb{K} , the product of all \mathbb{Z}_p -extensions of \mathbb{K} , with $\Delta = \text{Gal}(\mathbb{M}/\mathbb{K})$. If p is totally split in \mathbb{K} , then it was observed by Greenberg that $\mathbb{M}/\mathbb{K}_\infty$ is an unramified product of \mathbb{Z}_p -extension. This leads to a contradiction with Theorem 1 and the first result of the paper, which is:

Theorem 2. *Let \mathbb{K}/\mathbb{Q} be a totally real galois extension in which the rational prime p is completely split. Then Leopoldt's conjecture is true for \mathbb{K} .*

The main result proves the conjecture for arbitrary fields, using reflection and the Artin symbol. It is based on a direct investigation of *phantom* fields which arise as a consequence of $\mathcal{D}(\mathbb{K}) > 0$ and have, like inexistent objects tend to, very intriguing properties. The general case of the conjecture can be proved by showing that the *L-phantom* \mathbb{Z}_p -extensions, related to the Leopoldt defect, cannot exist. The paper covers a variety of correlated ideas; while one alone leads to a final proof, we found it important to keep the overview of the various approaches attempted.

1.2. Some notations and conventions. Throughout the paper, p is a rational prime, $\mathbb{K} = \mathbb{Q}[\alpha]$ is a finite galois extension of degree n and $r = r_1 + r_2 - 1$ is the \mathbb{Z} -rank of the global units $E = E(\mathbb{K}) = \mathcal{O}(\mathbb{K})$ and $P = \{\mathfrak{p} \subset \mathcal{O}(\mathbb{K}) : (\mathfrak{p}) \subset p\}$ are the primes of \mathbb{K} above p . The galois group is $G = \text{Gal}(\mathbb{K}/\mathbb{Q})$. The field \mathbb{K} may be more special in certain sections, and this is stated in the respective part of the text. We let \mathbb{K}/\mathbb{Q}_p be a finite extension which is isomorphic to the completions of \mathbb{K} at places above p .

The cyclotomic \mathbb{Z}_p -extension of \mathbb{K} is a tower $(\mathbb{K}_n)_{n \in \mathbb{N}} = (\mathbb{K} \cdot \mathbb{B}_n)_{n \in \mathbb{N}}$ with injective limit \mathbb{K}_∞ . We let $\mathbb{H}_n, \mathbb{M}_n$ be the maximal p -abelian unramified, respectively p -ramified extensions of \mathbb{K}_n and $X_n = \text{Gal}(\mathbb{H}_n/\mathbb{K}_n)$, $Y_n = \text{Gal}(\mathbb{M}_n/\mathbb{H}_n)$. From class field theory, one has ([6], Chapter 5, Theorem 5.1):

$$(1) \quad Y_n \cong p\text{-part of } U(\mathbb{K}_n)/\overline{E(\mathbb{K}_n)}.$$

It is known that there is a Minkowski unit $\delta \in E$ ([8], lemma 5.27), i.e. a unit such that

$$\mathbb{Z}\text{-rank}(\delta^{\mathbb{Z}[G]}) = r.$$

In the context of Leopoldt's conjecture we are interested in ranks and not in torsion of modules over rings. It is thus a useful simplification to tensor these modules with fields, so we introduce the following:

Definition 1. Let G be a finite group and A, B a \mathbb{Z} , respectively a \mathbb{Z}_p -module, which are torsion free. Let $a \in A, b \in B$. We denote

$$\begin{aligned}\hat{A} &= A \otimes_{\mathbb{Z}} \mathbb{Q}, & \hat{a} &= a \otimes 1, \\ \tilde{B} &= B \otimes_{\mathbb{Z}_p} \mathbb{Q}_p, & \tilde{b} &= a \otimes 1,\end{aligned}$$

We note that $\mathbb{Z} - \text{rank}(A) = \mathbb{Q} - \text{rank}(\hat{A})$ and $\mathbb{Z}_p - \text{rank}(B) = \mathbb{Q}_p - \text{rank}(\tilde{B})$. We shall simply write $\text{rank}(X)$ for the rank of a module when the ring of definition is clear (being one of \mathbb{Z}, \mathbb{Z}_p or \mathbb{Q}, \mathbb{Q}_p .)

2. Local theory

In this section we review the galois structure of the idèles that are trivial at all primes, except the ones above p and the ramified \mathbb{Z}_p -extensions of a finite extension of \mathbb{Q}_p .

2.1. Galois structure of some idèle-groups.

Theorem 3. Let $\mathbb{K} = \mathbb{Q}[\alpha]$, p, P and \mathbb{K}_p be like above, suppose that $f \in \mathbb{Z}[X]$ is a minimal polynomial of α and $\iota: \mathbb{Q} \hookrightarrow \mathbb{Q}_p$ is the natural embedding. Then

$$\mathbb{K}_p = \mathbb{Q}_p[X]/(\iota(f))$$

is a galois algebra with group $G = \text{Gal}(\mathbb{K}/\mathbb{Q})$ and the embedding ι extends to an embedding $\mathbb{K} \hookrightarrow \mathbb{K}_p$ which commutes with the galois action. The image $\iota(\mathbb{K}) \subset \mathbb{K}_p$ is dense in the product topology.

Proof. Let e, f, g denote as usual, the ramification index, the degree of the residual fields and the splitting index of the primes above p . The polynomial $\iota(f(X))$ is separable over \mathbb{Q}_p and splits in g polynomials of degree ef . Thus $\mathbb{K}_p = \mathbb{Q}_p[X]/(\iota(f))$ is the product of g isomorphic local, unramified extensions of degree ef . Each completion $\mathbb{K}_{\mathfrak{p}} \cong \mathbf{K}$ is a ramified extension of degree e of the unramified extension $\mathbf{K}_0/\mathbb{Q}_p$ of degree f .

It follows from the Chinese Remainder Theorem that $\iota: \mathbb{Q} \hookrightarrow \mathbb{Q}_p$ extends to an embedding $\iota: \mathbb{K} \hookrightarrow \mathbb{Q}_p[X]/(\iota(f))$ and that the image of \mathbb{K} is dense in \mathbb{K}_p . By continuity, the galois action of G extends to \mathbb{K}_p and commutes with the embedding.

Indeed for any $t \in \mathbb{K}_p$ there is a $h \in \mathbb{Q}_p[X]$ such that $t = h(\iota(\alpha))$. Let $h_n \in \mathbb{Q}[X]$ approximate h , so $\lim_{n \rightarrow \infty} \iota(h_n) = h$; setting $t_n = h_n(\alpha) \in \mathbb{K}$ we also have $\iota(t_n) = \iota(h_n(\alpha)) \rightarrow h(\iota(\alpha)) = t$. For any $\sigma \in H$ we define $\sigma(t) = h(\iota(\sigma(\alpha)))$. This action is well defined and commutes with the embedding, since for $t \in \mathbb{K}$ we have

$$\iota(\sigma(t)) = \iota(h(\sigma(\alpha))) = h(\iota(\sigma(\alpha))) = \sigma(\iota(t)).$$

□

The group under consideration is thus the multiplicative subgroup of idèles which are trivial at all places above rational primes different from p . By the Chinese Remainder Theorem we identify $u \in U$ with $(\iota_\varphi(u))_{\varphi \in P}$.

2.2. Algebra in the group ring, units and their presentation. We shall use multiplicative notation so all actions are from the right. If $A \subset \mathbb{Q}_p[G]$ is some module, then there is an idempotent $\alpha \in \mathbb{Q}_p[G]$ such that $A = (\alpha) = \alpha\mathbb{Q}_p[G]$. This follows from the proof of Maschke's Theorem [1], p. 116. The annihilator ideal of A is $(1 - \alpha)\mathbb{Q}_p[G]$ and conversely, A is the annihilator of $(1 - \alpha)$: thus $(\alpha) \cdot (1 - \alpha) = (1 - \alpha)\alpha$: this is a rephrasing of Maschke's theorem which makes explicit use of idempotents: $(1 - \alpha)\mathbb{Q}_p[G]$ is a complement of A .

If X is a ring and $R \subset \mathbb{Q}_p[G]$ is an ideal such that X is an R -module, for $x \in X$ we shall write $x^\top = \{a \in R : x^a = 1\}$ for its annihilator module. We shall work when possible with $\mathbb{Q}_p[G]$ -modules, which are endowed with a vector space structure. Note that elements $a \in \mathbb{Q}_p[G]$ act both from the left and from the right, thus generating left and right ideals; these ideals always have at least one generating idempotent. Elements $a \in R \subset \mathbb{Q}_p[G]$, can be regarded as linear maps of the \mathbb{Q}_p -vector space R and as such we have

$$(2) \quad \text{rank}(a) = \dim(aR) = \dim R - \text{rank}(1 - a).$$

We show now that there are local Minkowski units and describe their relation with global ones. Serre proves in [7], §1.4, Proposition 3, in the case when \mathbb{K}/\mathbb{Q}_p is a local field, that the group $U^{(1)}(\mathbb{K})$ contains a cyclic $\mathbb{Z}_p[G]$ module of finite index, which is thus isomorphic to $\mathbb{Z}_p[G]$. Using this result one easily constructs units of finite index in U . Let $\varphi \in P$ be fixed and $v \in \mathbb{K}_\varphi$ be a local Minkowski unit, according to Serre. Then we define $\xi = \xi(v) \in U$ by:

$$\iota_\tau \varphi(\xi) = \begin{cases} v & \text{for } \tau = 1, \\ 1 & \text{for } \tau \in G, \tau \neq 1. \end{cases}$$

Let D_φ be the decomposition group of φ and $C = D_\varphi \backslash G$ coset representatives. Then C acts on ξ and for $\sigma \in C$, the unit ξ^σ verifies:

$${}_{\iota_\tau \varphi}(\xi) = \begin{cases} v & \text{for } \tau = \sigma, \\ 1 & \text{for } \tau \in G, \tau \neq \sigma. \end{cases}$$

We denote units $u \in U$ such that $[U : u^{\mathbb{Z}_p[G]}] < \infty$ by *local Minkowski units*. The previous construction shows that such units exist and they generate a module which is isomorphic to $\mathbb{Z}_p[G]$. We define:

$$(3) \quad U' = \{u \in U^{(1)} : \mathbf{N}_{\mathbb{K}_p/\mathbb{Q}_p}(u) = 1\}$$

with $U^{(1)}/U' = U^{(1)}(\mathbb{Z}_p) \cong \mathbb{Z}_p$. Therefore $\tilde{U}' \cong (1 - N/|G|)\mathbb{Q}_p[G]$, a subalgebra of $\mathbb{Q}_p[G]$. For any \mathbb{K} we have $\bar{E}(\mathbb{K}) \subset U'$, so $U^{(1)}(\mathbb{Z}_p)$ is mapped injectively in Δ by the Artin map. By choosing $\delta \in E$ a global Minkowski unit, one may find a local one $\xi \in U'$ such that

$$(4) \quad \tilde{\xi}^\alpha = \tilde{\delta}, \quad \text{with } \alpha \in \mathbb{Q}_p[G].$$

We shall say the triple $(\xi, \delta, \alpha) \in U' \times E \times \mathbb{Q}_p[G]$ is a *presentation* of \bar{E} . If \mathbb{K}/\mathbb{Q} is a real extension, we have

$$(5) \quad (\tilde{U}')^\top = \hat{E}^\top \otimes_{\mathbb{Q}} \mathbb{Q}_p,$$

a canonic construction for a submodule which is isomorphic to \tilde{E} iff Leopoldt's conjecture is true for \mathbb{K}^+ .

3. Auxiliary fields

We assume in this section that \mathbb{K}/\mathbb{Q} is galois and contains the p -th roots of unity, so $\mu_{p^n} \subset \mathbb{K}_n$. The next lemma gives a canonic construction of a maximal extension $\mathbb{K} \subset \mathbb{M}_0 \subset \mathbb{M}$ which intersects \mathbb{K}_∞ in a finite extension of \mathbb{K} :

Lemma 1. *Let \mathbb{K}/\mathbb{Q} be galois with group G and \mathbb{M}/\mathbb{K} the product of all \mathbb{Z}_p extensions of \mathbb{K} , containing \mathbb{K}_∞ , the cyclotomic extension. There is a canonic subfield $\mathbb{M}_0 \subset \mathbb{M}$ with $\mathbb{M} \cap \mathbb{K}_\infty = \mathbb{K}_i$ for some $i \geq 0$ and $\text{Gal}(\mathbb{M}/\mathbb{M}_0)$ is a G -invariant group, isomorphic to \mathbb{Z}_p .*

Proof. Let $\Gamma' = \varphi(U^{(1)}(\mathbb{Z}_p)) \subset \Delta$, with $U^{(1)}(\mathbb{Z}_p) = (1+p)^{\mathbb{Z}_p}$. Since $U^{(1)}(\mathbb{Z}_p) \cap \bar{E} = \{1\}$, the group Γ' is isomorphic to \mathbb{Z}_p and G -invariant as a $\mathbb{Z}_p[G]$ -module. Therefore it acts by restriction on $\mathbb{K}_\infty/\mathbb{K}$ as \mathbb{Z}_p -subgroup of Γ , which implies the claim. It is

an interesting question to establish the cases when $i > 0$, but we do not need the answer here. \square

Definition 2. Let the \mathbb{Z}_p - torsion $A^\circ \subset A$. We define, for $*$ the Iwasawa involution (see 4.1):

$$\begin{aligned} F &= \varphi(A^\circ \cdot TA) \subset \text{Gal}(\mathbb{H}_\infty/\mathbb{K}_\infty), \\ F^* &= \varphi(A^\circ \cdot T^*A) \subset \text{Gal}(\mathbb{H}_\infty/\mathbb{K}_\infty), \end{aligned}$$

with φ the Artin symbol and \mathbb{H}_∞ the maximal unramified p - abelian extension of \mathbb{K}_∞ . With this we let

1.

$$\mathbb{H}_{T^*} = \mathbb{H}_\infty^{F^*}, \quad \mathbb{H}_T = \mathbb{H}_\infty^F$$

and let $A_{T^*}, A_T \subset A$ be some subsets with

$$\varphi(A_{T^*}) \cong \text{Gal}(\mathbb{H}_{T^*}/\mathbb{H}_\infty), \quad \varphi(A_T) \cong \text{Gal}(\mathbb{H}_T/\mathbb{H}_\infty).$$

The field \mathbb{H}_{T^*} is the L - phantom field and will also be denoted by Φ .

2. The unit extension of \mathbb{K}_∞ is

$$\mathbb{M}_E = \bigcup_{n=1}^{\infty} \mathbb{K}_n \left[E_n^{1/p^n} \right],$$

where E are the units of \mathbb{K}_n . In Lang's notation ([6], Chapter 6) we can also write $\mathbb{M}_E = \mathbb{K}_\infty[E^{1/p^\infty}]$.

3. For $X \in \{A_T, A_{T^*}\}$, the extensions

$$\mathbb{M}_E \left[X^{1/p^\infty} \right] = \bigcup_{n=1}^{\infty} \mathbb{M}_E[X_n^{1/p^n}],$$

with $X = \lim_n X_n \subset A_n$. The finite extensions $\mathbb{M}_E[X_n^{1/p^n}]$, depend in fact on the classes in X_n and can be built as follows: let $\mathfrak{Q}_n \in x_n \in X_n$ be an ideal of order p^m and $p^N = \max(p^m, p^n)$, $\mathfrak{q}_n = \mathfrak{Q}_n^{p^N}$. Then the extension $\mathbb{M}_{E,n}[\mathfrak{q}_n^{1/p^n}]$ depends by construction only on the class x_n .

By definition, the sets A_T, A_{T^*} are representatives of some factor groups which are free \mathbb{Z}_p - modules. Therefore, no information about torsion can be expected from the properties of the fields $\mathbb{H}_{T^*}, \mathbb{H}_T$. Since we are only interested in \mathbb{Z}_p - ranks, the particular choice of these sets, will have no impact on the results.

We shall use the following fundamental fact from Kummer theory:

Fact 1. Let $\mathbb{L} = \cup_{n=1}^{\infty} \mathbb{L}_n$ with $\mathbb{L}_n/\mathbb{K}_n$ Kummer extensions of exponent dividing p^n and $\mathbb{K}_{n+1} \supset \mathbb{K}_n$. If $\mathbb{L}/\mathbb{K}_{\infty}$ is p -ramified, then there are Kummer radicals $B_n \in \mathbb{K}_n^{\times}$ such that

1. $\mathbb{L}_n = \mathbb{K}_n[B_n^{1/p^n}]$.
2. For each $b_n \in B_n$ there is an ideal $\mathfrak{B} \subset \mathcal{O}(\mathbb{K}_n)$ and an ideal \mathfrak{p} which is divisible only by primes above p such that $(b) = \mathfrak{p} \cdot \mathfrak{B}^{p^n}$. In particular, b_n may be a unit.
3. If $\mathbb{L} \subset \mathbb{M}$, then $b_n^{T^*} \in (\mathbb{K}_m^{\times})^{p^m}$.

Proof. Point 1 is a consequence of \mathbb{L}_n being Kummer extensions. Since \mathbb{L}_n is p -ramified, we deduce point 2. Finally, if $\mathbb{L} \subset \mathbb{M}$, it is by definition abelian over \mathbb{K} . Therefore, if $\alpha \in \text{Gal}(\mathbb{L}_m/\mathbb{K}_m)$ is a generator, then $\alpha^T = 1$ and Kummer pairing yields

$$\langle a, \alpha^T \rangle = \langle a^{T^*}, \alpha \rangle = 1,$$

which confirms point 3, the Kummer pairing being non-degenerate. □

Most of the rest of this paper concerns the fields $\mathbb{M}_E[A_T^{1/p^{\infty}}]$ and $\mathbb{M}_E[A_{T^*}^{1/p^{\infty}}]$. In the CM case, the first extension is connected to the proof of Theorem 1. For the second extension, we show that it is non-trivial exactly when $\mathcal{D}(\mathbb{K}) > 0$.

3.1. An observation of Greenberg and proof of Theorem 2. In his seminal paper [4], at the end of §2, Greenberg remarks the following phenomenon (we adapt the notation to the one used above): *If \mathbb{K} is an arbitrary number field in which p splits completely, then ... we observe that by local class field theory, every \mathbb{Z}_p -extension of \mathbb{Q}_p is contained in the composite of two of them, the cyclotomic one and the unramified one. It follows easily that $\mathbb{M}/\mathbb{K}_{\infty}$ is an unramified extension of \mathbb{K}_{∞} .*

The remark implies that $\mathbb{M} \subset \mathbb{H}_{\infty}$ and $\text{Gal}(\mathbb{M}/\mathbb{K}_{\infty})$ is fixed by T . It follows that $\mathbb{M} = \mathbb{H}_T$ in the notation of Definition 2. The Theorem 1 implies that $\mathbb{H}_T = \mathbb{K}_{\infty}$ if \mathbb{K} is real and this proves Theorem 2.

4. Proof of Theorem 1

Let \mathbb{K} be a Galois extension and $\mathbb{K}_{\infty}/\mathbb{K}$ be the cyclotomic \mathbb{Z}_p -extension, A the projective limit of the p -parts A_n of the class groups of \mathbb{K}_n and τ be a topological generator of $\Gamma = \text{Gal}(\mathbb{K}_{\infty}/\mathbb{K})$. Let $\mathbb{H}_n, \mathbb{M}_n$ be the maximal p -abelian extensions of \mathbb{K}_n which are unramified, resp. p -ramified, thus $\mathbb{M}_n = \mathbb{M}(\mathbb{K}_n)$ in the notation of the introduction. If \mathbb{K} is CM, then complex conjugation acts naturally on Galois groups and induces plus and minus parts of $A_n, \mathbb{H}_n, \mathbb{M}_n$.

The first result we prove is independent of the action of complex conjugation on \mathbb{K} : it indicates that if Leopoldt's conjecture is true for \mathbb{K} , then $\mathbb{M} \subset \mathbb{M}_E$. More precisely, we show:

Theorem 4.

$$(6) \quad \mathbb{Z}_p\text{-rank}(\text{Gal}((\mathbb{M} \cap \mathbb{M}_E)/\mathbb{H}_\infty)) = r_2.$$

When \mathbb{K} is a CM field, by class field theory ([6], Chapter 5, Theorem 5.1) and since E^- is finite, being equal to the group of roots of unity, it follows that

$$(7) \quad \text{pro-}p \text{ part}(\text{Gal}(\mathbb{M}^-/\mathbb{H}^-)) = \text{pro-}p \text{ part}(U^-).$$

Thus (6) specializes to

$$(8) \quad \mathbb{M}^- \subset \mathbb{M}_E^-.$$

In both cases, Leopoldt's conjecture is equivalent to

$$(9) \quad \mathbb{M} \subset \mathbb{M}_E.$$

4.1. Proof of Theorem 4. Let $\zeta \in \mathbb{C}$ be a primitive p -th root of unity. We show that it suffices to prove (6) in the case when $\zeta \in \mathbb{K}$. Indeed, if \mathbb{K}/\mathbb{Q} is any galois field with group G , then $\mathbb{K}[\zeta]$ is a finite galois extension. Since \mathbb{M} and \mathbb{M}_E are both galois over \mathbb{Q} , so is their intersection. For a field \mathbf{k} , let $I(\mathbf{k}) = \mathbb{M}(\mathbf{k}) \cap \mathbb{M}_E(\mathbf{k})$ and $D(\mathbf{k}) = \text{Gal}(I(\mathbf{k})/\mathbf{k})$. The extension $I(\mathbb{K}[\zeta])/\mathbb{K}$ is galois and we let $D_0 = \{x \in \text{Gal}(\mathbb{K}[\zeta]/\mathbb{K}) : cx \in ([\mathbb{K}[\zeta] : \mathbb{K}]) - \mathbf{N}_{\mathbb{K}[\zeta]/\mathbb{K}} \cdot \text{Gal}(\mathbb{K}[\zeta]/\mathbb{K}), c \in \mathbb{Z}_p\}$ be a radical analog to the one defined in Lemma 1 and $J(\mathbb{K}) = I(\mathbb{K}[\zeta])^{D_0}$. Then $J(\mathbb{K}) \subset I(\mathbb{K})$ by construction and comparing \mathbb{Z}_p -ranks, one finds that if (6) holds for $I(\mathbb{K}[\zeta])$ it holds for $J(\mathbb{K})$.

We assume thus from now on that \mathbb{K} is galois with group G and it contains the q -th but not the qp -th roots of unity, for $q = p^n, n \geq 1$. In particular, $r_1 = 0, r_2 = [\mathbb{K} : \mathbb{Q}]/2$. We need some care with the numeration of the intermediate extensions of \mathbb{K}_∞ and shall simply write $\mathbb{K}_0 = \mathbb{K}_1 = \dots = \mathbb{K}_n \neq \mathbb{K}_{n+1}$, so $\mathbb{K}_{n+1} = \mathbb{K}[\zeta_q^{1/p}]$. We denote like usual by τ a topological generator of $\mathbb{K}_\infty/\mathbb{K}$ and $T = \tau - 1, \Lambda = \mathbb{Z}_p[[T]]$, which is a local ring with maximal ideal $\mathcal{M} = (p, T)$. Note that τ fixes \mathbb{K} , so it is the n -th power of a topological generator for the extension $\mathbb{K}_\infty/\mathbb{K}_{-1}$, with \mathbb{K}_{-1} the subfield of \mathbb{K} fixed by $\text{Gal}(\mathbb{Q}[\zeta_q]/\mathbb{Q}[\zeta_p])$.

Let the cyclotomic character act on Λ by $\kappa(\tau) = (q + 1)\tau$ so that the Iwasawa involution becomes:

$$T = \tau - 1 \mapsto T^* = \frac{q - T}{T + 1}.$$

The map $*$: $\Lambda \rightarrow \Lambda$ is an automorphism ([6], p. 150). For $m = n + l, l \geq 0$, we let $\Gamma_m = \text{Gal}(\mathbb{K}_m/\mathbb{K})$; the polynomial $\omega_m = (T + 1)^{p^l} - 1 \in \mathcal{M}^{p^m}$ annihilates \mathbb{K}_m^\times and

$$\mathbb{Z}_p[\Gamma'_m] = \Lambda/(\omega_m).$$

We denote this ring by $\Lambda_m = \mathbb{Z}_p[\Gamma'_m]$ and identify T with its image in Γ_m . Let $N_m = \omega_m/T$; then $p^l \in (\omega_m, \omega_m^*)$ and it follows that

$$(10) \quad N_m^* \cdot T^* \equiv 0 \pmod{p^l}.$$

We prove the existence of systems of units which generate $\mathbb{M}_E \cap \mathbb{M}$.

Lemma 2. *Let $m = l + n > 0$ and $\mathcal{E}_m = E_m^{N_m^*}$. Then*

$$\mathbb{F}_m = \mathbb{K}_m \left[\mathcal{E}^{1/p^m} \right]$$

is an abelian extension of \mathbb{K} and $\text{Gal}(\mathbb{F}_m/\mathbb{K}_m)$ has p - rank r_2 and exponent $e_m \geq p^{l/2}$. The inclusion $\mathbb{F}_m \subset \mathbb{F}_{m+1}$ holds for all $m > n$ and

$$\mathbb{F} = \cup_{l>0} \mathbb{F}_{n+l}$$

is an abelian extension of \mathbb{K}_∞ with galois group of \mathbb{Z}_p - rank r_2 .

Proof. Let $G_m = \text{Gal}(\mathbb{K}_m/\mathbb{Q})$ and $k = |G|$. An element $\alpha \in \mathbb{Z}[G_m]$ acting on E_m has the following development in the group ring:

$$\alpha = \sum_{i=0}^{k-1} A_i(T^*) \cdot \tau_i, \quad \tau_i \in \text{Gal}(\mathbb{K}_0/\mathbb{Q}),$$

where $A_i \in \mathbb{Z}[X]$ have degree $\text{deg}(A_i) < p^m$ and $\alpha_0 = \sum_{i=0}^{k-1} A_i(0)\tau_i$. We show that p - rank(\mathcal{E}_m) = r_2 . If $\delta_0 \in E$ is a Minkowski unit of the ground field, $H \subset G \setminus \{1\}$ is a maximal subset such that $\delta_0^{\mathbb{Z}[H]}$ is a free \mathbb{Z} - module of rank $r_2 - 1$. Let $D_0 = \{\delta_m^\sigma : \sigma \in H \cup \{1\}\}$, which is a system of relative units for E_m/E ; the identity automorphism accounts for $N_{\mathbb{K}_m/\mathbb{K}}^{-1}(1) \in E_m$. The system has \mathbb{Z} - rank r_2 and we write $D = \{d^{N_m^*} : d \in \langle D_0 \rangle_{\mathbb{Z}}\}$ for the \mathbb{Z} - module spanned by the $d^{N_m^*}, d \in D_0$. Thus $\text{rank}(\mathcal{E}_m) \geq r_2$. On the other hand, $\langle D_0 \rangle_{\mathbb{Z}[G]} = \delta_m^{\mathbb{Z}[G]}$ has by construction finite index in \mathcal{E}_m , so $\text{rank}(\mathcal{E}_m) \leq r_2$, which confirms the claim. Explicitly, \mathbb{F}_m will be generated by the radical of D in E_m .

We now show that the exponents of \mathbb{F}_m are diverging. For this we use the following observation of B. Anglès [2], Lemma 2.1, (2): let $l' = [l/2]$, then

$$(11) \quad \omega_m(T) = TN_m \in \left(p^{l'}, T^{p^{l'+1}} \right).$$

We may thus choose $a, b \in \Lambda_m$ with $a \in \Lambda_m^\times$ such that

$$N_m^* = ap^{l'} + bN_{l'+1}^*.$$

If $e \in E_m \setminus E_m^p$, such as for instance $\delta_m^\tau, \tau \in \text{Gal}(\mathbb{K}/\mathbb{Q})$, we see that $\mathbb{K}_m[e^{N_m^*/p^m}]/\mathbb{F}_{n+l'+1}$ is an extension of degree at least l' , so the exponents of \mathbb{F}_m diverge. The units $\delta_m^\sigma, \sigma \in H$ have \mathbb{F}_p independent images in E_m/E_m^p , thus they span a vector space of rank r_2 . It follows from (11), that the system $\{\delta^{\sigma N_m^*}, \sigma \in H\} \subset \mathcal{E}_m$ generates a module of full p -rank r_2 in the quotient $E_m/E_m^{l'+1}$, since a, b only depend of the norm, but not on the units. This leads to the conclusion that not only the exponents of $\text{Gal}(\mathbb{F}_m/\mathbb{K}_m)$ diverge, but the extension is the product of r_2 linearly disjoint extensions of degree at least l' . The extension \mathbb{F}_m/\mathbb{K} is abelian since $N_m^* \cdot T^* \equiv 0 \pmod{p^l}$ and since $N_m^* | N_{m+1}^*$ while $E_m \subset E_{m+1}$, it follows that $\mathbb{F}_m \subset \mathbb{F}_{m+1}$. The injective limit exists and is a product of r_2 independent abelian \mathbb{Z}_p -extensions of \mathbb{K} . \square

The lemma implies Theorem 4:

Corollary 1.

$$\mathbb{Z} - \text{rank}(\mathbb{M}_{E,m} \cap \mathbb{M}) = r_2, \quad \text{for all } m > 0.$$

In particular (6) holds for arbitrary galois extensions \mathbb{K}/\mathbb{Q} , relation (8) for CM extensions, and Theorem 4 is true.

Proof. We have shown that $p - \text{rank}(\text{Gal}(\mathbb{F}_m/\mathbb{K}_m)) = r_2$ and $\mathbb{F}_m \subset \mathbb{M}_{E,m}$ is the maximal subfield which is abelian over \mathbb{K} . The claim follows. \square

4.2. On to Theorem 1. In this section \mathbb{K} is a CM field containing the p -th root of unity, but not the p^2 -th and \mathbb{H}_T, A_T are like in Definition 2, with respect to \mathbb{K}^+ ; in particular the classes in A_T are real. Note that A_T is defined modulo torsion and is not canonic, the methods of this proof have no impact on the torsion part.

Let $a = (a_n) \in A_T$; then $\Lambda a = \mathbb{Z}_p a$ and Nakayama's lemma implies $\text{ord}(a_n) = p^{n+k}$ for $n \geq n_0$ and some $k \in \mathbb{Z}$ which depends on a but not on n . The fields $\mathbf{T}_n(a) = \mathbb{M}_{E,n}[a_n^{1/p^n}]$ and $\mathbf{T} = \cup_{n > n_0} \mathbf{T}_n(a)$ are defined like in the previous section. We shall distinguish the cases when \mathbf{T} is unramified and when it contains a totally ramified \mathbb{Z}_p -subextension. In both cases we show that $\mathbf{T} \subset \mathbb{M}_E$, thus reaching a contradiction with the fact that $\text{ord}(a) = \infty$. Let

$$(12) \quad \mathbb{T} = \mathbb{M}_E[A_T^{1/p^\infty}] = \prod_{a \in A_T} \mathbf{T}(a).$$

Our proof will relay on the following

Lemma 3. *Notations being like above, A_T is infinite if and only if \mathbb{T}/\mathbb{M}_E is infinite.*

Proof. Suppose that A_T is infinite and let $p^m = \max(\exp(A^\circ), \text{Gal}(\mathbb{T}/\mathbb{M}_E)^\circ)$, where X° is the \mathbb{Z} - torsion of the module X . Let $a \in A_T$, $n > m$ and $\Omega_n \in a_n, q_n$ be like above. Then $\mathbb{L}_n := \mathbb{K}_n[q^{1/p^n}]$ is a p - abelian p - ramified extension of degree p^n . If \mathbb{T}/\mathbb{M}_E is finite, $\mathbb{L}_n \subset \mathbb{M}_E$ and there is a unit $e \in E_n$ such that $\mathbb{L}_n = \mathbb{K}_n[e^{1/p^n}]$ and by Kummer theory, $q = e^c \cdot x^{p^n}$ with $(c, p) = 1$ and $x \in \mathbb{K}_n$. But then $\Omega^{p^{n+k}} = (q) = (x)^{p^n}$ and $(\Omega^{p^k}/(x))^{p^n} = (1)$. It follows that a_n has bounded order, in contradiction with the assumption. Thus \mathbb{T}/\mathbb{M}_E is infinite.

If \mathbb{T}/\mathbb{M}_E is infinite, by definition of \mathbb{T} it follows that A_T cannot be finite. This completes the proof. □

The Proposition 1 readily implies:

Lemma 4. *Notations being like above and with $\Delta = \text{Gal}(\mathbb{M}^+/\mathbb{K}_\infty^+)$, there is an isomorphism κ of $\mathbb{Z}_p[G]$ - modules*

$$(13) \quad \kappa : \Delta \rightarrow A_{T^*}^\bullet$$

and $\varphi(A_T) \subset \Delta$.

Proof. The relation (13) was proved in Proposition 1 and the inclusion $\varphi(A_T) \subset \Delta$ follows from the fact that $\varphi(A_T) = \text{Gal}(\mathbb{H}_T^+/\mathbb{K}_\infty^+)$ and we have shown that $\mathbb{H}_T^+ \subset \mathbb{M}^+$. Here φ is the Artin symbol $A \rightarrow \text{Gal}(\mathbb{H}_\infty/\mathbb{K}_\infty)$. □

The next result is

Lemma 5. *Let $\mathbb{T} = \mathbb{M}_E[A_T^{1/p^\infty}]$. Then $\mathcal{F}_1 = \text{Gal}(\mathbb{T} \cap \mathbb{H}_\infty/\mathbb{K}_\infty)$ is finite.*

Proof. By reflection, there is a subgroup $A' \subset A^-$ such that $\varphi(A^-) \cong \mathcal{F}_1$. If \mathcal{F}_1 is infinite, then $A' \subset A_{T^*}$ is infinite; but we have shown in the proof of the Proposition 1 that $\mathbb{H}_{T^*} \subset \mathbb{M}_E$, so $\mathbb{T} \cap \mathbb{H}_\infty \subset \mathbb{M}_E$, thus \mathcal{F}_1 is finite by the same argument used in the proof of Lemma 3. □

Finally we show

Lemma 6. *Notations being like in Lemma 5, $\text{Gal}(\mathbb{T}/\mathbb{M}_E)$ is finite.*

Proof. By Lemma 5, if \mathbb{T}/\mathbb{M}_E is infinite, then it is ramified, so $\mathbb{T} \subset \mathbb{M}^-$. Here we apply Theorem 4: since $\mathbb{M}^- \subset \mathbb{M}_E$ it follows a fortiori that $\mathbb{T} \subset \mathbb{M}_E$. □

The Theorem 1 now follows:

Proof. By Lemma 3, if A_T is infinite, then \mathbb{T}/\mathbb{M}_E is infinite. However we have shown in the last three lemmata that \mathbb{T}/\mathbb{M}_E is finite, and thus so must be A_T . □

Remark 1. *The proofs above show in fact that $A_T = \{1\}$. This comes from the definition of A_T which is not a subgroup of A but a system of representants for A/A° . Therefore our proof has no information about the torsion part of $A/(A^T)$.*

In presence of capitulation, thus for $b \in B$ of finite order, it is known that $\mathbb{K}_\infty[b_n^{1/p^n}] \subset \mathbb{M}_E$ for all $n > 0$ (e.g. [6], proof of Theorem 2.3, Chapter 6). The proof above shows that this cannot be the case if the order of b is not finite.

Greenberg mentions in [4] that Leopoldt's conjecture implies Theorem 1. We have shown here that they are in fact equivalent for totally real extensions in which p is totally split. The question whether the equivalence holds for arbitrary fields could not be solved in this paper. Therefore we give below a different approach for the general case. This involves a direct investigation of the L -phantom fields and yields a proof of Leopoldt's conjecture.

5. The Phantom Field

Let \mathbb{K}/\mathbb{Q} be a galois extension containing the p -th roots of unity and assume that $\mathcal{D}(\mathbb{K}) > 0$. We let (ξ_0, δ, α) be a fixed presentation of \bar{E} .

The following propositions explain the relation between the L -phantom field Φ and the Leopoldt defect and lay the ground for the proof of the conjecture.

Proposition 1. *There is a canonic field $\Phi_* = \mathbb{K}_\infty[A_{T^*}^{1/p^\infty}]$ with*

$$\mathbb{M} = \mathbb{M}_E \cdot \Phi_*$$

and such that

$$\Delta_h = \text{Gal}(\Phi_*/\mathbb{K}_\infty) \cong \text{Gal}\left(\mathbb{M}_E[A_{T^*}^{1/p^\infty}]/\mathbb{M}_E\right)$$

and Δ_h is a free \mathbb{Z}_p -module of \mathbb{Z}_p -rank $\mathcal{D}(\mathbb{K})$. Also, Δ_h is cyclic as a $\mathbb{Z}_p[G]$ -module and $\Delta_h \hookrightarrow \Delta$. In particular Δ_h is trivial if and only if Leopoldt's conjecture is true.

Proof. This follows by Kummer duality. Let \mathbb{M}_0 be defined by Lemma 1 and $B_m \subset \mathbb{K}_m^\times$ be the Kummer radicals which generate the maximal subfields $\mathbb{L}_m \subset \mathbb{K}_m \cdot \mathbb{M}_0$ with galois groups of exponent p^m over \mathbb{K}_m , for m sufficiently large. A Kummer radical for \mathbb{L}_m is a subgroup $(\mathbb{K}_m^\times)^{p^m} \subset B_m \subset \mathbb{K}_m^\times$ with $\mathbb{L}_m = \mathbb{K}_m[B_m^{1/p^m}]$. Two radicals B_m, B'_m are equivalent if $(B_m/\mathbb{K}_m^\times)^{p^m} = (B'_m/\mathbb{K}_m^\times)^{p^m}$. The Fact 1 together with reflection imply that $(B_m/\mathbb{K}_m^\times)^{p^m}$ is built up of either units or powers of ideals in A_{T^*} . Therefore $\mathbb{L}_m \subset \mathbb{M}_{E,m}[A_{T^*}^{1/p^m}]$, and taking inductive limits we find

$$\mathbb{M} \subset \mathbb{M}_E[A_{T^*}^{1/p^\infty}].$$

Let $\overline{\mathbb{M}}_E = \mathbb{M}_E[A_{T^*}^{1/p^\infty}]$, so $\mathbb{M} \subset \overline{\mathbb{M}}_E$ and Theorem 4 implies by comparing ranks, that $\mathbb{Z}_p - \text{rank}(\Delta_h) = \mathcal{D}(\mathbb{K})$. If the field \mathbb{K} is CM, we have the simple result

$$\mathbb{K}_\infty[(A_{T^*})^{1/p^\infty}] = \mathbb{M}(\mathbb{K}^+).$$

Indeed, in this case Theorem 4 implies that $\mathbb{M}^- = (\mathbb{M}_E \cap \mathbb{M})^-$ and thus, by complex conjugation, $\mathbb{M}^+ = \mathbb{M}(\mathbb{K}^+) \subset \mathbb{M}_E[A_{T^*}^{1/p^\infty}]$; since the defect is situated in the plus part, $A_{T^*} \subset A^-$ and

$$\overline{\mathbb{M}}_E = \Phi_* = \mathbb{K}_\infty[A_{T^*}^{1/p^\infty}].$$

We treat now the general case, when \mathbb{K} is not CM. The group $\Delta_f = \text{Gal}(\mathbb{F}/\mathbb{K}_\infty)$ is a $\mathbb{Z}_p[G]$ - module by construction of \mathbb{F} . We have seen that $\mathbb{M} \subset \overline{\mathbb{M}}_E$ and $\Delta_h = \text{Gal}(\overline{\mathbb{M}}_E/\mathbb{M}_E)$ has $\mathbb{Z}_p - \text{rank } \mathcal{D}(K)$. Since $\overline{\mathbb{M}}_E/\mathbb{K}_\infty$ is the injective limit of Kummer extensions of \mathbb{K}_m , it is an abelian extension with group $\overline{\Delta}_E$. Let $\Delta_E \subset \overline{\Delta}_E$ be the subgroup fixing \mathbb{M} ; then

$$\mathbb{M}_E^{\Delta_E} = \mathbb{M}_E \cap \mathbb{M} = \mathbb{F},$$

by construction of \mathbb{F} and $\text{Gal}(\mathbb{M}/\mathbb{F}) \cong \text{Gal}(\overline{\mathbb{M}}_E/\mathbb{M}_E) = \Delta_h$. Kummer duality implies that $\mathbb{M} = \mathbb{F}[A_{T^*}^{1/p^\infty}]$, so the field on the right hand side is well defined. We now use the fact that Δ_f - is a cyclic $\mathbb{Z}_p[G]$ - module of $\mathbb{Z}_p - \text{rank } r_2$. Let $v_0 = \varphi(\xi_0) \in \Delta$ and $v := v_0|_{\Delta_f} \in \Delta_f$, with φ the global Artin symbol: then v generates the $\mathbb{Z}_p[G]$ - module Δ_f and we let $\alpha \in \mathbb{Q}_p[G]$ be an idempotent generating $\hat{v}^\top \subset \mathbb{Q}_p[G]$; in particular, the group $\Delta'_f = \varphi\left(v_0^{(1-\alpha)\mathbb{Z}_p[G]}\right) \subset \Delta$ is a lift of Δ_f to Δ . The field $\Phi_* = \mathbb{M}^{\Delta'_f}$ has the desired properties. Indeed, by definition $\Phi_* \cdot \mathbb{F} = \mathbb{M}$ and $[\Phi_* \cap \mathbb{F} : \mathbb{K}_\infty] < \infty$; the group

$$\text{Gal}(\Phi_*/\mathbb{K}_\infty) \cong \text{Gal}(\mathbb{M}/\mathbb{F}) \cong \text{Gal}(\overline{\mathbb{M}}_E/\mathbb{M}_E) = \Delta_h$$

is by construction the complement of $\Delta'_f = (1 - \alpha)\Delta$, so $\Delta_h = \alpha\Delta$ is a cyclic $\mathbb{Z}_p[G]$ - module. This completes that proof. \square

The field Φ_* is dual to Φ by Kummer pairing, which motivates the following

Definition 3. *The field Φ_* is the dual L - phantom field. If $\alpha \in \mathbb{Q}_p[G]$ is the idempotent defined in the proof above, we let*

$$(14) \quad R(\mathbb{K}) = \alpha\mathbb{Q}_p[G], \quad R^\top(\mathbb{K}) = (1 - \alpha)\mathbb{Q}_p[G],$$

thus $\mathbb{Q}_p[G] = R^\top \oplus R$ and $\Delta_h = \text{Gal}(\Phi_*/\mathbb{K}_\infty) \cong \Delta^R$.

The second proposition investigates the Kummer radicals of Φ and relates the galois groups of the two phantom fields to the annihilator θ of δ .

Proposition 2. *There is an isomorphism $\Delta_d = \text{Gal}(\Phi/\mathbb{K}_\infty) \cong \Delta_h^\bullet$ and an increasing chain of Kummer radicals $D_m \subset E(\mathbb{K})$ such that the intermediate fields of Φ are $\Phi_m = \mathbb{K}_m[D_m^{1/p^{m+1}}]$. In particular $\theta R \cong \Delta_h$.*

Proof. We have shown that $\Phi = \mathbb{H}_{T^*} \subset \mathbb{M}_E$ and $\Delta_d = \text{Gal}(\Phi/\mathbb{K}_\infty)$ is a free \mathbb{Z}_p -module of rank $\mathcal{D}(\mathbb{K})$ which is cyclic as a $\mathbb{Z}_p[G]$ -module. Since $\text{Gal}(\Phi/\mathbb{K}_\infty) \cong A_{T^*}$ via the Artin symbol, while $\text{Gal}(\mathbb{H}_*/\mathbb{K}_\infty)^\bullet \cong A_{T^*}$ by Kummer duality, the claim $\Delta_d \cong \Delta_h^\bullet$ follows. Let $\Phi_m \supset \mathbb{K}_m$ be the maximal subfields of $\Phi = \mathbb{H}_{T^*}$ with galois group of exponent p^{m+1} over \mathbb{K}_m . We show that the Kummer radicals of these extensions are units related to the annihilation of $\delta \in \bar{E}$.

Let $\delta, \alpha, \theta, R$ be defined like above. Thus $\text{rank}(\theta R) = \mathcal{D}(\mathbb{K})$; let $k = p^i - 1$ be such that $\delta^k \equiv 1 \pmod{p\mathcal{O}(\mathbb{K})}$. Then $\theta\mathbb{Q}_p[G]$ is a $\mathbb{Q}_p[G]$ module of rank $\mathcal{D}(\mathbb{K})$ and we may choose $H \subset G$ a set such that $\{\theta\sigma : \sigma \in H\}$ is a base for this vector space. Let $\theta_m \in \mathbb{Z}[G]$ be approximations of θ to the p^{m+1} -th order and $d_m = \delta^{k\theta_m} \in E(\mathbb{K}) \setminus E(\mathbb{K})^p$, while $\iota(d_m) \in U^{p^{m+1}}$. Let $D_m = \text{Span}(\{d_m^\sigma : \sigma \in H\})$, where the span is taken over \mathbb{Z} ; the p -rank is $p - \text{rank}(D) = \text{rank}(D/D^p) = \text{rank}(\text{Span}(\bar{\delta}^\sigma)) = \mathcal{D}(\mathbb{K})$ – here $\bar{\delta}$ is the image of δ in E/E^p . It follows that $\mathbb{K}_m[D_m^{1/p^{m+1}}]$ is an unramified extension of \mathbb{K}_m contained in Φ_m and all the maximal cyclic subextensions $\mathbb{K}_m \subset \mathbb{L}_m \subset \mathbb{K}_m[D_m^{1/p^{m+1}}]$ have the degree p^{m+1} . Note also that $\mathbb{K}_m[D_m^{1/p^{m+1}}] \subset \mathbb{K}_{m+1}[D_{m+1}^{1/p^{m+2}}]$.

Since $\text{Gal}(\Phi_m/\mathbb{K}_m)$ and $\text{Gal}(\mathbb{K}_m[D_m^{1/p^{m+1}}]/\mathbb{K}_m)$ have the same p -rank and are both products of $\mathcal{D}(\mathbb{K})$ copies of $C_{p^{m+1}}$, while $\mathbb{K}_m[D_m^{1/p^{m+1}}] \subseteq \Phi_m$, it follows that the two fields are equal and $D_m/D_m^{p^{m+1}}$ is the Kummer radical of Φ_m , which completes the proof. \square

We note for future reference the following fact which follows from the construction of D_n :

$$(15) \quad \dim \left(D_n^{p^n} \cdot E^{p^{n+1}} / E^{p^{n+1}} \right) = \mathcal{D}(\mathbb{K}) = p - \text{rank}(D_n/D_n^{p^{n+1}}).$$

Equivalently, $\text{Gal}(\Phi_n/\mathbb{K}_n) \cong (\mathbb{Z}/(p^{n+1} \cdot \mathbb{Z}))^{\mathbb{K}}$. Furthermore, there is a submodule $A' \subset A$ such that the elements of a have infinite order and for each $n > n_0$, we have the isomorphism $\varphi(A'_n) \cong \text{Gal}(\Phi_n^*/\mathbb{K}_n)$ via the Artin symbol. Furthermore A'_n is a product of $\mathcal{D}(\mathbb{K})$ disjoint cyclic groups of order p^{n+1} , being isomorphic to the group D_n , which has this property by construction.

We also assume from now on that $k = 1$. It is proved in the second paper of these Proceedings, that if Leopoldt's conjecture holds for a field \mathbb{K} , then it holds also for solvable – and in particular abelian – extensions thereof. Conversely, if it does not

hold for \mathbb{K}_n , then it does not hold for \mathbb{K} either, so the assumption $k = 1$ is not a restriction of generality.

5.1. Capitulation in the Phantom field. Since $D_m \subset \mathbb{K}_0$ for all $m > 0$ and $D_{m+n} \subset D_m \cdot E^{p^{m+1}}$, we may construct arbitrarily large sequences of iterated class fields over \mathbb{K}_m , which are albeit not abelian over \mathbb{K}_m . Indeed:

Lemma 7. *Notations being like above, for each $m > 0$, each cyclic subextension $\mathbb{L}_m \subset \Phi_m^*$ and arbitrarily large $M > m$ there are telescoping sequences of extensions $\mathbb{K}_m \subset \mathbb{F}_1 \subset \mathbb{F}_2 \subset \dots \subset \mathbb{F}_M$, such that*

1. *The degrees $[F_i : F_{i-1}] = p, M \geq i \geq 1$, where $F_0 = \mathbb{K}_m$ and \mathbb{L}_m is one of the extension F_i .*
2. *For each $i \leq M - (m + k)$, the extension F_{i+m+k}/F_i is cyclic abelian.*
3. *If $a \in A_m^-$ is a class such that $\varphi(a)$ generates $\text{Gal}(\mathbb{L}/\mathbb{K}_m)$, then the primes $\varphi \in a$ are inert in F_M/\mathbb{K}_m and do not capitulate.*
4. *For each $r = p^i \leq p^{m+1}$, the primes in $a_m^{p^i}$ are totally split in F_i/\mathbb{K}_m . If $\mathfrak{q} \in a_m^{p^i}$ and $\Omega \subset F_i$ is a prime above it, then $\Omega \sim \varphi \mathcal{O}(F_i)$. Furthermore, Ω is inert and does not capitulate in F_M/F_i .*

Proof. Let $d_m = \delta^{(c\theta)_m} \in D_m, c \in \mathbb{Z}_p$ be such that $\mathbb{L} = \mathbb{K}_m[d_m^{1/p^{m+1}}]$ and let $d_M = \delta^{(c\theta)_M}$ and F_M be an embedding of $\mathbb{K}_m[d_M^{1/p^M}]$. Then we define $F_j \subset F_M$ as the subfields of degree p^j over \mathbb{K}_m . This implies the properties 1. and 2. Indeed, $\mathbb{L} = F_{m+1}$, the relative degree are p and since $d_M \in U^{p^M}$, the extension F_M/\mathbb{K}_m is unramified, while every successive extension of degree at most p^{m+1} is also Kummer abelian.

Recall that if $\mathbf{K} \supset \mathbf{L} \supset \mathbf{k}$ is an extension tower, such that \mathbf{K}/\mathbf{k} is cyclic and $\varphi \in \mathcal{O}(\mathbf{k})$ is a prime which does not ramify in \mathbf{K} , then if φ is inert in \mathbf{L}/\mathbf{k} then it is inert in \mathbf{K}/\mathbf{k} . Therefore, if a_n is a class like in the hypothesis of the Lemma and $\varphi \in a_n$, then φ is inert in \mathbb{L}/\mathbb{K}_m . For $j < m$, we define $\varphi_j = \varphi \mathcal{O}(F_j)$ the lift of φ to F_j , which is thus a prime ideal. Its Artin symbol $\left(\frac{F_{m+j}/F_j}{\varphi_j}\right) \Big|_{\mathbb{L}} = \left(\frac{\mathbb{L}/F_j}{\varphi}\right)$ and the previous observation implies that φ_j must be inert in F_{m+j}/F_j and thus φ is inert in F_{m+j}/\mathbb{L}_m . In particular, φ does not capitulate. Repeating the argument inductively, we obtain the claim of 3.

The field F_i is the fixed field of the Artin symbol of the primes in $a_m^{p^i}$ so these primes are totally split in F_i and inert in \mathbb{L}/F_i . By using the same argument as for φ above, we conclude that a prime $\Omega \subset \mathcal{O}(F_i)$ as defined in 4. will be inert in F_M/F_i and does not capitulate. In particular, it must have maximal order p^{m+1} (recall

that this order is bounded solely by the order of the p -power roots of unity in \mathbb{K}_m . Since $\varphi(\Omega)$ and $\varphi(\wp_i)$ generate $\text{Gal}(\mathbb{F}_{m+i}/\mathbb{F}_i)$, these symbols differ by a power c , $(c, p) = 1$. Using the fact that $\Omega \cap \mathcal{O}(\mathbb{K}) = \mathfrak{q}$ and $\mathfrak{q}^{p^i} \sim \wp$, we conclude that $c = 1$, which completes the proof. \square

The properties of these telescoping extensions are directly connected to the particular situation in which the Kummer radicals of all the extensions Φ_n^* are subsets of \mathbb{K}_0 . They imply the remarkable fact, that ideals of A_{T^*} will only capitulate in subextensions of the Hilbert class field in which they are totally split.

We now combine this observation with the construction of Thaine. Let $a_n \in A'_n \setminus A_n^p$ and $\mathfrak{q} \in a_n$ a prime with the following properties:

1. The prime $q \in \mathbb{Z}$ below \mathfrak{q} splits completely in \mathbb{K}_N/\mathbb{Q} for some $N > n$ and is unramified in \mathbb{K} .
2. Let ζ_q be a primitive q -th root of unity and $\mathbb{F}_n = \mathbb{K}_n[\eta] \subset \mathbb{K}_n[\zeta_q]$ be the field of degree $[\mathbb{F}_n : \mathbb{K}_n] = p^N$. Then \mathfrak{q} is totally ramified in $\mathbb{F}_n/\mathbb{K}_n$: if $\Omega \subset \mathbb{F}_n$ is the prime above \mathfrak{q} , then $\mathfrak{q} = \Omega^{p^N}$ and we let $a'_n = [\Omega]$ be the p -part of the class of Ω .

We shall assume that

$$(16) \quad \text{ord}(a'_n) > \text{ord}(a_n),$$

Observe that $a'_n \in A(\mathbb{F}_n) \setminus A^p(\mathbb{F}_n)$, as a consequence of the fact that a_n is p -indivisible. If \mathbb{K} is CM and galois, then $a_n \in A^-$ and \mathfrak{q} will not capitulate in the extension $\mathbb{F}_n/\mathbb{K}_n$, which is also CM, thus (16) holds. Based on this relation, we use the construction of Thaine sketched in point 2. and prove:

Lemma 8. *If $f_{\mathcal{D}}(\mathbb{K}) > 0$, there is no extension \mathbb{F}_n verifying the premises above and such that $\text{ord}(a'_n) > \text{ord}(a_n)$.*

Proof. We assume that $\text{ord}(a'_n) = r \text{ord}(a_n)$ for some $r = p^j$, $j \geq 1$ and n sufficiently large. We shall prove that $a'_n{}^{T^*} = 1$. This implies the claim. Indeed, if $\mathbb{L} \subset \mathbb{H}_n(\mathbb{F})$ is a cyclic extension with group generated by $\varphi(a_n)$, then $\mathbb{L} \cdot \mathbb{K}_{n+j} \subset \Phi_{n+j}^*[\eta]$ is a Kummer extension of maximal degree p^{n+1+j} . But we have seen that these extensions are not abelian over \mathbb{K}_n , and thus \mathbb{L}/\mathbb{F}_n cannot be abelian: a contradiction of the assumption $\text{ord}(a'_n) > \text{ord}(a_n)$ (or equivalently, to $j > 0$).

Since q is unramified in \mathbb{K} , \mathbb{K} and $\mathbb{Q}[\eta]$ are linearly disjoint and we let $\mathbb{F}_m = \mathbb{K}_m[\eta]$ for all $m \geq 0$. Thus $\text{Gal}(\mathbb{F}_m/\mathbb{Q}) = \text{Gal}(\mathbb{K}_m/\mathbb{Q}) \times \text{Gal}(\mathbb{Q}[\eta]/\mathbb{Q})$ and $\cup_m \mathbb{F}_m$ is the cyclotomic \mathbb{Z}_p -extension of $\mathbb{K}[\eta]$. The group A'_n is $\mathbb{Z}_p[G]$ -cyclic and we assume

that a_n is a generator. Let $A'_n(\mathbb{F}) = \mathbf{N}_{\mathbb{F}_n/\mathbb{K}_n}^{-1}(A'_n)$ be the inverse image of norm for A'_n . Since Ω is totally ramified, $a'^{v-1} = 1$ for v a generator of $\text{Gal}(\mathbb{F}_n/\mathbb{K}_n)$; the group A'_n being $\mathbb{Z}_p[G]$ cyclic, it follows that $A'_n(\mathbb{F})$ has the same p -rank $\mathcal{D}(\mathbb{K})$ as A'_n , and a'_n is a generator of this $\mathbb{Z}_p[G]$ -module. We let $\tilde{\Phi}_n^* = \Phi_n^* \cdot \mathbb{F}_n \subset \mathbb{H}_n(\mathbb{F})$, the Hilbert class field of \mathbb{F}_n .

Let $a' = (a'_m) \in A(\mathbb{F})$ be a norm coherent sequence containing a'_n . Then $a' \notin A(\mathbb{F})^p$. Consider now $\wp \in a'_n$ a prime ideal with $\wp^r \sim \Omega$ and let $\mathbb{L}/\mathbb{F}_n, \mathbb{L} \subset \mathbb{H}_n(\mathbb{F})$ an unramified cyclic extension with $[\mathbb{L} : \mathbb{F}_n] = \text{ord}(a'_n)$ in which \wp is inert. Then $\mathbb{M}_1 = KL \cap \tilde{\Phi}_n^*$ is an extension of degree p^{n+1} with galois group generated by $\varphi(a'_n)$ and thus $\varphi(a_n)$ fixes some subfield $\mathbb{M} \subset \mathbb{L}$ with $[\mathbb{M} : \mathbb{F}_n] = r$, and for $p^{n+1} > r$ we have $\mathbb{M} \subset \mathbb{M}_1$. Now $\mathbb{L} \cdot \mathbb{F}_{n+j} \subset \Phi_{n+j}^*[\eta]$ and thus $\text{Gal}(\mathbb{L} \cdot \mathbb{F}_{n+j}/\mathbb{F}_{n+j})^{T^*} = \{1\}$; let $\wp \in a'_n$ be a prime splitting in $\mathbb{F}_{n+j}/\mathbb{F}_n$, and $\wp' \subset \mathcal{O}(\mathbb{F}_{n+j})$ a prime above it, so

$$\left(\frac{\mathbb{L} \cdot \mathbb{F}_{n+j}/\mathbb{F}_{n+j}}{\wp'} \right) = \left(\frac{\mathbb{L}/\mathbb{F}_n}{\wp} \right),$$

by restriction (cf. [6], Chapter X, p. 198, **A2**). But the Artin symbol on the left is canceled by T^* , and thus $a'^{T^*} = 1$, as claimed. □

As a direct consequence, since $\text{ord}(a'_n) > \text{ord}(a_n)$ holds for CM fields, we have:

Corollary 2. *Leopoldt's conjecture holds for CM galois extensions \mathbb{K}/\mathbb{Q} .*

If \mathbb{K}/\mathbb{Q} is galois but not CM, we need the following

Assumption 1. *Let $\alpha \in \mathbb{Z}_p[G]$ be any idempotent and $\alpha^* \in \mathbb{Z}_p[G]$ be its image through the Leopoldt involution. Then we assume that the units $E \subset \mathbb{K}$ are such that for each idempotent and each $n > 0$*

$$\begin{aligned} \alpha(\mathbb{Z}[G]/(p^n\mathbb{Z}[G])) \cdot (E/p^n E) &= \{1\} \quad \text{or} \\ \alpha^*(\mathbb{Z}[G]/(p^n\mathbb{Z}[G])) \cdot (E/p^n E) &= \{1\}. \end{aligned}$$

Then one can show that assumption (16) and thus Leopoldt's conjecture follows for \mathbb{K} .

Acknowledgments: I take full responsibility for the proofs presented here. However, without the valuable support received in the last months, which helped test and sort out a variety of arguments, thus leading to the approaches presented in these Proceedings, this work might still have been on the work bench for a long period of time.

I am grateful to Hendrik Lenstra for a stimulating email exchange which helped bring to the surface the central requirements for applying Baker theory. I thank

Ehud De Shallit for his careful proof reading of an earlier version of this paper and suggestive examples, which signaled the need of Baker theory for bootstrapping the proof. I am indebted to Oliver Bräunling, Laurent Bartholdi, Filippo Nuccio and Francisco Pappalardi for fruitful help on preliminary versions of the draft and attentive discussion on the algebraic aspects of noncommutative group rings. Many final insights and clarifications grew from discussions with Bruno Anglès and David Vauclair in Caen. I also thank the first for pointing me out Waldschmidt's result.

I thank all my colleagues in Göttingen for their support and motivating discussions; my gratitude goes in particular to Hans-Christian Graf von Bothmer, Christian Böhning, Ina Kersten, Thomas Schick, Ulrich Stuhler, Victor Vuletescu and Ingo Witt. I am most grateful to Samuel Patterson who followed during a seminar in summer 2008 the development of some of the ideas presented here, with his attention, remarks and advice.

Last but not least I thank Theres and Seraina for the innocent patience with which they dealt with my ambig presence over longer periods of time.

References

- [1] J. ALPERIN & R. BELL – *Groups and representations*, Graduate Texts in Mathematics, vol. 162, Springer, 1995.
- [2] B. ANGLÈS – “On the p -adic Leopoldt transformation of a power series”, *Acta Arithmetica* **134** (2008), no. 4, p. 349–368.
- [3] A. BRUMER – “On the units of algebraic number fields”, *Mathematika* **14** (1967), p. 121–124.
- [4] R. GREENBERG – “On the Iwasawa invariants of totally real fields”, *American Journal of Mathematics* **98** (1973), p. 263–284.
- [5] M. LAURENT – “Rang p -adique d'unités et action de groupes”, *J. reine angew. Math.* **399** (1989), p. 81–108.
- [6] S. LANG – *Cyclotomic fields I and II*, combined second edition ed., Graduate Texts in Mathematics, vol. 121, Springer, 1990.
- [7] J. SERRE – “Local class field theory”, *Algebraic Number Theory* (Cassels & Fröhlich, eds.), Academic Press, 1967, p. 129–161.
- [8] L. WASHINGTON – *Introduction to cyclotomic fields*, Graduate Texts in Mathematics, vol. 83, Springer, 1996.

HOMOLOGICAL ALGEBRA FOR SCHWARTZ ALGEBRAS

Ralf Meyer

Mathematisches Institut, Georg-August Universität Göttingen
E-mail: rameyer@uni-math.gwdg.de

Abstract. Let G be a reductive group over a non-Archimedean local field. For two tempered smooth representations, it makes no difference for the Ext-groups whether we work in the category of tempered smooth representations of G or of all smooth representations of G . Similar results hold for certain discrete groups. We explain the basic ideas from functional analysis and geometric group theory that are needed to state this result correctly and prove it.

1. Introduction

These notes are based on my lecture at the conference “Symmetries in Algebra and Number Theory” in Göttingen in October 2008, where I discussed results of [5, 7, 9]. Since details are available in these articles, our presentation will sometimes be informal and limited to the most basic ideas.

First I briefly introduce some categories of representations studied in representation theory, before focussing on the categories of smooth and tempered smooth representations of reductive p -adic groups.

It was observed for such groups that it makes no difference for homological algebra in which of these two categories we work: both $\mathrm{Ext}_G^*(V, W)$ and $\mathrm{Tor}_*^G(V, W)$ agree in both worlds if V and W are tempered smooth representations. Even more is true: the derived category of tempered smooth representations is a full subcategory of the derived category of smooth representations. All this can be deduced from

October 2008.

2000 Mathematics Subject Classification. 20G05, 18E30.

the exactness of a single chain complex, namely, the chain complex that computes $\mathrm{Tor}_*^{\mathbb{C}_c^\infty(G)}(\mathcal{S}(G), \mathcal{S}(G))$ for the Schwartz algebra $\mathcal{S}(G)$ of G , where $\mathbb{C}_c^\infty(G)$ denotes the Hecke algebra of G .

The result as stated above is false, however: to get a correct statement we must incorporate some functional analysis into our homological algebra in order to replace tensor products by complete tensor products. I explain this for the Schwartz algebra $\mathcal{S}(\mathbb{Z})$ for the group of integers, where the same problem appears. After the necessary excursion into bornological vector spaces and homological algebra for them, we can correctly state our main result.

Let A be a dense subalgebra of a bornological algebra B . By density, a map between two B -modules is B -linear once it is A -linear. Hence the category of B -modules is always a full subcategory of the category of A -modules. This becomes false when we pass to derived categories. But there are many cases where the canonical functor from the derived category of B -modules to the derived category of A -modules is fully faithful. I first met this phenomenon along the way in [8] and studied it more systematically in [6]. The same phenomenon has been studied under different names in slightly different contexts by other authors, as kindly pointed out to me by Alexei Yu. Pirkovskii and Henning Krause. The first instance I know is the notion of *absolute localisation* in Joseph L. Taylor's work on the functional calculus for several commuting operators on a Banach space ([14]); this notion is formulated for continuous homomorphisms of topological algebras. Another instance is the notion of a *homological epimorphism* introduced by Werner Geigle and Helmut Lenzing in [2]; this notion is formulated for homomorphisms of rings and has been applied to the representation theory of finite-dimensional algebras. Amnon Neeman and Andrew Ranicki call such homomorphisms *stably flat* and use them in connection with algebraic K-theory ([10], see also [4]). The same name is used by Alexei Yu. Pirkovskii in [11]. I call such maps *isocohomological* because they preserve cohomology.

The proof that the embedding $\mathbb{C}_c^\infty(G) \rightarrow \mathcal{S}(G)$ of the Hecke algebra into the Schwartz algebra of a reductive p -adic group is isocohomological is based on an idea from geometric group theory. To make this point, I also discuss a similar result for discrete groups from [5], which deals with the group ring $\mathbb{C}[G]$ of a finitely generated discrete group G and a Schwartz algebra $\mathcal{S}(G)$ defined by weighted ℓ_1 -estimates. It turns out that the chain complex whose contractibility is crucial for an isocohomological embedding $\mathbb{C}[G] \rightarrow \mathcal{S}(G)$ for a discrete group G is a coarse invariant of G .

This leads to a recipe for contracting it, based on the notion of a *combing* from geometric group theory.

Reductive p -adic groups have such a combing because they act cocompactly on a CAT(0)-space – their affine Bruhat–Tits building. This led me to prove the result for reductive p -adic groups. After I established results on isocohomological embeddings for Abelian groups in [8], I wanted to extend them to reductive p -adic groups such as $Sl_n(\mathbb{Q}_p)$ but found this difficult. Therefore, I first worked out a similar problem for discrete groups in [5], in a way that ought to generalise to reductive p -adic groups; then I carried out this generalisation in [7].

In these notes I only sketch some rough ideas of the proof of the result for discrete groups. After this sketch of a proof, I turn to some applications. The first is a vanishing result for certain Ext and Tor-groups for square-integrable representations. Together with results of Peter Schneider and Ulrich Stuhler from [12], this provides a combinatorial formula for the formal dimension of a square-integrable representation, which implies that these dimensions are quantised. As a consequence, the number of square-integrable irreducible representations that contain a U -invariant vector for a compact open subgroup U of G grows at most linearly in $\text{vol}(U)^{-1}$.

2. Categories of representations

Here we discuss some classes of representations that have been studied in representation theory. Later on, we will focus on smooth representations and tempered smooth representations.

The first class of representations to be studied were the finite-dimensional ones. Infinite-dimensional representations came into focus because of quantum mechanics, which required understanding unitary representations of certain Lie groups on Hilbert spaces.

For any locally compact group G , we know one basic example of a unitary representation: the regular representation on the Hilbert space $L_2(G)$, given by $g \cdot f(x) := f(g^{-1}x)$ for all $g, x \in G$, $f \in L_2(G)$. If G is a compact group, then any irreducible representation of G is contained in the regular representation. For non-compact groups, we must restrict to unitary representations – unitarity is automatic for compact groups – and weaken our notion of containment: already for the Abelian group \mathbb{R} , irreducible representations are only *weakly* contained in the regular representation. (A unitary representation π of G on a Hilbert space \mathcal{H} is weakly contained in another unitary representation ρ of G if its matrix coefficients $g \mapsto \langle \pi_g \vec{v}, \vec{w} \rangle$ for

$\vec{v}, \vec{w} \in \mathcal{H}$ can be approximated locally uniformly by linear combinations of matrix coefficients of ρ .)

All non-compact semi-simple Lie groups have unitary representations that are not weakly contained in the regular representations – the simplest example is the trivial representation of $\mathrm{Sl}(2, \mathbb{R})$. Unitary representations that are weakly contained in the regular representation are called *tempered* unitary representations. The systematic study of the representation theory of Lie groups showed that tempered unitary representations are much easier to classify than general unitary representations – we still lack a complete description of the latter for general semi-simple Lie groups, while the tempered unitary representations are, in principle, classified.

Lie algebra methods are one of the major tools for studying finite-dimensional representations of Lie groups. These do not directly apply to unitary representations on Hilbert space because the Lie algebra is only represented by unbounded operators. But they do apply nicely once we pass to a suitable subset of smooth vectors. Many results in representation theory are established first in this category of smooth representations and then translated to unitary Hilbert space representations.

This lecture mainly deals with reductive p -adic groups instead of Lie groups. A *reductive p -adic group* is a reductive linear algebraic group over a non-Archimedean local field. Their representation theory is remarkably similar to the representation theory of Lie groups. For the following, it suffices to think of basic examples of reductive p -adic groups such as the special linear groups $\mathrm{Sl}_n(\mathbb{Q}_p)$ or $\mathrm{Sl}_n(\mathbb{F}_q[[t^{-1}, t]])$ over the fields \mathbb{Q}_p of p -adic integers or over a local function field $\mathbb{F}_q[[t^{-1}, t]]$. All groups we consider are locally compact and totally disconnected, that is, their topology has a basis consisting of subsets that are both compact and open. Moreover, the unit element has a neighbourhood basis of compact open subgroups.

Definition 2.1. A representation of a locally compact, totally disconnected group (on a \mathbb{C} -vector space) is called *smooth* if each vector is fixed by some open subgroup. Let $\mathrm{Rep}(G)$ denote the category of smooth representations of G .

Example 2.2. Let $G = \mathrm{Sl}_2(\mathbb{Q}_p)$ and let V be the space of locally constant functions on the projective line $\mathbb{P}^1\mathbb{Q}_p$, equipped with the induced action of G . This is a smooth representation of G . The constant function is G -invariant, so that V contains a subrepresentation isomorphic to the trivial representation. The quotient $V/\mathbb{C} \cdot 1$ is a tempered, irreducible smooth representation called the *Steinberg representation* of G . The representation V itself is not tempered because the trivial representation of G is not tempered.

The definition of tempered smooth representations is more subtle. It may seem natural to require the existence of an invariant inner product such that the Hilbert space completion carries a tempered unitary representation. But it is better not to require tempered smooth representations to be unitary at all.

One reason for this is the following *desideratum*: the category of tempered smooth representations should be closed under extensions. In particular, a smooth representation of finite length (that is, one with a finite Jordan–Hölder series) should be tempered if all its irreducible subquotients are tempered. But the property of being unitary is not closed under extensions.

Example 2.3. Represent the group \mathbb{Z} on C^2 by unipotent matrices: $n \mapsto \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}$. This representation is a non-trivial extension of the trivial representation by itself and should therefore be tempered because the trivial representation of \mathbb{Z} is a tempered unitary representation. But a unipotent matrix is not unitary for any inner product.

The correct definition of tempered smooth representations uses Schwartz algebras. Smooth representations of a group G may be viewed as non-degenerate modules over the convolution algebra $C_c^\infty(G)$ of smooth, compactly supported functions on G . For totally disconnected groups, “smooth” means “locally constant,” and $C_c^\infty(G)$ is also called the *Hecke algebra* of G . For Lie groups, the above statement is only literally correct if we incorporate some functional analysis into our definitions – we should study smooth representations and $C_c^\infty(G)$ -modules in the category of bornological vector spaces (see [9]).

Tempered representations are defined as modules over a certain completion of $C_c^\infty(G)$, generically called *Schwartz algebra*. The cases $G = \mathbb{Z}$ and $G = \mathbb{R}$ are most familiar: here the Schwartz algebra $\mathcal{S}(G)$ is the convolution algebra of rapidly decreasing functions of Laurent Schwartz. Recall that the Fourier transform provides algebra isomorphisms between $\mathcal{S}(\mathbb{Z})$ and the algebra $C^\infty(\hat{\mathbb{Z}})$ of smooth functions on the Pontrjagin dual $\hat{\mathbb{Z}} \cong \mathbb{T}$ of \mathbb{Z} with pointwise multiplication, and between $\mathcal{S}(\mathbb{R})$ with convolution and $\mathcal{S}(\mathbb{R})$ with pointwise product. We will soon use $\mathcal{S}(\mathbb{Z})$ to motivate our results for reductive p -adic groups.

The definition of the Schwartz algebra $\mathcal{S}(G)$ for a reductive p -adic group G is due to Harish–Chandra and involves two ingredients: uniform smoothness and rapid decay. A function on G is *uniformly smooth* if it is U -invariant on the left and on the right for some compact-open subgroup U in G , so that it descends to a function on the double coset space $G // U$. A uniformly smooth function f on G has *rapid decay* if and only if $f \cdot (\ell + 1)^k \in L_2(G)$ for all $k \in \mathbb{N}$, where ℓ is an appropriate length

function on G ; for $G = \mathrm{Sl}_n(\mathbb{Q}_p)$, we may take

$$\ell(g) := \log_p \max\{\|g\|_\infty, \|g^{-1}\|_\infty\}$$

where $\|g\|_\infty$ denotes the maximum of the p -adic norms of the matrix entries of g . It is non-trivial that convolutions of uniformly smooth functions of rapid decay are well-defined and have rapid decay.

The above description of $\mathcal{S}(G)$ by L_2 -estimates is due to Marie-France Vignéras [15]. Harish–Chandra defines $\mathcal{S}(G)$ using a weighted supremum norm instead. For uniformly smooth functions, an L_2 -estimate is equivalent to suitable weighted L_p -estimates for $p \in [0, \infty]$ because the set $G//U$ of double cosets – unlike G itself or G/U – has polynomial growth with respect to the length function ℓ .

Roughly speaking, a tempered smooth representation of G is a module over the Schwartz algebra $\mathcal{S}(G)$. But we must modify this definition because as stated above, the derived category of $\mathcal{S}(G)$ does *not* embed into the derived category of $C_c^\infty(G)$.

Before we discuss this problem, we briefly define Schwartz algebras for finitely generated discrete groups. If ℓ is a word-length function on such a group G , we let $f \in \mathcal{S}(G)$ if $f \cdot (\ell + 1)^k \in \ell_1(G)$ for all $k \in \mathbb{N}$, that is,

$$\sum_{g \in G} |f(g)| (\ell(g) + 1)^k < \infty.$$

We use this definition although modules over $\mathcal{S}(G)$ have little to do with tempered unitary representations in general; they are more closely related to uniformly bounded Banach space representations. For groups of rapid decay, the Jolissaint algebra [3] is an interesting alternative to $\mathcal{S}(G)$ that is closely related to tempered unitary representations. But our main results are false for the Jolissaint algebra, that is, its derived category does *not* embed into the derived category of the group ring.

3. Why do we need bornological modules?

Our main result asserts that the derived category of tempered smooth representations of a reductive p -adic group is a full subcategory of the category of all its smooth representations. In particular, if V and W are both tempered smooth representations, then it makes no difference for $\mathrm{Ext}_G^*(V, W)$ and $\mathrm{Tor}_G^*(V, W)$ in which of the two categories of smooth representations we work.

In this section, we explain why this theorem is false and how to rectify it. The issue is that we cannot work in a purely algebraic setting; we must complete tensor products, forcing us to incorporate some functional analysis into our setup.

The problem is the same for reductive p -adic groups and Abelian groups and already appears for the group $G = \mathbb{Z}$. Hence we study this very simple case here. Let $A := \mathbb{C}[\mathbb{Z}]$ and $B := \mathcal{S}(\mathbb{Z})$. The basic input for all homological computations with A -modules is a free A -bimodule resolution of A : once we know such a resolution, we can get free resolutions for all left or right A -modules and compute derived functors.

The algebra A is isomorphic to the algebra $\mathbb{C}[t, t^{-1}]$ of Laurent polynomials. There is a very small free A -bimodule resolution of A ,

$$0 \rightarrow A \otimes A \xrightarrow{d} A \otimes A \xrightarrow{m} A \rightarrow 0,$$

where $d(f \otimes g) := tf \otimes g - f \otimes tg$ and $m(f \otimes g) = f \cdot g$ for $f, g \in \mathbb{C}[t, t^{-1}]$. It is routine to check that the above chain complex is exact.

The basic question is whether it remains exact when we replace A by the completion B . The resulting chain complex

$$(3.1) \quad 0 \rightarrow B \otimes B \xrightarrow{d} B \otimes B \xrightarrow{m} B \rightarrow 0$$

computes $\mathrm{Tor}_A^*(B, B)$ and should be exact if $\mathrm{Tor}_A^*(B, B) \cong \mathrm{Tor}_B^*(B, B)$. Conversely, if this single chain complex is exact, then $\mathrm{Ext}_A^n(V, W) \cong \mathrm{Ext}_B^n(V, W)$ for all B -modules V and W , similarly for Tor , and the canonical functor from the derived category of B to the derived category of A is fully faithful (see [6]).

Unfortunately, the chain complex in (3.1) is not exact. Clearly, d is injective and m is surjective, but $\ker m$ is bigger than the range of d . To remedy this, we have to complete our tensor products and replace $B \otimes B = \mathcal{S}(\mathbb{Z}) \otimes \mathcal{S}(\mathbb{Z})$ by $B \hat{\otimes} B := \mathcal{S}(\mathbb{Z} \times \mathbb{Z})$. This is isomorphic by the Fourier transform to $C^\infty(\mathbb{T}^2)$, and the completed chain complex

$$0 \rightarrow C^\infty(\mathbb{T}^2) \xrightarrow{d} C^\infty(\mathbb{T}^2) \xrightarrow{m} C^\infty(\mathbb{T}) \rightarrow 0$$

is exact. Here $df(x, y) := (x - y)f(x, y)$ and $mf(x) = f(x, x)$.

Thus we need to modify our setup and consider a category of modules where the algebraic tensor product is replaced by a completed tensor product with $B \hat{\otimes} B = \mathcal{S}(\mathbb{Z} \times \mathbb{Z})$.

The most familiar choice is the category of complete locally convex topological vector spaces with the complete projective topological tensor product. While this may still work well enough for the example \mathbb{Z} , we get serious problems for the Schwartz algebra $\mathcal{S}(G)$ of a reductive p -adic group G because the product in $\mathcal{S}(G)$ is not jointly continuous. To accomodate this, we may follow [1] and use the complete inductive topological tensor product, which is designed to be universal for separately continuous bilinear maps. But this tensor product behaves rather badly for general

locally convex topological vector spaces – even associativity is unclear. Everything works fine if we also restrict attention, say, to the category of nuclear LF-spaces. This is enough to cover $\mathcal{S}(G)$ and the modules we need.

There is, however, a better way to combine functional analysis with homological algebra, where the problems with completed tensor products mentioned above disappear. In other words, these problems are mere artefacts produced by unsuitable definitions. Instead of complete locally convex topological vector spaces, we should use complete convex bornological vector spaces. A bornology on a vector space is a collection of bounded subsets with certain properties – thus bounded subsets replace open subsets in bornological analysis. Correspondingly, boundedness replaces continuity for linear and bilinear maps. Any complete convex bornological vector space is an inductive limit of Banach spaces in a natural way.

The category of complete bornological vector spaces has very good algebraic properties. For instance, the complete projective tensor product in this category and the internal Hom functor are related by the familiar adjointness isomorphism

$$\mathrm{Hom}(X, \mathrm{Hom}(Y, Z)) \cong \mathrm{Hom}(X \hat{\otimes} Y, Z).$$

Any vector space carries a canonical bornology, called the *fine bornology*. Its bounded subsets are the bounded subsets of finite-dimensional subspaces. This defines a fully faithful, fully exact embedding of the category of vector spaces into the category of complete bornological vector spaces. Furthermore, the embedding is symmetric monoidal, that is, compatible with (complete) tensor products. Roughly speaking, nothing happens when we equip a vector space with the fine bornology.

Topological vector spaces also carry canonical bornologies. The most useful choice is the precompact bornology, consisting of all precompact subsets. This defines a fully faithful, fully exact, symmetric monoidal embedding of the category of Fréchet spaces into the category of complete bornological vector spaces. The precompact bornology also defines such an embedding on the category of nuclear LF-spaces. On this category, the complete projective bornological tensor product agrees with the complete inductive topological tensor product – exactly what we need. Thus $\mathcal{S}(G) \hat{\otimes} \mathcal{S}(G) \cong \mathcal{S}(G \times G)$ for any reductive p -adic group G .

We now modify our definitions to allow smooth representations on (complete, convex) bornological vector spaces; from now on, all bornological vector spaces are required complete and convex. A group representation of a locally compact, totally disconnected group G on a bornological vector space V is called *smooth* if, for each bounded subset S , there is a compact-open subgroup U of G with $u \cdot \bar{v} = \bar{v}$

for all $u \in U$, $\vec{v} \in S$. The category of smooth representations of G on bornological vector spaces is isomorphic to the category of essential bornological $C_c^\infty(G)$ -modules (a bornological A -module is essential if the multiplication map $A \hat{\otimes} M \rightarrow M$ is a bornological quotient map).

The *tempered smooth representations* of G are the essential $\mathcal{S}(G)$ -modules, where we equip $\mathcal{S}(G)$ with the obvious bornology: a subset of $\mathcal{S}(G)$ is bounded if its elements are uniformly smooth and of uniformly rapid decay.

When doing homological algebra in this setting, it is better to replace exactness by a stronger condition, namely, the existence of a bounded contracting homotopy. Otherwise, we would get a complicated derived category even for the trivial algebra \mathbb{C} because there are non-trivial extensions of bornological vector spaces. To get rid of these complications, we do *relative homological algebra* with respect to the category of bornological vector spaces. Formally, this means that we turn categories of bornological vector spaces and modules into *exact categories* in the sense of Quillen. These exact categories have enough injective and enough projective objects, so that homological algebra works as usual. In particular, we can form derived categories. The main point to remember is that resolutions are required to have a bounded contracting homotopy.

We can now state our main theorem:

Theorem 3.2. *The embedding $C_c^\infty(G) \rightarrow \mathcal{S}(G)$ induces a fully faithful functor between the derived categories of non-degenerate bornological modules over $C_c^\infty(G)$ and $\mathcal{S}(G)$. In particular, if both V and W are essential $\mathcal{S}(G)$ -modules, then*

$$\mathrm{Ext}_{C_c^\infty(G)}^n(V, W) \cong \mathrm{Ext}_{\mathcal{S}(G)}^n(V, W) \quad \text{and} \quad \mathrm{Tor}_n^{C_c^\infty(G)}(V, W) \cong \mathrm{Tor}_n^{\mathcal{S}(G)}(V, W).$$

Various equivalent conditions for such an embedding of derived categories are given in [6]. The one that is practical to check is the following:

Theorem 3.3. *Let A and B be bornological algebras and let $f: A \rightarrow B$ be an essential bounded algebra homomorphism; that is, B is essential as an A -bimodule. Let $P_\bullet \rightarrow A$ be a projective A -bimodule resolution of A . The functor between the derived categories of essential bornological modules over A and B induced by f is fully faithful if and only if $B \hat{\otimes}_A P_\bullet \hat{\otimes}_A B \rightarrow B$ has a bounded contracting homotopy.*

We also call f *isocohomological* if this is the case. The argument above shows that the embedding $\mathbb{C}[\mathbb{Z}] \rightarrow \mathcal{S}(\mathbb{Z})$ is isocohomological. Our main theorem states that the embedding $C_c^\infty(G) \rightarrow \mathcal{S}(G)$ for a reductive p -adic group is isocohomological.

The proof of Theorem 3.3 is not hard. Only the sufficiency of the condition is relevant here. The main point is that if $B \hat{\otimes}_A P \cdot \hat{\otimes}_A B \rightarrow B$ has a bounded contracting homotopy, then it is a projective B -bimodule resolution of B . Hence all derived functors for B can be computed using this bimodule resolutions. When we compare them with similar computations for A , we get the same answer because the two resolutions are so closely related.

4. Some ideas from the proof of the main theorem

We have now stated our main theorem correctly and established an analogous result for the discrete group \mathbb{Z} . To find a proof for reductive p -adic groups, we identify a geometric property for discrete groups related to non-negative curvature that ensures that the embedding $\mathbb{C}[G] \rightarrow \mathcal{S}(G)$ is isocohomological.

What does “geometric” mean here? Any finitely generated group G becomes a metric space with respect to a word-length function. This metric is not unique, but it is unique up to quasi-isometry. Moreover, it frequently happens that a group G is quasi-isometric to a nice geometric object like a smooth manifold. In general, if G acts cocompactly, properly, and by isometries on a metric space X , then G and X are quasi-isometric. For instance, the fundamental group of a Riemannian manifold is quasi-isometric to the universal covering of this manifold.

Any reductive p -adic group G acts cocompactly and properly on a nice geometric space – its affine Bruhat–Tits building. It is known that such buildings have non-positive curvature – formally, they have the CAT(0)-property that triangles in them are thinner than comparison triangles in flat Euclidean space. Hence a geometric non-positive curvature condition covers reductive p -adic groups as well.

It is shown in [5] that the question whether or not the embedding $\mathbb{C}[G] \rightarrow \mathcal{S}(G)$ is isocohomological for a discrete group G depends only on the quasi-isometry type of G . This can be formalised as follows: to any metric space we may associate a certain chain complex that is a quasi-isometry invariant, and for the underlying metric space of a finitely generated discrete group G , this chain complex has a bounded contracting homotopy if and only if the embedding $\mathbb{C}[G] \rightarrow \mathcal{S}(G)$ is isocohomological.

The chain complex in question is a certain completion of the reduced bar complex on X . More precisely, we take the reduced chain complex $C_\bullet(X)$ of the simplicial set with X^{n+1} as its set of n -simplices, and face and degeneracy maps deleting or adding one of the entries. The completion involves chains with controlled support

– there is $R > 0$ such that $f(x_0, \dots, x_n)$ vanishes if $d(x_i, x_j) > R$ for some i, j – and with rapid decay in the sense that the function $(\ell(x_0) + 1)^k f(x_0, \dots, x_n)$ on X^{n+1} is absolutely summable for each $k \in \mathbb{N}$. For a discrete group G , this chain complex has a bounded contracting homotopy if and only if the embedding $\mathbb{C}[G] \rightarrow \mathcal{S}(G)$ is isocohomological. Thus our task is to find a contracting homotopy for the uncompleted chain complex $C_\bullet(X)$ that is bounded with respect to the appropriate bornology and therefore extends to the completion.

The reduced bar complex above is huge and may therefore appear impractical for cohomology computations. Nevertheless, it is ideal for our current purpose because of its good functoriality properties. By design, the bar construction makes sense for any discrete set X , and any map $f: X \rightarrow Y$ induces a chain map. If $f, g: X \rightarrow Y$ are two maps, then the induced chain maps are chain homotopic – simply because the bar complex is contractible. But there is, in fact, an explicit and simple formula for a chain homotopy between the chain maps induced by f and g .

In particular, this chain homotopy for the identity map and a constant map provides a contracting homotopy for $C_\bullet(X)$. This is just the standard contracting homotopy of the reduced bar complex $C_\bullet(X)$, and it is unbounded for the relevant bornology. To get a bounded contracting homotopy, we must contract our group more gently: we need a sequence of maps $f_n: X \rightarrow X$ that, roughly speaking, move each $x \in X$ to the base point 0 in small steps; typically, we just let $f_n(x)$ be points on a quasi-geodesic in X from x to 0; “moving in small steps” means that there is $S > 0$ with $d(f_n(x), f_{n+1}(x)) < R$ for all $n \in \mathbb{N}$, $x \in X$.

Given such a sequence of maps (f_n) , we sum up the canonical chain homotopies between the chain maps induced by the f_n and hope that the sum remains bounded. This requires further geometric conditions. Since our homotopy must preserve controlled supports, we need that $f_n(x)$ and $f_n(y)$ remain close for all $n \in \mathbb{N}$ if x and y are close; more precisely, for each $R > 0$ there is $S > 0$ such that if $d(x, y) < R$, then $d(f_n(x), f_n(y)) < S$ for all $n \in \mathbb{N}$. A sequence of maps (f_n) with the two properties described above is called a (synchronous) *combing* on G . In addition, to preserve the rapid decay condition, we need that the number of $n \in \mathbb{N}$ with $f_n(x) \neq f_{n+1}(x)$ grows at most polynomially in $\ell(x) = d(x, 0)$; if the points $f_n(x)$ follow a quasi-geodesic, then this number automatically grows linearly (polynomial growth rules out some groups that only admit more complicated combings where each element follows a huge detour). We can now formulate the main result of [5]:

Theorem 4.1. *Let G be a finitely generated discrete group that has a combing of polynomial growth. Then the embedding $\mathbb{C}[G] \rightarrow \mathcal{S}(G)$ is isocohomological.*

In a Riemannian metric, geodesics remain close to each other if the curvature is non-positive. For instance, this happens in flat \mathbb{R}^n . Since the group \mathbb{Z}^n with word-length metric is quasi-isometric to \mathbb{R}^n , the group \mathbb{Z}^n has a combing as well. This explains why the embedding $\mathbb{C}[\mathbb{Z}^n] \rightarrow \mathcal{S}(\mathbb{Z}^n)$ is isocohomological. Another class of groups that admit combings are hyperbolic groups, fundamental groups of non-positively curved manifolds, and cocompact lattices in Lie groups – the latter act cocompactly and properly on non-positively curved Riemannian manifolds.

For discrete groups, categories of representations are usually not well-behaved, so that Ext- and Tor-groups for them are usually hard to compute. The most interesting case of Theorem 4.1 seems to concern the trivial representation. The groups $\text{Ext}_{\mathbb{C}[G]}^*(\mathbb{C}, \mathbb{C})$ are group cohomology, while $\text{Ext}_{\mathcal{S}(G)}^*(\mathbb{C}, \mathbb{C})$ are group cohomology with polynomial growth; that is, we take the standard bar complex computing group cohomology and take the cohomology of the subcomplex of cochains of polynomial growth. Theorem 4.1 implies that both chain complexes are homotopy equivalent if G has a combing of polynomial growth. In particular, every class in the group cohomology is represented by a cocycle of polynomial growth.

Any reductive p -adic group acts cocompactly on its affine Bruhat–Tits building, which is another example of a non-positively curved space – formally, a CAT(0)-space. Hence reductive p -adic groups have combings of linear growth. This suggested to me that Theorem 4.1 should remain true for reductive p -adic groups. In fact, I checked this for $\text{Sl}_2(\mathbb{Q}_p)$ by hand and added a remark to this extent in the introduction of [5]. This intrigued Peter Schneider who had tried to prove such a statement but hit the problem described in §3 and concluded that the statement was false.

As expected, the proof of Theorem 4.1 carries over to reductive p -adic groups. But two new problems appear that still requires a significant amount of additional work.

One issue is that the trivial representation of a reductive p -adic group is not tempered. In the discrete case, the trivial representation is always a module over $\mathcal{S}(G)$ because the latter is defined using ℓ_1 -estimates. This allows some simplification in the chain complexes to be considered. The chain complex described involves chains on X^n that are compactly supported in all but one direction. To treat reductive p -adic groups, we must allow functions that are compactly supported in all but *two* directions, and that satisfy a certain growth condition in the other two directions.

A second issue is the uniform smoothness in the definition of the Schwartz algebra. The contracting homotopy must be constructed more carefully in order to preserve this property, and the argument requires results of Bruhat and Tits about stabilisers of points in the building.

5. Some applications

In this section, we let G be a semi-simple p -adic group such as $\mathrm{Sl}_n(\mathbb{Q}_p)$ or such as $\mathrm{Sl}_n(\mathbb{F}_q[[t^{-1}, t]])$. An extension to reductive groups such as $\mathrm{Gl}_n(\mathbb{Q}_p)$ or $\mathrm{Gl}_n(\mathbb{F}_q[[t^{-1}, t]])$ is possible but requires more notation, which I do not want to introduce here.

What can we learn from our main result that the embedding $C_c^\infty(G) \rightarrow \mathcal{S}(G)$ is isocohomological? First of all, this implies that the subcategory of tempered smooth representations is closed under extensions in the category of all smooth representations because Ext^1 is the same in both categories. We already observed this important property of the category of tempered smooth representations in §2. Example 2.3 shows that the categories of unitary representations or of uniformly bounded Banach space representations are not closed under extensions. Hence the embedding $\mathbb{C}[\mathbb{Z}] \rightarrow \ell_1(\mathbb{Z})$ is not isocohomological.

Another important general consequence is that $\mathcal{S}(G)$ has a projective bimodule resolution of finite length – the same length as for $C_c^\infty(G)$. This means that derived functors for $\mathcal{S}(G)$ vanish above some dimension.

In principle, this projective bimodule resolution can be used to compute the Hochschild and cyclic homology of $\mathcal{S}(G)$. While such a computation for $C_c^\infty(G)$ is feasible, the case of $\mathcal{S}(G)$ is considerably more complicated because it requires careful estimates about the growth of the length function on conjugacy classes. It is known that $C_c^\infty(G)$ and $\mathcal{S}(G)$ have isomorphic periodic cyclic homology ([13]), but it seems very hard to prove this using the isocohomological embedding. There is no general theorem to this effect.

Square-integrable representations are special tempered representations that are isolated among tempered representations. That is, they are both projective and injective in the category of tempered smooth representations. Hence $\mathrm{Ext}_{\mathcal{S}(G)}^n(V, W)$ vanishes for $n \neq 0$ if V or W is square-integrable and the other one tempered. Our main theorem yields the same for $\mathrm{Ext}_{C_c^\infty(G)}^n(V, W)$. For instance, this applies if V is the Steinberg representation of $\mathrm{Sl}_2(\mathbb{Q}_p)$ (see Example 2.2). This vanishing result is remarkable because, by its very definition, the Steinberg representation V has a non-trivial extension by the trivial representation \mathbb{C} , so that $\mathrm{Ext}^1_{C_c^\infty(G)}(V, \mathbb{C}) \neq 0$.

Vanishing results for square-integrable representations are particularly important because these are the atoms of the Plancherel measure on the set of irreducible representations of G . Moreover, since $\text{Ext}_{C_c^\infty(G)}^n(V, W)$ vanishes unless V and W have the same central character, the support of the function $V \mapsto \dim \text{Ext}_{C_c^\infty(G)}^n(V, W)$ is always a finite set. Hence this function vanishes almost everywhere if W is tempered and $n \neq 0$. In contrast, the function $V \mapsto \dim \text{Ext}_{C_c^\infty(\text{Sl}_2 \mathbb{Q}_p)}(V, \mathbb{C})$ does not vanish almost everywhere because it is non-zero at the Steinberg representation, which is an atom of the Plancherel measure.

Another interesting application is the quantisation of formal dimensions of square-integrable representations:

Theorem 5.1. *There is $\alpha > 0$ such that the formal dimension of any irreducible square-integrable representation of G belongs to $\alpha \cdot \mathbb{N}_{\geq 1}$.*

This implies a bound on the number of irreducible square-integrable representations that contain U -fixed vectors for some compact-open subgroup U : any such representation is contained in $L_2(G/U)$, which has formal dimension $\text{vol}(U)^{-1}$. Since formal dimensions are additive and positive, we conclude that there are at most $1/(\alpha \text{vol } U)$ square-integrable representations that contain U -fixed vectors.

How is our main result related to formal dimensions? The following conceptual explanation is taken from [7]. The ring $C_c^\infty(G)$ is a regular Noetherian ring, that is, any finitely generated $C_c^\infty(G)$ -module has a resolution of finite length by finitely generated projective modules. Hence any finitely generated $C_c^\infty(G)$ -module determines a class in the algebraic K-theory of $C_c^\infty(G)$: take the Euler characteristic $\sum_{n=0}^\infty (-1)^n [P_n]$ of a finite type projective resolution P_\bullet . Since the formal dimension of representations defines a linear map $K_0(C_c^\infty G) \rightarrow \mathbb{R}$, this yields a notion of formal dimension for all finitely generated $C_c^\infty(G)$ -modules.

There seems to be no general formula for such a resolution for a general finitely generated module. But if we restrict to representations of finite length, then Peter Schneider and Ulrich Stuhler construct an explicit projective resolution in [12]. This can be used to compute the formal dimension mentioned above and shows that it is quantised. But it is not clear whether this new notion of formal dimension agrees with the usual one that is based on traces on group von Neumann algebras.

This is exactly where our main theorem is needed. It implies that a projective $C_c^\infty(G)$ -module resolution of an $\mathcal{S}(G)$ -module remains a resolution when we base change to $\mathcal{S}(G)$. If we start with a square-integrable representation V , then V is a projective $\mathcal{S}(G)$ -module. Hence the resolution $\mathcal{S}(G) \hat{\otimes}_{C_c^\infty(G)} P_\bullet$ of V must split, so

that $[V] = \sum_{n=0}^{\infty} (-1)^n [\mathcal{S}(G) \hat{\otimes}_{C_c^\infty(G)} P_n]$ in $K_0(\mathcal{S}G)$. This implies that the combinatorial formal dimension computed using the class of V in $K_0(C_c^\infty G)$ agrees with the usual formal dimension.

6. Conclusion

We have seen that the derived category of tempered smooth representations of a reductive p -adic group G is a full subcategory of the derived category of all smooth representations G . To achieve this, we had to incorporate some functional analysis into our categories of modules. The proof of this result is inspired by the proof of a similar result for discrete groups that uses ideas from geometric group theory.

This comparison result is useful in two ways: it shows that the modules over the Schwartz algebra have reasonably simple projective resolutions because this happens over the Hecke algebra. And it shows that Ext and Tor vanish for tempered representations if one of them is square-integrable. We have also seen one consequence – the quantisation of formal dimensions of square-integrable representations – whose statement does not involve any homological algebra.

References

- [1] Jacek Brodzki and Roger Plymen, *Periodic cyclic homology of certain nuclear algebras*, C. R. Acad. Sci. Paris Sér. I Math. **329** (1999), no. 8, 671–676 (English, with English and French summaries). MR 1724090
- [2] Werner Geigle and Helmut Lenzing, *Perpendicular categories with applications to representations and sheaves*, J. Algebra **144** (1991), no. 2, 273–343. MR 1140607
- [3] Paul Jolissaint, *Rapidly decreasing functions in reduced C^* -algebras of groups*, Trans. Amer. Math. Soc. **317** (1990), no. 1, 167–196. MR 943303
- [4] Henning Krause, *Cohomological quotients and smashing localizations*, Amer. J. Math. **127** (2005), no. 6, 1191–1246. MR 2183523
- [5] Ralf Meyer, *Combable groups have group cohomology of polynomial growth*, Q. J. Math. **57** (2006), no. 2, 241–261. MR 2237601
- [6] ———, *Embeddings of derived categories of bornological modules* (2004), eprint. arXiv:math.FA/0410596.
- [7] ———, *Homological algebra for schwartz algebras of reductive p -adic groups* (2006), 263–300. MR 2327309
- [8] ———, *On a representation of the idele class group related to primes and zeros of l -functions*, Duke Math. J. **127** (2005), no. 3, 519–595. MR 2132868
- [9] ———, *Smooth group representations on bornological vector spaces*, Bull. Sci. Math. **128** (2004), no. 2, 127–166 (English, with English and French summaries). MR 2039113
- [10] Amnon Neeman and Andrew Ranicki, *Noncommutative localisation in algebraic k -theory. i*, Geom. Topol. **8** (2004), 1385–1425 (electronic). MR 2119300
- [11] Alexei Yu. Pirkovskii, *Stably flat completions of universal enveloping algebras*, Dissertationes Math. (Rozprawy Mat.) **441** (2006), 60. MR 2283621

- [12] Peter Schneider and Ulrich Stuhler, *Representation theory and sheaves on the Bruhat–Tits building*, Inst. Hautes Études Sci. Publ. Math. (1997), no. 85, 97–191. MR 1471867
- [13] Maarten Solleveld, *Periodic cyclic homology of reductive p -adic groups* (2007), eprint. [arXiv:0710.5815](https://arxiv.org/abs/0710.5815).
- [14] Joseph L. Taylor, *A general framework for a multi-operator functional calculus*, Advances in Math. **9** (1972), 183–252. MR 0328625
- [15] Marie-France Vignéras, *On formal dimensions for reductive p -adic groups*, Festschrift in honor of I. I. Piatetski-Shapiro on the occasion of his sixtieth birthday, Part I (Ramat Aviv, 1989), 1990, pp. 225–266. MR 1159104

WEIGHTS IN GENERALIZATIONS OF SERRE'S CONJECTURE AND THE MOD p LOCAL LANGLANDS CORRESPONDENCE

Michael M. Schein

Department of Mathematics, Bar-Ilan University
E-mail: mschein@math.biu.ac.il

Abstract. In this mostly expository article we give a survey of some of the generalizations of Serre's conjecture and results towards them that have been obtained in recent years. We also discuss recent progress towards a mod p local Langlands correspondence for p -adic fields and its connections with Serre's conjecture. A theorem describing the structure of some mod p Hecke algebras for GL_n is proved.

1. Introduction

1.1. The classical Serre conjecture. Algebraic number theory is, in some sense, the study of the group $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. In particular, we are interested in the continuous representations of this group and its finite index subgroups. One source of such representations is the cohomology of algebraic varieties. For instance, if X is a variety defined over a number field F and \mathcal{F} is an étale sheaf on X , then the étale cohomology $H_{\acute{e}t}^*(X \otimes \overline{\mathbb{Q}}, \mathcal{F})$ carries an action of $\text{Gal}(\overline{\mathbb{Q}}/F)$. Serre's conjectures and its generalizations tell us which representations arise in this way.

In this section, we will briefly review the original conjecture of Serre. The reader is directed to the excellent expository article [26] for details. For each prime l , fix a

2000 Mathematics Subject Classification. 11F80, 11F85.

The author is grateful to Florian Herzig and David Kazhdan for comments on earlier versions of this paper. All errors are the author's.

decomposition subgroup $G_l \subset \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ at l , let $I_l \subset G_l$ be the corresponding inertia subgroup, and let $P_l \subset I_l$ be the pro- l -Sylow wild inertia subgroup. Let $N \geq 1$ and recall that $\Gamma_1(N) \subset \text{SL}_2(\mathbb{Z})$ is the subgroup

$$\Gamma_1(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{SL}_2(\mathbb{Z}) : a - 1 \equiv c \equiv d - 1 \equiv 0 \pmod{N} \right\}.$$

A modular form of weight k and level N is a holomorphic function $f : \mathcal{H} \rightarrow \mathbb{C}$, where \mathcal{H} is the complex upper half plane, satisfying the following automorphy condition, as well as some growth conditions at infinity:

$$f\left(\frac{az+b}{cz+d}\right) = (cz+d)^k f(z) \quad \forall z \in \mathcal{H}, \forall \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_1(N).$$

Note that $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in \Gamma_1(N)$, whence $f(z+1) = f(z)$ for all $z \in \mathcal{H}$, so that f has a Fourier expansion. If f is, moreover, cuspidal and an eigenform for the Hecke operators, then the expansion has only positive terms: $f(z) = \sum_{n \geq 1} a_n q^n$, where $q = e^{2\pi iz}$. The classical construction of Eichler and Shimura associates to such an f a two-dimensional Galois representation $\rho_f : \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \rightarrow \text{GL}_2(\overline{\mathbb{F}}_p)$ which is unramified at all primes $l \nmid Np$. In other words, $I_l \subset \ker \rho_f$; note that I_l is defined up to conjugation, but this condition is well-defined. It follows that if Frob_l is an arithmetic Frobenius element for l , the characteristic polynomial of $\rho_f(\text{Frob}_l)$ is well-defined for $l \nmid Np$. It is $x^2 - a_l x + l^{k-1}$, and by the Chebotarev density theorem the facts stated here determine ρ_f up to isomorphism. The Eichler-Shimura construction essentially comes down to finding ρ_f inside the cohomology of a suitable modular curve. We say that a Galois representation $\rho : \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \rightarrow \text{GL}_2(\overline{\mathbb{F}}_p)$ is *modular* if $\rho \simeq \rho_f$ for some modular form f .

Suppose we are given $\rho : \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \rightarrow \text{GL}_2(\overline{\mathbb{F}}_p)$, and let $c \in \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ be the element induced by complex conjugation. Since c^2 is the identity, we must have $\det \rho(c) = \pm 1$. We say that ρ is *odd* if $\det \rho(c) = -1$ and *even* otherwise. In the early 1970's, J.-P. Serre conjectured that

Conjecture 1.1. Suppose that $\rho : \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \rightarrow \text{GL}_2(\overline{\mathbb{F}}_p)$ is continuous, irreducible, and odd. Then ρ is modular.

Moreover, Serre gave a combinatorial recipe for the weights of the modular forms giving rise to such ρ . The quantitative statement that a continuous, irreducible, odd two-dimensional Galois representation is modular *of the specified weights* is

called the strong Serre conjecture. Conjecture 1.1 was recently proved by Khare and Wintenberger [18], [19], [20], relying on work of Kisin [22], [21]. However, the implication that the strong Serre conjecture follows from Conjecture 1.1 (i.e. that if ρ is modular, then it is modular of precisely the predicted weights) was known much earlier, except for a few cases with $p = 2$. It was established by work of Mazur, Ribet, Deligne, Fontaine, Carayol, Gross, Coleman, Voloch, and Edixhoven, among many others. An example of a statement towards this result is the following theorem of Fontaine. It was proved in a letter to Serre in 1979, and a somewhat different proof eventually appeared in print in [8]. Recall that $I_p \simeq \text{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p^{nr})$, where \mathbb{Q}_p^{nr} is the maximal unramified extension of \mathbb{Q}_p . A character $\psi : I_p \rightarrow \overline{\mathbb{F}}_p^*$ is said to be of level two if it factors through the quotient $\text{Gal}(K_2/\mathbb{Q}_p^{nr}) \simeq \mathbb{F}_{p^2}^*$ of I_p , where $\mathbb{Q}_p^{nr} \subset K_2 \subset \overline{\mathbb{Q}}_p$ is the unique subextension with $[K_2 : \mathbb{Q}_p^{nr}] = p^2 - 1$.

Theorem 1.2 (Fontaine). *Let $\rho : \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \rightarrow \text{GL}_2(\overline{\mathbb{F}}_p)$ be modular of weight k and level N , with $2 \leq k \leq p + 1$. Suppose that $\rho|_{G_p}$ is irreducible. Let $\psi, \psi' : I_p \rightarrow \overline{\mathbb{F}}_p^*$ be the characters of level two induced by the two field embeddings $\mathbb{F}_{p^2} \hookrightarrow \overline{\mathbb{F}}_p$. Then,*

$$\rho|_{I_p} \sim \begin{pmatrix} \psi^{k-1} & 0 \\ 0 & (\psi')^{k-1} \end{pmatrix}.$$

1.2. Generalizations of Serre's conjecture. How can this picture be generalized? For any number field F and any $n \geq 1$, we need to have a notion of a Galois representation $\rho : \text{Gal}(\overline{\mathbb{Q}}/F) \rightarrow \text{GL}_n(\overline{\mathbb{F}}_p)$ being modular; roughly this should mean that ρ arises “geometrically.” In most cases it is not known how to associate Galois representations to automorphic objects, making it difficult to motivate a natural definition of modularity. But given a notion of modularity, one can seek to formulate analogues of the Serre and strong Serre conjectures.

If F is a totally real number field, then it is known how to construct compatible families of Galois representations associated to Hilbert modular forms over F (see [4], [32]). The first generalizations of Serre's conjecture dealt with this case. When p is unramified in F , a conjecture was formulated by Buzzard, Diamond, and Jarvis [7]. It was extended by the author to totally real fields F where p ramifies arbitrarily, but only when ρ is tamely ramified at all places above p . This conjecture is discussed in the next section.

Serre's conjecture has also been generalized in another direction, to Galois representations $\rho : \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \rightarrow \text{GL}_n(\overline{\mathbb{F}}_p)$ for arbitrary n . Ash, Doud, D. Pollack, and Sinnott [2], [1] conjectured a combinatorial recipe for some modular weights in this

case, but never claimed to have found all the modular weights. When ρ is tamely ramified at p , their work was improved by Herzig [15] who conjectured a complete list of regular modular weights that is defined more conceptually. A more general conjecture (without the assumption of tame ramification) was made by Gee [10], but it specifies the modular weights in terms of the existence of certain crystalline lifts and does not usually allow them to be written down explicitly. In fact, the conjecture of Buzzard, Diamond, and Jarvis was already phrased in terms of crystalline lifts, and hence non-explicit, when ρ had wild ramification at p .

The framework exists for stating Serre’s conjecture in an even more general context. This work was begun by Gross in [12] and [13]. Let G/\mathbb{Q} be a reductive group such that all the arithmetic subgroups of $G(\mathbb{Q})$ are finite, and suppose that G is an inner form of a split group over \mathbb{Q} . Let \hat{G} be the split dual group over \mathbb{Z} . To each weight and level (a weight in this general context is an irreducible $\overline{\mathbb{F}}_p$ -representation of $G(\mathbb{F}_p)$) Gross associated a space of modular forms with a Hecke-algebra action and conjectured ([12], Conj. 1.1) that to any Hecke eigenform one can associate a Galois representation $\rho : \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \rightarrow \hat{G}(\overline{\mathbb{F}}_p)$ satisfying certain properties. This provides a notion of modularity. One can now ask when a representation $\rho : \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \rightarrow \hat{G}(\overline{\mathbb{F}}_p)$ is modular and what the weights are of the modular forms giving rise to it.

2. Serre’s conjecture for Hilbert modular forms

2.1. Weights and modularity. Let F be a number field and $n \geq 1$. For any place v of F , let \mathcal{O}_v denote the ring of integers of the completion F_v , and let k_v be the residue field. A *Serre weight* is an irreducible $\overline{\mathbb{F}}_p$ -representation of the finite group $\text{GL}_n(\mathcal{O}_F/p)$. Any Serre weight factors through the quotient map

$$\text{GL}_n(\mathcal{O}_F/p) \rightarrow \Delta = \prod_{v|p} \text{GL}_n(k_v),$$

since the kernel is a p -group. Therefore Serre weights have the form $\sigma = \otimes_{v|p} \sigma_v$, where σ_v is an irreducible $\overline{\mathbb{F}}_p$ -representation of $\text{GL}_n(k_v)$. We call such σ_v *local Serre weights at v* .

Suppose for the rest of this section that $n = 2$ and F is totally real. We now introduce a notion of modularity of a mod p Galois representation $\rho : \text{Gal}(\overline{\mathbb{Q}}/F) \rightarrow \text{GL}_2(\overline{\mathbb{F}}_p)$. Let D/F be a quaternion algebra that splits at exactly one infinite place and at all places dividing p . Consider the reductive group $G = \text{Res}_{F/\mathbb{Q}}(D^*)$ and let $U \subset G(\mathbb{A}^\infty)$ be an open compact subgroup. Let X_U/F be the associated Shimura curve; its complex points are $X_U(\mathbb{C}) = G(\mathbb{Q}) \backslash G(\mathbb{A}^\infty) \times (\mathbb{C} - \mathbb{R})/U$. Recall that the

relative Picard scheme of X_U parametrizes line bundles locally of degree zero, and let $\text{Pic}^0(X_U)/F$ be its component containing the identity. This is an abelian variety.

Let $U'_p = \ker((D \otimes \hat{\mathbb{Z}})_p^* = \prod_{v|p} \text{GL}_2(\mathcal{O}_v) \rightarrow \text{GL}_2(\mathcal{O}_F/p))$, and let $U''_p = \ker(\prod_{v|p} \text{GL}_2(\mathcal{O}_v) \rightarrow \prod_{v|p} \text{GL}_2(k_v))$. Clearly $U'_p \subset U''_p$. We say that an open compact $U \subset G(\mathbb{A}^\infty)$ is of type $(*)$ if $U = U'_p \times U^p$, where $U^p \subset G(\mathbb{A}^{\infty,p})$. Let $V = \prod_{v|p} \text{GL}_2(\mathcal{O}_v) \times U^p$. If U^p is sufficiently small in the sense of section 3.1 of [29], then $X_U \rightarrow X_V$ is a Galois cover with group $V/U = \text{GL}_2(\mathcal{O}_F/p)$. Hence we have an action of V/U on $\text{Pic}^0(X_U)$.

Definition 2.1. Let σ be a Serre weight. An irreducible Galois representation $\rho : \text{Gal}(\bar{\mathbb{Q}}/F) \rightarrow \text{GL}_2(\bar{\mathbb{F}}_p)$ is *modular of weight σ* if there exists a quaternion algebra D/F as above and an open compact $U \subset (D \otimes \hat{\mathbb{Z}})^* \subset G(\mathbb{A}^\infty)$ of type $(*)$, such that $(\text{Pic}^0(X_U)[p] \otimes_{\bar{\mathbb{F}}_p} \sigma)^{\text{GL}_2(\mathcal{O}_F/p)} = (\text{Pic}^0(X_{U''_p \times U^p})[p] \otimes_{\bar{\mathbb{F}}_p} \sigma)^\Delta$ has ρ as a Jordan-Hölder constituent.

If $[F : \mathbb{Q}] = d$, then each of the d embeddings of F into $\mathbb{R} \subset \mathbb{C}$ induces a “complex conjugation” in $\text{Gal}(\bar{\mathbb{Q}}/F)$. Denote these complex conjugations c_1, \dots, c_d . We say that ρ is *totally odd* if $\det \rho(c_i) = -1$ for all $1 \leq i \leq d$. The qualitative Serre conjecture generalizes to our situation as follows.

Conjecture 2.2. Suppose that $\rho : \text{Gal}(\bar{\mathbb{Q}}/F) \rightarrow \text{GL}_2(\bar{\mathbb{F}}_p)$ is continuous, irreducible, and totally odd. Then ρ is modular.

To formulate an analogue of the strong Serre conjecture we must, given a ρ , specify its modular weights. The Langlands philosophy suggests that the modular weights should be determined by local information, as was indeed the case for $F = \mathbb{Q}$. We fix a decomposition subgroup $G_v \subset \text{Gal}(\bar{\mathbb{Q}}/F)$ for each place $v|p$ and will define a set $W_v(\rho)$ of local Serre weights at v that depends only on the restriction $\rho|_{G_v}$. Then we will conjecture that

Conjecture 2.3. Let $\rho : \text{Gal}(\bar{\mathbb{Q}}/F) \rightarrow \text{GL}_2(\bar{\mathbb{F}}_p)$ be a Galois representation. Its set of modular weights is

$$W(\rho) = \left\{ \sigma = \bigotimes_{v|p} \sigma_v : \forall v|p, \sigma_v \in W_v(\rho) \right\}.$$

Fix a place \mathfrak{p} of F dividing p , let the cardinality of $k_{\mathfrak{p}}$ be $q = p^f$, and let e be the ramification index of $F_{\mathfrak{p}}$ over \mathbb{Q}_p . When $n = 2$, it is easy to give explicit models for the local Serre weights at \mathfrak{p} . Let I be the set of field embeddings $\tau : k_{\mathfrak{p}} \hookrightarrow \bar{\mathbb{F}}_p$. Let

inertia by conjugation. It follows that $\{\varphi^q, (\varphi')^q\} = \{\varphi, \varphi'\}$, so that two possible cases arise:

1. The characters φ, φ' have niveau $2f$, so that $\varphi' = \varphi^q$ and $(\varphi')^q = \varphi$. This implies the irreducibility of $\rho|_{G_p}$.
2. The characters φ and φ' are of niveau f , and $\rho|_{G_p}$ is reducible.

We are now ready to provide a recipe for $W_p(\rho)$ when ρ is tamely ramified at \mathfrak{p} .

Definition 2.4. Let $\rho : \text{Gal}(\overline{\mathbb{Q}}/F) \rightarrow \text{GL}_2(\overline{\mathbb{F}}_p)$ be a Galois representation.

1. Suppose that $\rho|_{G_p}$ is irreducible. Then the local Serre weight

$$(1) \quad \sigma_{\mathfrak{p}} = \bigotimes_{\tau \in I} (\det^{w_{\tau}} \text{Sym}^{k_{\tau}-2} k_{\mathfrak{p}}^2) \otimes_{k_{\mathfrak{p},\tau}} \overline{\mathbb{F}}_p$$

is contained in $W_p(\rho)$ if and only if for each $\tau \in I$ there exists a labeling $\{\tilde{\tau}, \tilde{\tau}'\}$ of the two lifts of τ to $k'_{\mathfrak{p}}$ and an integer $0 \leq \delta_{\tau} \leq e-1$ such that

$$\rho|_{I_p} \sim \prod_{\tau \in I} \psi_{\tau}^{w_{\tau}} \begin{pmatrix} \prod_{\tilde{\tau}} \psi_{\tilde{\tau}}^{k_{\tau}-1+\delta_{\tau}} \psi_{\tilde{\tau}'}^{e-1-\delta_{\tau}} & 0 \\ 0 & \prod_{\tilde{\tau}} \psi_{\tilde{\tau}}^{e-1-\delta_{\tau}} \psi_{\tilde{\tau}'}^{k_{\tau}-1+\delta_{\tau}} \end{pmatrix}.$$

2. Suppose that $\rho|_{G_p}$ is reducible and that ρ is tamely ramified at \mathfrak{p} (i.e. $P_{\mathfrak{p}} \subset \ker \rho$). Then $W_p(\rho)$ consists precisely of the Serre weights as in (1) for which there exists a subset $J \subset I$ and an integer $0 \leq \delta_{\tau} \leq e-1$ for each $\tau \in I$ such that

$$\rho|_{I_p} \sim \prod_{\tau \in I} \psi_{\tau}^{w_{\tau}} \begin{pmatrix} \prod_{\tau \in J} \psi_{\tau}^{k_{\tau}-1+\delta_{\tau}} \prod_{\tau \notin J} \psi_{\tau}^{e-1-\delta_{\tau}} & 0 \\ 0 & \prod_{\tau \in J} \psi_{\tau}^{e-1-\delta_{\tau}} \prod_{\tau \notin J} \psi_{\tau}^{k_{\tau}-1+\delta_{\tau}} \end{pmatrix}.$$

In particular, if $e \geq p-1$, then all Serre weights of suitable central character should be modular.

2.2. Evidence. There are some results available towards this conjecture. Dembélé's computations of modular weights of Hilbert modular forms over $\mathbb{Q}(\sqrt{5})$ for $p = 5$ agree with Conjecture 2.3 (see section 4 of [28]). The following theoretical results have also been established.

Theorem 2.5 ([28], Theorem 3.4). *Suppose that $e < p-1$ and let $\rho : \text{Gal}(\overline{\mathbb{Q}}/F) \rightarrow \text{GL}_2(\overline{\mathbb{F}}_p)$ be such that $\rho|_{G_p}$ is irreducible and ρ is modular of weight $\sigma = \otimes_{\nu|p} \sigma_{\nu}$, where $\sigma_{\mathfrak{p}}$, written as in (1), satisfies $k_{\tau} - 2 + e \leq p-1$ for all $\tau \in I$. Then $\sigma_{\mathfrak{p}} \in W_p(\rho)$.*

If p is unramified in F , then a stronger result has been proved by Gee. We say that a local Serre weight at \mathfrak{p} , written as in (1), is *strongly regular* if $3 \leq k_{\tau} \leq p-1$ for all $\tau \in I$ and *regular* if $2 \leq k_{\tau} \leq p$ for all $\tau \in I$. Gee uses a variant definition of modularity,

working with definite quaternion algebras, but his results should be translatable to our setting.

Theorem 2.6 ([11], Thms. 5.1.2 and 5.1.3). *Suppose that p is unramified in F , that $\rho : \text{Gal}(\overline{\mathbb{Q}}/F) \rightarrow \text{GL}_2(\overline{\mathbb{F}}_p)$ is modular of weight σ , and that σ is strongly regular. Then $\sigma \in W(\rho)$.*

If $\sigma \in W(\rho)$ is strongly regular and non-ordinary, then it is a modular weight of ρ .

We refer the reader to Gee’s paper for the definition of “non-ordinary,” which is a technical condition; “generically” weights are non-ordinary. We will briefly explain why the hypothesis of $k_\tau - 2 + e \leq p - 1$ in Theorem 2.5 and that of strong regularity in Gee’s theorem ultimately stem from the same source.

A representation $\rho : \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \rightarrow \text{GL}_2(\overline{\mathbb{F}}_p)$ that arises from a modular form of weight $2 \leq k \leq p + 1$ and nebentype w is modular of weight $\det^w \otimes \text{Sym}^{k-2} \overline{\mathbb{F}}_p^2$. It is now clear that Theorem 2.5 is a generalization of Fontaine’s Theorem 1.2 above, and its proof follows the same method, although it works with Shimura curves rather than modular curves and deals with complications introduced by the extra ramification. Let $B_2(k_p) \subset \text{GL}_2(k_p)$ be the upper triangular Borel subgroup, and choose a character $\theta : B_2(k_p) \rightarrow \overline{\mathbb{F}}_p^*$ such that σ_p is a subquotient of $\text{Ind}_{B_2(k_p)}^{\text{GL}_2(k_p)} \theta$. Then we find ρ inside a suitable piece of $\text{Jac}(X_{U_1^{bal}(p) \times U^p})[p^\infty]$, where

$$U_1^{bal}(p) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{GL}_2(\mathcal{O}_p) : a - 1, c, d - 1 \in \mathfrak{p} \right\}$$

and U^p is the part of U away from \mathfrak{p} . This suitable piece is a vector space scheme to which we can apply Raynaud’s theory [25] to obtain combinatorial restrictions on ρ . Effectively we have lifted a mod p Hilbert modular form giving rise to ρ to trivial weight, but at the price of raising its level to $U_1^{bal}(p) \times U^p$. In the process we have lost information about σ_p and retained only θ . In fact, at this step of the proof ([28], Prop. 3.3) ρ is restricted to being precisely one of the Galois representations for which one of the subquotients of $\text{Ind}_{B_2(k_p)}^{\text{GL}_2(k_p)} \theta$ is conjectured to be a modular weight, which is the best possible result. Then we consider all characters θ such that σ_p appears in $\text{Ind}_{B_2(k_p)}^{\text{GL}_2(k_p)} \theta$ and intersect the sets of permitted ρ ’s. If the hypothesis that $k_\tau - 2 + e \leq p - 1$ for all $\tau \in I$ does not hold, then this intersection is too large and contains ρ ’s for which σ is not conjectured to be modular.

Gee's proof follows a completely different method and hinges on a variant of one of Kisin's modularity lifting results. Kisin's functor from crystalline Galois representations to a certain category of \mathfrak{S} -modules is only essentially surjective if the Hodge-Tate numbers on both sides are restricted to $\{0, 1\}$, and the needed modularity lifting theorem is only available in this case. This forces Gee also to lift to weight 2, work in weight 2, and deal with the same combinatorics at the end. The point is that Theorems 2.5 and 2.6 are the best results obtainable by their methods of proof with the available technology, except for ad hoc tricks such as the result of [27], which allows one to prove a bit more when the residue field k_p is small. In order to move forward, it appears that one will need more general modularity lifting theorems.

2.3. A more conceptual formulation of the sets of modular weights. In this section we will briefly discuss the structure of Herzig's conjecture and how his ideas can be used to restate Conjecture 2.3 more conceptually.

Let $\rho : \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \rightarrow \text{GL}_n(\overline{\mathbb{F}}_p)$ be a Galois representation that is tamely ramified at p . Then $\rho|_{I_p}$ factors through the abelian group I_p/P_p , so it is a sum of characters. The action of G_p on I_p/P_p by conjugation implies that if a character $\psi : \mathbb{F}_p^* \rightarrow \overline{\mathbb{F}}_p^*$ appears in $\rho|_{I_p}$, then so do all of its Galois conjugates. Hence there is a partition $n_1 + \dots + n_r = n$ such that

$$\rho|_{I_p} \sim \begin{pmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_r \end{pmatrix},$$

where each A_i is an $n_i \times n_i$ diagonal matrix whose diagonal entries are a Galois conjugacy class $[\psi_i]$ of characters $\mathbb{F}_p^* \rightarrow \overline{\mathbb{F}}_p^*$. This conjugacy class defines a cuspidal characteristic zero representation $\Theta([\psi_i])$ of $\text{GL}_{n_i}(\mathbb{F}_p)$. Now let $P \subset \text{GL}_n(\mathbb{F}_p)$ be the standard parabolic subgroup whose Levi subgroup is $\text{GL}_{n_1}(\mathbb{F}_p) \times \dots \times \text{GL}_{n_r}(\mathbb{F}_p)$. We define a characteristic zero representation $V(\rho|_{I_p})$ of $\text{GL}_n(\mathbb{F}_p)$ by

$$V(\rho|_{I_p}) = \text{Ind}_P^{\text{GL}_n(\mathbb{F}_p)} (\Theta([\psi_1]) \otimes \dots \otimes \Theta([\psi_r])).$$

Let $JH(\overline{V(\rho|_{I_p})})$ be the set of Jordan-Hölder constituents of the reduction modulo p of $V(\rho|_{I_p})$. Its elements are irreducible $\overline{\mathbb{F}}_p$ -representations of $\text{GL}_n(\mathbb{F}_p)$ – in other words, weights. Herzig defines a class of *regular* weights analogous to that for GL_2 . Roughly, a weight is regular if it does not lie on some boundaries of alcoves; for $n = 2$ the notion of regularity coincides with the one defined above. Then he defines an

operator \mathcal{R} sending the set of Serre weights into that of regular Serre weights and conjectures ([15], Conj. 6.9) that the set of regular modular Serre weights of ρ is $\mathcal{R}(JH(\overline{V(\rho|_{I_p})}))$.

Herzig also reformulates the conjecture of Buzzard, Diamond, and Jarvis in this language. Given a tamely ramified $\rho : \text{Gal}(\overline{\mathbb{Q}}/F) \rightarrow \text{GL}_2(\overline{\mathbb{F}}_p)$, for F a totally real field in which p does not ramify, he defines a representation $V(\rho|_{I_p})$ of $\text{GL}_2(k_p)$ such that $W_p(\rho) = \mathcal{R}(JH(\overline{V(\rho|_{I_p})}))$ for an operator \mathcal{R} acting on the appropriate set of local Serre weights. Herzig's restatement of the conjecture accounts for the non-regular weights as well, at the price of \mathcal{R} being multi-valued.

If F_p has ramification index e over \mathbb{Q}_p , then given a tamely ramified Galois representation $\rho : \text{Gal}(\overline{\mathbb{Q}}/F) \rightarrow \text{GL}_2(\overline{\mathbb{F}}_p)$ it is shown in [30] that one can define a collection $V(\rho)$ of characteristic zero representations of $\text{GL}_2(k_p)$ such that

$$W_p(\rho) = \bigcup_{V \in V(\rho)} \mathcal{R}(JH(\overline{V})).$$

The collection $V(\rho)$ has at most e^f elements, where $f = [k_p : \mathbb{F}_p]$.

3. Representations of $\text{GL}_n(M)$, for M a p -adic field

3.1. Notation. Let M/\mathbb{Q}_p be a finite extension, and denote by \mathcal{O} its ring of integers, by π a uniformizer, and by k the residue field. Set $G = \text{GL}_n(M)$, and let K denote the maximal compact subgroup $\text{GL}_n(\mathcal{O})$. Let $Z \simeq M^*$ be the center of G . Write $\Gamma \subset K$ for subgroup of matrices congruent to the identity modulo π . Let N be the collection of non-decreasing n -tuples of integers $\nu = (\nu_1, \dots, \nu_n)$, where $0 = \nu_1 \leq \nu_2 \leq \dots \leq \nu_n$. For $\nu \in N$, let $\alpha_\nu \in G$ be the matrix

$$\alpha_\nu = \begin{pmatrix} 1 & & & \\ & \pi^{\nu_2} & & \\ & & \ddots & \\ & & & \pi^{\nu_n} \end{pmatrix}.$$

For $0 \leq i \leq n$, let $\overline{P}_i \subset \text{GL}_n$ be the parabolic subgroup

$$\overline{P}_i = \left\{ \begin{pmatrix} A & 0 \\ C & D \end{pmatrix} : A \in \text{GL}_i, D \in \text{GL}_{n-i} \right\}.$$

In particular, $\overline{P}_0 = \overline{P}_n = \text{GL}_n$.

Recall the Cartan decomposition of G into a disjoint union of double cosets of KZ :

$$(2) \quad G = \coprod_{v \in N} KZ\alpha_v^{-1}K.$$

Let $B \subset G$ be the upper triangular Borel subgroup, and let $T \subset B$ and $U \subset B$ be the diagonal torus and the unipotent radical, respectively. Let $S = \{\beta_i : 1 \leq i \leq n-1\}$ be the standard set of simple roots of GL_n ; here $\beta_i = e_i - e_{i+1}$, where X_{ij} is the dual basis of $M_n(M)^*$ and $e_i = X_{ii}|_T$.

Let $W \simeq S_n$ be the Weyl group of GL_n . We consider W as a subgroup of K via the usual realization by permutation matrices. Let $\hat{w} \in W$ be the element of greatest length. Let $\omega : K \rightarrow \mathrm{GL}_n(k)$ be the natural surjection. The Iwahori subgroup $I \subset K$ is the preimage of $B(k)$ under ω , and we denote by $I(1) \subset I$ the pro- p -Sylow subgroup.

If $H \subset G$ is an open subgroup and (τ, V_τ) is a smooth representation of H , then we recall that the compact induction $\mathrm{ind}_H^G \tau$ is a smooth representation of G . A model for it is the space $S(G, \tau)$ of compactly supported locally constant functions $f : G \rightarrow V_\tau$ such that $f(hg) = \tau(h)f(g)$ for all $g \in G$ and $h \in H$. The G -action is given by $(g'f)(g) = f(gg')$. For each $v' \in V_\tau$ we can define an element $\hat{f}_{v'} \in \mathrm{ind}_H^G \tau$ by

$$(3) \quad \hat{f}_{v'}(g) = \begin{cases} \tau(g)v' & : g \in H \\ 0 & : g \notin H. \end{cases}$$

We observe that the set $\{\hat{f}_{v'} : v' \in V_\tau\}$ generates $\mathrm{ind}_H^G \tau$ as a G -module.

3.2. Hecke algebras. In this section we recall some basic facts about Hecke algebras. Let $H \subset G$ be an open subgroup, and let (τ, V_τ) be a smooth representation of H . Then we define

$$\mathcal{H}(H, \tau) = \mathrm{End}_G(\mathrm{ind}_H^G \tau).$$

Recall from section 2.2 of [3] that $\mathcal{H}(H, \tau)$ can be interpreted as a convolution algebra as follows. Let $\mathcal{H}_H(\tau)$ be the space of functions $\varphi : G \rightarrow \mathrm{End}_{\mathbb{F}_p}(V_\tau)$ such that

1. For each $v \in V_\tau$, the function $\varphi_v : G \rightarrow V_\tau$ given by $\varphi_v(g) = \varphi(g)v$ is locally constant on G and supported on a set of the form HC , where $C \subset G$ is compact.
2. For all $g \in G$ and all $h_1, h_2 \in H$, we have $\varphi(h_1gh_2) = \tau(h_1)\varphi(g)\tau(h_2)$.

Given $\varphi_1, \varphi_2 \in \mathcal{H}_H(\tau)$, we define their convolution $\varphi_2 * \varphi_1$ to be

$$(\varphi_2 * \varphi_1)(g)(v) = \sum_{y \in G/H} \varphi_2(y)\varphi_1(y^{-1}g)(v)$$

for all $g \in G$ and $v \in V_\tau$. We recall the following result, which states that $\mathcal{H}(H, \tau)$ and $\mathcal{H}_H(\tau)$ are isomorphic and that composition in the former corresponds to convolution in the latter.

Proposition 3.1 ([3], **Proposition 5**). *Given $\varphi \in \mathcal{H}_H(\tau)$, define $T_\varphi \in \mathcal{H}(H, \tau)$ by*

$$(4) \quad T_\varphi(f)(g) = \sum_{Hx \in H \backslash G} \varphi(gx^{-1})f(x) = \sum_{yH \in G/H} \varphi(y)f(y^{-1}g)$$

for all $g \in G$ and $f \in S(G, \tau)$. Then the map

$$\begin{aligned} \mathcal{H}_H(\tau) &\rightarrow \mathcal{H}(H, \tau) \\ \varphi &\mapsto T_\varphi \end{aligned}$$

is an isomorphism of $\overline{\mathbb{F}}_p$ -modules. Moreover, if $\varphi_1, \varphi_2 \in \mathcal{H}_H(\tau)$, then $T_{\varphi_2} \circ T_{\varphi_1} = T_{\varphi_2 * \varphi_1}$.

Let σ be an irreducible $\overline{\mathbb{F}}_p$ -representation of $GL_n(k)$. We can view σ as a representation of KZ after inflating via ω and letting π act trivially. Let V_σ be the representation space, and choose a highest weight vector v . Then $v' = \hat{w} \cdot v$ is a lowest weight vector.

3.3. The structure of $\mathcal{H}(KZ, \sigma)$. Our immediate aim is to prove Proposition 3.5, which states that the Hecke algebra $\mathcal{H}(KZ, \sigma)$ is isomorphic to a polynomial algebra $\overline{\mathbb{F}}_p[T_1, \dots, T_{n-1}]$ in $n - 1$ canonical generators. This was proved independently by Herzig, who recently extended his result for GL_n to a Satake isomorphism determining the structure of Hecke algebras $\mathcal{H}(KZ, \sigma)$, where G is any connected split reductive group, $K \subset G$ is a maximal compact subgroup, and σ is an irreducible mod p representation of KZ (see [16]). Nevertheless, we believe it is useful to write down an explicit proof for GL_n ; in particular, readers who are unfamiliar with the representation theory of reductive groups may benefit from comparing the two arguments.

Given $v \in N$, we will define a linear transformation U_v of the vector space V_σ . Let $\omega : K \rightarrow GL_n(k)$ be the natural projection, and set $\overline{P}_v = \omega(K \cap \alpha_v^{-1}K\alpha_v)$. This is a parabolic subgroup of $GL_n(k)$. Let \overline{N}_v be its unipotent radical, let $L_v \in \overline{N}_v$ be the Levi subgroup, and let $P_v = L_v N_v$ be the opposite parabolic. Let $J \subset S$ be the set of simple roots corresponding to L_v .

Let $\lambda : T(k) \rightarrow \overline{\mathbb{F}}_p^*$ be the highest weight of $\sigma = F(\lambda)$. For a weight μ , let $(V_\sigma)_\mu \subset V_\sigma$ be the corresponding weight subspace. Then [17], II.2.11a states that

$$V_\sigma^{N_v} = \bigoplus_{\alpha \in ZJ} (V_\sigma)_{\lambda - \alpha}.$$

It is easy to derive the analogous result for the negative standard parabolic subgroups. Recall that $\hat{w} \in W$ is the longest element of the Weyl group, and define $J' \subset S$ by $J' = \{\beta_i : \beta_{n-i} \in J\}$. If $P_{J'} \subset \mathrm{GL}_n(k)$ is the corresponding parabolic subgroup, then $\overline{P}_v = \hat{w}P_{J'}\hat{w}^{-1}$. Therefore, if $\hat{\lambda} = \hat{w}\lambda$ is the lowest weight of V_σ , we have

$$(5) \quad V_\sigma^{\overline{N}_v} = \hat{w} \left(\bigoplus_{\alpha' \in ZJ'} (V_\sigma)_{\lambda - \alpha'} \right) = \bigoplus_{\alpha \in ZJ} (V_\sigma)_{\hat{\lambda} + \alpha}.$$

We define an endomorphism U_v of V_σ as follows. Let \mathcal{B} be a basis of V_σ consisting of weight vectors. Then for $v \in \mathcal{B}$ we set

$$U_v(v) = \begin{cases} v & : v \in V_\sigma^{\overline{N}_v} \\ 0 & : v \notin V_\sigma^{\overline{N}_v}. \end{cases}$$

By (5) this definition is independent of the choice of \mathcal{B} . For each $v \in N$, let $\varphi_v \in \mathcal{H}_{KZ}(\sigma)$ be the function supported on the double coset $KZ\alpha_v^{-1}K$ that is determined by $\varphi_v(\alpha_v^{-1}) = U_v \in \mathrm{End}_{\overline{\mathbb{F}}_p}(V_\sigma)$.

Lemma 3.2. *The set $\{\varphi_v : v \in N\}$ is a basis of the $\overline{\mathbb{F}}_p$ -vector space $\mathcal{H}_{KZ}(\sigma)$.*

Proof. The φ_v are linearly independent, so we only need to show that they span $\mathcal{H}_{KZ}(\sigma)$. Clearly it suffices to show that any non-zero element $\varphi \in \mathcal{H}_{KZ}(\sigma)$ whose support consists of a sole double coset $KZ\alpha_v^{-1}K$ is a scalar multiple of φ_v .

As above, let L_v be the Levi subgroup of $\overline{P}_v \subset \mathrm{GL}_n(k)$. It is easy to see that $V_\sigma^{\overline{N}_v}$ is preserved by L_v . It is an irreducible L_v -module by the result of [31].

The definition of $\mathcal{H}_{KZ}(\sigma)$ implies that for all $k_1, k_2 \in K$ such that $k_1\alpha_v^{-1} = \alpha_v^{-1}k_2$, we have

$$(6) \quad \sigma(k_1)\varphi(\alpha_v^{-1}) = \varphi(\alpha_v^{-1})\sigma(k_2).$$

In particular, for every $\bar{k} \in \overline{N}_v$ we may choose $k_1 \in \omega^{-1}(\bar{k})$ and $k_2 \in \Gamma$ such that $k_1\alpha_v^{-1} = \alpha_v^{-1}k_2$. Hence the image of $\varphi(\alpha_v^{-1})$ lies in $V_\sigma^{\overline{N}_v}$. Similarly, we can take k_1 and k_2 to lie in $\omega^{-1}(l)$ for any $l \in L_v$. Thus $\varphi(\alpha_v^{-1})|_{V_\sigma^{\overline{N}_v}} : V_\sigma^{\overline{N}_v} \rightarrow V_\sigma^{\overline{N}_v}$ is an L_v -module homomorphism, so by Schur's lemma it is a scalar $c \in \overline{\mathbb{F}}_p$.

Now suppose that $v \in V_\sigma$ is a weight vector that is not contained in $V_\sigma^{\overline{N}_v}$, and let μ be the weight of v . Any $t \in T(\mathcal{O}) \subset K$ commutes with α_v^{-1} , and hence we find that

$$\sigma(t)\varphi(\alpha_v^{-1})(v) = \varphi(\alpha_v^{-1})(\sigma(t)v) = \mu(t)\varphi(\alpha_v^{-1})(v).$$

Hence $\varphi(\alpha_v^{-1})(v) \in (V_\sigma)_\mu$. But $(V_\sigma)_\mu \cap V_\sigma^{\overline{N}_v} = 0$, whence $v \in \ker \varphi(\alpha_v^{-1})$. We have shown that $\varphi(\alpha_v^{-1}) = cU_v$ and hence that $\varphi = c\varphi_v$. Moreover, it is easy to check that φ_v is indeed an element of $\mathcal{H}_{KZ}(\sigma)$. \square

Lemma 3.3. *The Hecke algebra $\mathcal{H}(KZ, \sigma)$ is commutative.*

Proof. We will show that $\mathcal{H}_{KZ}(\sigma)$ is commutative using a Gelfand-type argument suggested by Florian Herzig. Denote the transpose of a matrix g by g^T . Recall that V_σ is the representation space of σ . Following [17] II.2.12, we define V_σ^T as follows: as a vector space it is the dual space $\text{Hom}(V_\sigma, \overline{\mathbb{F}}_p)$, but the $\text{GL}_n(k)$ -action is defined by $(gf)(v) = f(g^T v)$ for all $g \in \text{GL}_n(k)$ and $f \in V_\sigma^T$. It is easy to see that V_σ^T is an irreducible $\text{GL}_n(k)$ -module. Since V_σ and V_σ^T have the same highest weight, they are isomorphic. Let $\alpha : V_\sigma \rightarrow V_\sigma^T$ be an isomorphism and let $\langle, \rangle : V_\sigma \times V_\sigma \rightarrow \overline{\mathbb{F}}_p$ be the corresponding non-degenerate bilinear form; for every $v, v' \in V_\sigma$, we have $(\alpha(v))(v') = \langle v, v' \rangle$. It is easy to see that for all $g \in \text{GL}_n(k)$ we must have

$$(7) \quad \langle v, v' \rangle = \langle gv, (g^{-1})^T v' \rangle.$$

As usual, we view V_σ^T as a KZ -module. If $A \in \text{End}_{\overline{\mathbb{F}}_p}(V_\sigma)$, let $A^T : V_\sigma \rightarrow V_\sigma$ be given by $A^T(v) = \alpha^{-1}(\alpha(v) \circ A)$. Equivalently, A^T is characterized by the relation $\langle A^T v, v' \rangle = \langle v, Av' \rangle$ for all $v, v' \in V_\sigma$. From this it is easy to check that $(AB)^T = B^T A^T$ for $A, B \in \text{End}_{\overline{\mathbb{F}}_p}(V_\sigma)$. Moreover, if $\rho : KZ \rightarrow \text{Aut}(V_\sigma)$ is the action of KZ on V_σ and $A = \rho(h)$ for $h \in KZ$, then $\langle A^T v, v' \rangle = \langle v, hv' \rangle = \langle h^T v, v' \rangle$ by (7), and hence $A^T = \rho(h^T)$.

If v and v' are weight vectors with weights $\lambda : T(k) \rightarrow \overline{\mathbb{F}}_p^*$ and $\lambda' : T(k) \rightarrow \overline{\mathbb{F}}_p^*$, respectively, with $\lambda \neq \lambda'$, then by (7) for all $t \in T(k)$ we have $\langle v, v' \rangle = \langle tv, t^{-1}v' \rangle = \lambda(t)(\lambda')^{-1}(t)\langle v, v' \rangle$, and hence $\langle v, v' \rangle = 0$. Therefore, if B is a basis for V_σ consisting of weight vectors and $v, v' \in B$, then for any $v \in N$ we have

$$\langle U_v v, v' \rangle = \langle v, U_v v' \rangle = \begin{cases} \langle v, v' \rangle & : v, v' \in V^{\overline{N}_v} \\ 0 & : \text{otherwise.} \end{cases}$$

It follows that $U_v^T = U_v$. We now define an involution $\varphi \mapsto \tilde{\varphi}$ of $\mathcal{H}_{KZ}(\sigma)$. For all $\varphi \in \mathcal{H}_{KZ}(\sigma)$ and $g \in G$ we put $\tilde{\varphi}(g) = (\varphi(g^T))^T$. If $h_1, h_2 \in KZ$ and $g \in G$, then we see that $\tilde{\varphi}(h_1 g h_2) = (\varphi(h_2^T g^T h_1^T))^T = \sigma(h_1^T)^T \varphi(g^T)^T \sigma(h_2^T)^T = \sigma(h_1) \tilde{\varphi}(g) \sigma(h_2)$, so that

indeed $\tilde{\varphi} \in \mathcal{H}_{KZ}(\sigma)$. Moreover, the map $\varphi \mapsto \tilde{\varphi}$ is an anti-automorphism of $\mathcal{H}_{KZ}(\sigma)$. Indeed, as y runs over a set of coset representatives of G/KZ , so does $z = g(y^T)^{-1}$, and hence we have:

$$(8) \quad \begin{aligned} \overline{(\varphi_1 * \varphi_2)}(g) &= \left(\sum_{y \in G/KZ} \varphi_1(y) \varphi_2(y^{-1}g^T) \right)^T = \sum_{y \in G/KZ} \varphi_2(y^{-1}g^T)^T \varphi_1(y)^T \\ &= \sum_{z \in G/KZ} \varphi_2(z^T)^T \varphi_1(g^T(z^T)^{-1})^T = \sum_{z \in G/KZ} \tilde{\varphi}_2(z) \tilde{\varphi}_1(z^{-1}g) = (\tilde{\varphi}_2 * \tilde{\varphi}_1)(g). \end{aligned}$$

We claim that $\tilde{\varphi} = \varphi$ for all $\varphi \in \mathcal{H}_{KZ}(\sigma)$. Indeed, for any $\nu, \mu \in N$ we have $(\alpha_\nu^{-1})^T = \alpha_\nu^{-1}$ and hence

$$\tilde{\varphi}_\mu(\alpha_\nu^{-1}) = (\varphi_\mu(\alpha_\nu^{-1}))^T = \begin{cases} U_\mu & : \nu = \mu \\ 0 & : \nu \neq \mu. \end{cases}$$

Therefore $\tilde{\varphi}_\mu = \varphi_\mu$. Now (8) implies the commutativity of $\mathcal{H}_{KZ}(\sigma)$. □

If $1 \leq i \leq n - 1$, define $\nu(i) \in N$ to be the vector $(0, \dots, 0, 1, \dots, 1)$ consisting of i zeroes followed by $n - i$ ones. We will write α_i and φ_i for $\alpha_{\nu(i)}$ and $\varphi_{\nu(i)}$, respectively. Let $T_\nu \in \mathcal{H}(KZ, \sigma)$ be the operator corresponding to φ_ν under the isomorphism of Proposition 3.1, and write T_i for $T_{\nu(i)}$. Note that $\omega(K \cap \alpha_i^{-1} K \alpha_i)$ is the parabolic subgroup $\overline{P}_i \subset \text{GL}_n(k)$ that was defined in the introduction.

Let $W_i \subset W$ be the subgroup of permutations that preserve the subsets $\{1, \dots, i\}$ and $\{i + 1, \dots, n\}$, and let \mathcal{W}_i be a set of coset representatives of W/W_i . Then by the Bruhat decomposition in $\text{GL}_n(k)$ we see that

$$\text{GL}_n(k) = \coprod_{w \in \mathcal{W}_i} \overline{U}(k) w \overline{P}_i,$$

where \overline{U} is the unipotent radical of the lower triangular Borel subgroup. Lifting to K , we can obtain a set of coset representatives for $K/(K \cap \alpha_i^{-1} K \alpha_i)$ of the form $\cup_{w \in \mathcal{W}_i} \{uw : u \in \Lambda_w\}$, where Λ_w is a subset of the lower triangular pro- p -Iwahori subgroup $I(1) \subset K$.

Consider the reverse lexicographical ordering on N defined as follows. If $\nu = (\nu_1, \dots, \nu_n)$ and $\nu' = (\nu'_1, \dots, \nu'_n)$, then $\nu' < \nu$ if and only if there is some $1 \leq r \leq n$ such that $\nu'_r < \nu_r$ and $\nu_i = \nu'_i$ for $i > r$. This gives a complete linear ordering of N . Observe that for every $\nu \in N$ there exist only finitely many $\nu' \in N$ such that $\nu' < \nu$. Note that this ordering is not the opposite of the usual lexicographical ordering; perhaps it would be clearer to call it the Hebrew lexicographical ordering.

Lemma 3.4. *Let $v = v(i)$ for some $1 \leq i \leq n - 1$, and suppose that $\mu, \lambda \in N$ are such that $(\varphi_i * \varphi_\mu)(\alpha_\lambda^{-1}) \neq 0$. Then $\lambda \leq \mu + v(i)$, with respect to the ordering defined above. Moreover, for each $\lambda < \mu + v(i)$ there exists $c_\lambda \in \overline{\mathbb{F}}_p$ such that*

$$\varphi_i * \varphi_\mu = \varphi_{\mu+v(i)} + \sum_{\lambda < \mu+v(i)} c_\lambda \varphi_\lambda.$$

Proof. By definition of the convolution, we have that

$$(9) \quad (\varphi_i * \varphi_\mu)(\alpha_\lambda^{-1}) = \sum_{y \in G/KZ} \varphi_i(y) \varphi_\mu(y^{-1} \alpha_\lambda^{-1}) = \sum_{y \in K\alpha_i^{-1}K/K} \varphi_i(y) \varphi_\mu(y^{-1} \alpha_\lambda^{-1}).$$

A set of coset representatives of $K\alpha_i^{-1}K/K$ is given by $\{z\alpha_i^{-1}\}$, where z runs over coset representatives of $K/(K \cap \alpha_i^{-1}K\alpha_i) \cong \text{GL}_n(k)/\overline{P}_i$. We saw above that such a set is given by $\bigcup_{w \in \mathcal{W}_i} \Lambda_w w$, where Λ_w is a set of matrices that may be taken to be lower triangular with ones on the diagonal. Let $x \in \Lambda_w$, and let $u = x^{-1}$. Then $(xw\alpha_i^{-1})^{-1} \alpha_\lambda^{-1} = \alpha_i w^{-1} u \alpha_\lambda^{-1} = w^{-1} (w\alpha_i w^{-1}) u \alpha_\lambda^{-1} = w^{-1} C$, where $C = (c_{ij})$ is a lower triangular matrix whose elements are given by

$$c_{ij} = u_{ij} \pi^{v_{w^{-1}(i)} - \lambda_j}.$$

Since $\lambda_1 \leq \dots \leq \lambda_n$, we see that by adding integer multiples of the rightmost column to the other columns, we can clear out the non-diagonal terms in the bottom row. Proceeding in this way, we find that $C = \text{diag}(c_{11}, c_{22}, \dots, c_{nn})C'$, where $C' \in K$ is some lower triangular matrix. Now let $s \in W$ be a permutation such that

$$v_{w^{-1}(s(1))} - \lambda_{s(1)} \geq v_{w^{-1}(s(2))} - \lambda_{s(2)} \geq \dots \geq v_{w^{-1}(s(n))} - \lambda_{s(n)}$$

and let $\kappa_j = v_{w^{-1}(s(1))} - \lambda_{s(1)} - (v_{w^{-1}(s(j))} - \lambda_{s(j)})$. Then $\kappa = (\kappa_1, \dots, \kappa_n) \in N$, and we have

$$(10) \quad (xw\alpha_i^{-1})^{-1} \alpha_\lambda^{-1} = w^{-1} s \alpha_\kappa^{-1} s^{-1} C' \in K\alpha_\kappa^{-1}KZ.$$

Observe that $\kappa = \kappa(\lambda, w)$ depends on λ and w , although we did not indicate this in the notation to avoid encumbering it further. It follows from the above that if $(\varphi_i * \varphi_\mu)(\alpha_\lambda^{-1}) \neq 0$, then $\mu = \kappa(\lambda, w)$ for some $w \in \mathcal{W}_i$. It is easy to see that $v_{w^{-1}(s(1))} - \lambda_{s(1)} \in \{0, 1\}$ and that we may assume that $\lambda(s(j)) = \lambda_j$ for all $1 \leq j \leq n$. Therefore $\kappa_j = \mu_j \geq \lambda_j - 1$. This proves the first part of the lemma.

Moreover, we find that $\kappa(\lambda, w) = \lambda - v(i)$ if and only if $w \in \mathcal{W}_i$ belongs to the class of the identity permutation. Then we may take $\Lambda_w = \{I_n\}$ to consist of the identity matrix. Therefore, if $\lambda = \mu + v(i)$, then the only non-zero term of the sum in (9) is the one corresponding to $y = \alpha_i^{-1}$, and this term is clearly equal to $U_i U_\mu = U_\lambda$. The second part of our claim follows. \square

Proposition 3.5. *As an $\overline{\mathbb{F}}_p$ -algebra, $\mathcal{H}(KZ, \sigma)$ is isomorphic to the ring $\overline{\mathbb{F}}_p[T_1, \dots, T_{n-1}]$ of polynomials in $n - 1$ variables.*

Proof. We will prove the equivalent statement that $\mathcal{H}_{KZ}(\sigma) \simeq \overline{\mathbb{F}}_p[\varphi_1, \dots, \varphi_{n-1}]$. We saw in Lemma 3.3 that $\mathcal{H}_{KZ}(\sigma)$ is commutative. By Lemma 3.2 the set $\{\varphi_\nu : \nu \in N\}$ spans $\mathcal{H}_{KZ}(\sigma)$, so it suffices to prove that each φ_ν is contained in the subalgebra generated by $\{\varphi_1, \dots, \varphi_{n-1}\}$.

We will argue by induction on ν with respect to the reverse lexicographical ordering. The claim is trivial for $\nu = (0, \dots, 0)$. Otherwise, let $i \leq n - 1$ be the largest integer such that $\nu_i = 0$. Then $\mu = \nu - \nu(i) \in N$. By Lemma 3.4, we know that

$$\varphi_\nu = \varphi_i * \varphi_\mu - \sum_{\lambda < \nu} c_\lambda \varphi_\lambda$$

for some $c_\lambda \in \overline{\mathbb{F}}_p$. Since $\mu < \nu$, our claim follows by induction. \square

Remark 3.6. We remark that using the decomposition of $V_\sigma^{\overline{N}_\nu}$ into weight spaces from [17], II.2.11, it is possible to obtain an explicit recursive formula expressing T_ν in terms of T_1, \dots, T_{n-1} . Such a formula would be useful for any work involving computations on the Hecke algebra $\mathcal{H}(KZ, \sigma)$.

3.4. Realization of irreducible admissible representations of G . As before, if σ is an irreducible $\overline{\mathbb{F}}_p$ -representation of $\mathrm{GL}_n(k)$, we inflate it to K via the map ω . Letting the uniformizer π act trivially, we obtain a representation of KZ that we continue to denote σ .

Lemma 3.7. *Let (ρ, V_ρ) be a smooth irreducible representation of KZ . Then there exist a unique unramified character $\chi : M^* \rightarrow \overline{\mathbb{F}}_p^*$ and irreducible representation σ of $\mathrm{GL}_n(k)$ such that $\rho = (\chi \circ \det) \otimes \sigma$.*

Proof. This is the same proof as in [3], Proposition 4. Since the uniformizer π lies in the center of KZ , the operator $\rho(\pi)$ is a KZ -module automorphism of V_ρ . Hence, by Schur's Lemma, π acts as a scalar $\lambda \in \overline{\mathbb{F}}_p$. So $\rho|_K$ is still irreducible. Since the congruence subgroup $\Gamma \subset K$ is a pro- p group, the space of invariants V_ρ^Γ is non-trivial. Since it is a normal subgroup, V_ρ^Γ is preserved by K , so by irreducibility it must be all of V_ρ . Thus $\rho|_K$ factors through $K/\Gamma \simeq \mathrm{GL}_n(k)$, so it is the inflation of a unique irreducible representation σ of $\mathrm{GL}_n(k)$. Let $\chi : F^* \rightarrow \overline{\mathbb{F}}_p^*$ be the unramified character defined by setting $\chi(\pi) = \lambda$. Clearly, $\rho = (\chi \circ \det) \otimes \sigma$. \square

Now let (ρ, V_ρ) be an irreducible $\overline{\mathbb{F}}_p$ -representation of $G = \mathrm{GL}_n(M)$ with central character. Since $I(1)$ is a pro- p -group, V_ρ has non-zero $I(1)$ -invariants. Clearly I preserves $V_\rho^{I(1)}$ and its action factors through the abelian quotient $I/I(1)$, so there is an eigenvector $v' \in V_\rho^{I(1)}$ on which I acts via a character $\varepsilon : I \rightarrow \overline{\mathbb{F}}_p^*$. The map $V_\varepsilon \rightarrow V_\rho|_I$ given by $1 \mapsto v'$ corresponds by Frobenius reciprocity to a non-zero element of $\mathrm{Hom}_K(\mathrm{ind}_I^K \varepsilon, V_\rho|_K)$. Let σ be an irreducible subrepresentation of $\mathrm{ind}_I^K \varepsilon$. Since ρ has a central character, there is a KZ -module homomorphism $(\chi \circ \det) \otimes \sigma \rightarrow V_\rho|_{KZ}$ for a suitable unramified character χ . Applying Frobenius reciprocity again, we get a non-zero, hence surjective, G -module homomorphism $\Phi : (\chi \circ \det) \otimes \mathrm{ind}_{KZ}^G \sigma \rightarrow V_\rho$.

The functor $V_\rho \mapsto V_\rho \otimes (\chi \circ \det)$ is an equivalence of categories from the category of G -modules to itself. Hence $\mathrm{End}_G((\chi \circ \det) \otimes \mathrm{ind}_{KZ}^G \sigma) \simeq \mathcal{H}(KZ, \sigma)$. Recall from Proposition 3.5 that $\mathcal{H}(KZ, \sigma)$ is isomorphic to a polynomial ring $\overline{\mathbb{F}}_p[T_1, \dots, T_{n-1}]$.

Theorem 3.8. *Let (ρ, V_ρ) be an irreducible and admissible $\overline{\mathbb{F}}_p$ -representation of $G = \mathrm{GL}_n(M)$ such that there exists a surjective G -module map $\Phi : \mathrm{ind}_{KZ}^G \sigma \rightarrow V_\rho$, where σ is an irreducible $\overline{\mathbb{F}}_p$ -representation of $\mathrm{GL}_n(k)$. Then there exist scalars $\lambda_1, \dots, \lambda_{n-1} \in \overline{\mathbb{F}}_p$ for which there is a G -module surjection*

$$\mathrm{ind}_{KZ}^G \sigma / (T_1 - \lambda_1, \dots, T_{n-1} - \lambda_{n-1}) \mathrm{ind}_{KZ}^G \sigma \rightarrow V_\rho.$$

Proof. We need to show that the space $\mathrm{Hom}_G(\mathrm{ind}_{KZ}^G \sigma, V_\rho)$, which is non-trivial by assumption, contains an eigenvector for the action of $\mathcal{H}(KZ, \sigma)$. Since this Hecke algebra is commutative by Lemma 3.3, it suffices to show that $\mathrm{Hom}_G(\mathrm{ind}_{KZ}^G \sigma, V_\rho)$ is finite-dimensional.

Recall that $\Gamma \subset K$ acts trivially on σ . By Frobenius reciprocity we have

$$\mathrm{Hom}_G(\mathrm{ind} KZ^G \sigma, V_\rho) \simeq \mathrm{Hom}_{KZ}(\sigma, V_\rho|_{KZ}) = \mathrm{Hom}_{KZ}(\sigma, V_\rho^\Gamma).$$

The subspace $V_\rho^\Gamma \subset V$ of invariants is finite-dimensional by admissibility of ρ , and the theorem follows. □

Remark 3.9. In view of the remark before the statement of Theorem 3.8, the theorem remains true if we replace $\mathrm{ind}_{KZ}^G \sigma$ by $(\chi \circ \det) \otimes \mathrm{ind}_{KZ}^G \sigma$ for an unramified character $\chi : M^* \rightarrow \overline{\mathbb{F}}_p^*$.

Remark 3.10. When $n = 2$, the previous result was proved by Barthel and Livné [3] without assuming admissibility. All irreducible complex representations of $\mathrm{GL}_2(M)$ are admissible, but it is not known whether this is true for $\overline{\mathbb{F}}_p$ -representations.

4. Towards the mod p local Langlands correspondence

In this section, all representations are over $\overline{\mathbb{F}}_p$. Let M/\mathbb{Q}_p be a finite extension, and maintain all the notation from the previous section. The Langlands philosophy predicts roughly that there is a natural bijection as follows:

$$\left\{ \begin{array}{l} n\text{-dimensional reps.} \\ \text{of } \text{Gal}(\overline{M}/M) \end{array} \right\} \xrightarrow{\pi} \left\{ \begin{array}{l} \text{certain smooth admissible} \\ \text{irreducible reps. of } \text{GL}_n(M) \end{array} \right\}.$$

For complex representations, such a bijection was established by Harris and Taylor [14]; on the left hand side in that case one considers Weil-Deligne representations rather than Galois representations. A mod l correspondence for p -adic fields, when $l \neq p$, has been given by work of Vignéras [33] and Emerton [9]. The $l = p$ case is considerably more involved. The bijection is then understood only when $n = 2$ and $M = \mathbb{Q}_p$. We briefly describe the mod p local Langlands correspondence in this case.

We just showed in Theorem 3.8 above that any irreducible admissible representation V of $\text{GL}_n(M)$ is a quotient of $\text{ind}_{KZ}^G \sigma / (T_1 - \lambda_1, \dots, T_{n-1} - \lambda_{n-1}) \text{ind}_{KZ}^G \sigma$ for some scalars $\lambda_1, \dots, \lambda_{n-1} \in \overline{\mathbb{F}}_p$. We say that V is *supersingular* if this is true for $\lambda_1 = \dots = \lambda_{n-1} = 0$. All $\overline{\mathbb{F}}_p$ -representations of $\text{GL}_1(M)$ are also considered supersingular.

The non-supersingular representations of $\text{GL}_2(M)$ were classified by Barthel and Livné for arbitrary M . When $M = \mathbb{Q}_p$, Breuil [5] proved that the $\text{ind}_{KZ}^G \sigma / (T_1) \text{ind}_{KZ}^G \sigma$ are all irreducible, and that the only isomorphisms between them are precisely those that are required to make the following definition well-defined.

Definition 4.1. Given an irreducible representation $\rho_p : \text{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p) \rightarrow \text{GL}_2(\overline{\mathbb{F}}_p)$, set $\pi(\rho_p) = \text{ind}_{KZ}^G \sigma / (T_1) \text{ind}_{KZ}^G \sigma$, where σ is a modular weight of ρ_p .

Recall that we may speak, somewhat abusively, of the modular weights of a local Galois representation, since the modular weights of $\rho : \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \rightarrow \text{GL}_2(\overline{\mathbb{F}}_p)$ are determined by the restriction of ρ to $G_p \simeq \text{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$.

Thus we have obtained a natural bijection between irreducible two-dimensional representations $\rho_p : \text{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p) \rightarrow \text{GL}_2(\overline{\mathbb{F}}_p)$ and supersingular representations of $\text{GL}_2(\mathbb{Q}_p)$. In all other cases, rather little is known. In particular, there appear to be far more supersingular representations of $\text{GL}_n(M)$ than there are irreducible n -dimensional representations of $\text{Gal}(\overline{M}/M)$. The supersingular representations are far from being classified; the $\text{ind}_{KZ}^G \sigma / (T_1, \dots, T_{n-1}) \text{ind}_{KZ}^G \sigma$ are in general of infinite length, even when $n = 2$, and their constituents are not understood. The first construction of supersingular representations of $\text{GL}_2(M)$ was by Paskunas [24].

Together with Breuil [6], they construct some representations, for p unramified in M , of the form that one would expect to find on the right-hand side of the local Langlands correspondence, but the picture is still very murky. In particular, it is not known how to characterize the representations of $GL_n(M)$ that appear in the image of the mod p local Langlands correspondence.

The strong connection between the mod p local Langlands correspondence and Serre’s conjecture that is evident in Definition 4.1 should hold in general. An argument of Emerton strongly suggests that for any irreducible $\rho : Gal(\overline{M}/M) \rightarrow GL_n(\overline{\mathbb{F}}_p)$, a non-zero surjection

$$\text{ind}_{KZ}^G \sigma / (T_1, \dots, T_{n-1}) \text{ind}_{KZ}^G \sigma \rightarrow \pi(\rho)$$

exists if and only if σ is a modular weight of ρ .

We conclude with some questions for future work on the representation theory of $GL_n(M)$. Recall that $S = \{\beta_i : 1 \leq i \leq n - 1\}$ is the standard set of simple roots of GL_n ; here $\beta_i = e_i - e_{i+1}$, where X_{ij} is the dual basis of $M_n(M)^*$ and $e_i = X_{ii}|_T$.

Given an irreducible representation V of $GL_n(M)$ and a surjective G -module homomorphism $\Phi : \text{ind}_{KZ}^G \sigma / (T_1 - \lambda_1, \dots, T_{n-1} - \lambda_{n-1}) \text{ind}_{KZ}^G \sigma \rightarrow V$, let $I_{V,\Phi} = \{i \in [1, n - 1] : \lambda_i = 0\}$. This induces a subset $J_{V,\Phi} \subset S$ consisting of the roots β_i such that $i \in I_{V,\Phi}$. Let $P_{V,\Phi} \subset GL_n$ be the standard parabolic subgroup associated to the set of roots $J_{V,\Phi}$. We conjecture that V arises from induction to $GL_n(M)$ of a representation of $P_{V,\Phi}(M)$.

More precisely, suppose that $P_{V,\Phi} \subset GL_n$ is the parabolic subgroup consisting of elements of the form

$$\begin{pmatrix} A_1 & * & & * \\ 0 & A_2 & & \\ & & \ddots & * \\ 0 & & 0 & A_R \end{pmatrix},$$

where $n = \sum_{r=1}^R n_r$ is a partition and $A_r \in GL_{n_r}$. For each $1 \leq r \leq R$, we let $G_r \subset G$ be the image of $GL_{n_r}(M)$ under the embedding

$$\begin{aligned} GL_{n_r}(M) &\rightarrow P_{V,\Phi} \subset GL_n(M) \\ A_r &\mapsto \begin{pmatrix} 1 & & & 0 \\ & A_r & & \\ & & \ddots & \\ 0 & & & 1 \end{pmatrix}. \end{aligned}$$

Let $Z_r \subset G_r$ be the center, and let $K_r = \text{GL}_{n_r}(\mathcal{O}) \subset G_r$ be a maximal compact subgroup. For convenience, we define $N_r = \sum_{s=1}^r n_s$. If V_r is a representation of $\text{GL}_{n_r}(M)$ for each $1 \leq r \leq R$, then we write $V_1 \otimes \cdots \otimes V_R$ for the inflation to $P_{V,\Phi}$ of the obvious representation of its Levi subgroup $\text{GL}_{n_1}(M) \times \cdots \times \text{GL}_{n_R}(M)$.

For each $1 \leq r \leq R$, we let σ_r be the following $\overline{\mathbb{F}}_p$ -representation of $\text{GL}_{n_r}(k)$, in the standard Weyl-module notation:

$$\sigma_r = F(a_{N_r-(n_r-1)}, a_{N_r-(n_r-2)}, \dots, a_{N_r}).$$

In other words, σ_r is generated as a $G_r(k)$ -module by the highest weight vector $v \in V_\sigma$. Note that if $n_r = 1$, then $\sigma_r : M^* \rightarrow \overline{\mathbb{F}}_p^*$ is just the character inflated from the map $k^* \rightarrow \overline{\mathbb{F}}_p^*$ defined by $x \mapsto x^{a_{N_r}}$.

Conjecture 4.2. Maintain the notations defined above. For each $1 \leq r \leq R$, there exists a supersingular irreducible $\overline{\mathbb{F}}_p$ -representation V_r of $\text{GL}_{n_r}(M)$ such that V is a subquotient of the parabolic induction

$$\text{ind}_{P_{V,\Phi}}^G (V_1 \otimes \cdots \otimes V_R).$$

Moreover, each V_r is a quotient of $\text{ind}_{K_r Z_r}^{G_r} \sigma_r / (T_1^r, \dots, T_{n_r-1}^r) \text{ind}_{K_r Z_r}^{G_r} \sigma_r$, where $T_1^r, \dots, T_{n_r-1}^r$ are the canonical generators of $\mathcal{H}(K_r Z_r, \sigma_r)$.

Some progress towards results of this type has been made recently by Florian Herzig. He is able to express the quotient $\text{ind}_{KZ}^G \sigma / (T_1 - \lambda_1, \dots, T_{n-1} - \lambda_{n-1}) \text{ind}_{KZ}^G \sigma$, for many representations σ of $\text{GL}_n(k)$, as a parabolic induction from the predicted parabolic subgroup of G .

The conjecture above expects the modular representation theory of $\text{GL}_n(M)$ to have the same general structure as the complex representation theory. The basic objects are the supersingular representations, and everything else can be built from them by parabolic induction. At the present time, we have no understanding of either the supersingular representations of $\text{GL}_n(F)$ or the structure of the parabolic inductions (except for the criterion of Rachel Ollivier [23] specifying exactly when the induction of a character from a Borel subgroup is irreducible), but these are beautiful and fruitful questions for research in the near future.

References

- [1] A. ASH, D. DOUD & D. POLLACK – “Galois representations with conjectural connections to arithmetic cohomology”, *Duke Math. J.* **112** (2002), no. 3, p. 521–579.
- [2] A. ASH & W. SINNOTT – “An analogue of Serre’s conjecture for Galois representations and Hecke eigenclasses in the mod p cohomology of $\mathrm{GL}(n, \mathbf{Z})$ ”, *Duke Math. J.* **105** (2000), no. 1, p. 1–24.
- [3] L. BARTHEL & R. LIVNÉ – “Irreducible modular representations of GL_2 of a local field”, *Duke Math. J.* **75** (1994), no. 2, p. 261–292.
- [4] D. BLASIUŞ & J. ROGAWSKI – “Galois representations for Hilbert modular forms”, *Bull. Amer. Math. Soc. (N.S.)* **21** (1989), no. 1, p. 65–69.
- [5] C. BREUIL – “Sur quelques représentations modulaires et p -adiques de $\mathrm{GL}_2(\mathbf{Q}_p)$. I”, *Compositio Math.* **138** (2003), no. 2, p. 165–188.
- [6] C. BREUIL & V. PASKUNAS – “Towards a modulo p Langlands correspondence for GL_2 ”, Preprint, 2007.
- [7] K. BUZZARD, F. DIAMOND & F. JARVIS – “On Serre’s conjecture for mod l Galois representations over totally real fields”, Preprint, available at <http://www.unet.brandeis.edu/~fdiamond/bdj12.pdf>.
- [8] B. EDIXHOVEN – “The weight in Serre’s conjectures on modular forms”, *Invent. Math.* **109** (1992), no. 3, p. 563–594.
- [9] M. EMERTON – “The local Langlands correspondence for $\mathrm{GL}_2(\mathbf{Q}_l)$ in p -adic families, and local-global compatibility for mod p and p -adic modular forms”, In preparation.
- [10] T. GEE – “Automorphic lifts of prescribed types”, Preprint, available at <http://arxiv.org/pdf/0810.1877>.
- [11] ———, “On weights of mod p Hilbert modular forms”, Preprint, 2008.
- [12] B. H. GROSS – “Modular forms (mod p) and Galois representations”, *Internat. Math. Res. Notices* (1998), no. 16, p. 865–875.
- [13] ———, “Algebraic modular forms”, *Israel J. Math.* **113** (1999), p. 61–93.
- [14] M. HARRIS & R. TAYLOR – *The geometry and cohomology of some simple Shimura varieties*, Annals of Mathematics Studies, vol. 151, Princeton University Press, Princeton, NJ, 2001, With an appendix by Vladimir G. Berkovich.
- [15] F. HERZIG – “The weight in a Serre-type conjecture for tame n -dimensional Galois representations”, Preprint, 2007.
- [16] ———, “A Satake isomorphism in characteristic p ”, In preparation, 2008.
- [17] J. C. JANTZEN – *Representations of algebraic groups*, second ed., Mathematical Surveys and Monographs, vol. 107, American Mathematical Society, Providence, RI, 2003.
- [18] C. KHARE – “Serre’s modularity conjecture: the level one case”, *Duke Math. J.* **134** (2006), no. 3, p. 557–589.
- [19] C. KHARE & J.-P. WINTENBERGER – “Serre’s modularity conjecture (I)”, Preprint.
- [20] ———, “Serre’s modularity conjecture (II)”, Preprint.
- [21] M. KISIN – “Modularity of 2-adic Barsotti-Tate representations”, Preprint, 2007.

- [22] ———, “Moduli of finite flat group schemes and modularity”, *Ann. of Math. (2)* (To appear).
- [23] R. OLLIVIER – “Critère d’irréductibilité pour les séries principales de $GL_n(F)$ en caractéristique p ”, *J. Algebra* **304** (2006), no. 1, p. 39–72.
- [24] V. PASKUNAS – “Unicity of types for supercuspidal representations of GL_N ”, *Proc. London Math. Soc. (3)* **91** (2005), no. 3, p. 623–654.
- [25] M. RAYNAUD – “Schémas en groupes de type (p, \dots, p) ”, *Bull. Soc. Math. France* **102** (1974), p. 241–280.
- [26] K. A. RIBET & W. A. STEIN – “Lectures on Serre’s conjectures”, Arithmetic algebraic geometry (Park City, UT, 1999), IAS/Park City Math. Ser., vol. 9, Amer. Math. Soc., Providence, RI, 2001, p. 143–232.
- [27] M. M. SCHEIN – “On modular weights of Galois representations”, *Math. Res. Lett.* **15** (2008), p. 537–542.
- [28] ———, “Weights in Serre’s conjecture for Hilbert modular forms: the ramified case”, *Israel J. Math.* **166** (2008), p. 369–391.
- [29] ———, “Weights of Galois representations associated to Hilbert modular forms”, *J. Reine Angew. Math* **622** (2008), p. 57–94.
- [30] ———, “Reduction modulo p of cuspidal representations and weights in Serre’s conjecture”, *Bull. London Math. Soc.* (To appear).
- [31] S. D. SMITH – “Irreducible modules and parabolic subgroups”, *J. Algebra* **75** (1982), no. 1, p. 286–289.
- [32] R. TAYLOR – “On Galois representations associated to Hilbert modular forms”, *Invent. Math.* **98** (1989), no. 2, p. 265–280.
- [33] M.-F. VIGNÉRAS – “La conjecture de Langlands locale pour $GL(n, F)$ modulo l quand $l \neq p$, $l > n$ ”, *Ann. Sci. École Norm. Sup. (4)* **34** (2001), no. 6, p. 789–816.

TOWARDS LANGLANDS CORRESPONDENCE OVER FUNCTION FIELDS FOR SPLIT REDUCTIVE GROUPS

Yakov Varshavsky

Mathematics Institute, The Hebrew University of Jerusalem

E-mail: vyakov@math.huji.ac.il

Abstract. In this note I will describe our joint work with David Kazhdan on the global Langlands correspondence over function fields for arbitrary split reductive groups.

Our main result asserts that for every pair (π, ω) , where π is a cuspidal representation of G one of whose local components is a cuspidal Deligne-Lusztig representation, and ω is a representation of the dual group, there exists a virtual Galois representation $\rho_{\pi, \omega}$, whose L -function equals the L -function of the pair (π, ω) .

1. Langlands correspondence

1.1. Notation.

- X is a smooth projective geometrically connected curve over a finite field \mathbb{F}_q of characteristic p ;
- $K = \mathbb{F}_q(X)$ is a field of rational functions of X ;
- $\Gamma_K = \text{Gal}(\overline{K}/K)$ is the absolute Galois group of K , $W_K \subset \Gamma_K$ is the Weil group of K , that is, W_K is a dense subgroup of Γ_K defined by $W_K := \{\sigma \in \Gamma_K \mid \sigma|_{\overline{\mathbb{F}}_q} \in \text{Frob}_{\overline{\mathbb{F}}_q}^{\mathbb{Z}}\}$.
- $\mathbb{A} = \mathbb{A}_K$ is the ring of adèles of K , that is, \mathbb{A} is the restricted product $\mathbb{A} = \prod'_{u \in X} K_u$, where K_u is the completion of K at u ;
- l is a prime number prime, different from p ;
- G is a split reductive group over \mathbb{F}_q ;

October 2008.

2000 Mathematics Subject Classification. 11R39 (11F70, 14G35, 14H60, 22E55).

- $\widehat{G} = {}^L G^0$ is the connected Langlands dual group over $\overline{\mathbb{Q}}_l$, that is, \widehat{G} is a reductive group over $\overline{\mathbb{Q}}_l$ such that the root datum of \widehat{G} is dual to the root datum of G . For example, $\widehat{\mathrm{GL}}_r = \mathrm{GL}_r, \widehat{\mathrm{SL}}_r = \mathrm{PGL}_r, \widehat{\mathrm{PGL}}_r = \mathrm{SL}_r, \widehat{\mathrm{Sp}}_{2n} = \mathrm{SO}_{2n+1}$ (for a more detailed discussion see, for example, [1]).

1.2. Langlands conjecture. For every automorphic representation π of $G(\mathbb{A})$ over $\overline{\mathbb{Q}}_l$ there exists a continuous homomorphism $\rho_\pi : W_K \rightarrow \widehat{G}(\overline{\mathbb{Q}}_l)$, which is unramified almost everywhere, such that $[\rho_\pi(\mathrm{Fr}_u)]^{ss} = c_u(\pi)$ for almost all places u of K .

1.3. Explanations.

(i) An *automorphic representation* is roughly speaking an irreducible subquotient of the space of smooth functions $G(K) \backslash G(\mathbb{A}) \rightarrow \overline{\mathbb{Q}}_l$ (for a precise definition, see [2]);

(ii) An irreducible admissible $\pi \in \mathrm{Rep}(G(\mathbb{A}))$ decomposes as a restricted tensor product $\otimes_{u \in X}^l \pi_u$, where

- each π_u is a smooth irreducible representation of $G(K_u)$;
- π_u is *unramified* for almost all u , that is, π_u is a representation $G(K_u) \rightarrow \mathrm{Aut}(V)$ such that $V^{G(\mathcal{O}_u)} \neq 0$ (see [8]).

(iii) The Satake isomorphism associates to each smooth irreducible unramified representation π_u of $G(K_u)$ a semi-simple conjugacy class $c_u(\pi) = c(\pi_u)$ in $\widehat{G}(\overline{\mathbb{Q}}_l)$. Namely, the representation π_u defines a homomorphism $\mathcal{H}(G(K_u), G(\mathcal{O}_u)) \rightarrow \overline{\mathbb{Q}}_l$, where $\mathcal{H}(G(K_u), G(\mathcal{O}_u))$ is the spherical Hecke algebra, hence by the Satake isomorphism (see for example [15, pp.18–24]) π_u defines an element of

$$\mathrm{Hom}(\overline{\mathbb{Q}}_l[\widehat{T}/W], \overline{\mathbb{Q}}_l) \cong \widehat{T}(\overline{\mathbb{Q}}_l)/W = \widehat{G}^{ss}(\overline{\mathbb{Q}}_l)/\mathrm{conjugacy}.$$

(iv) A homomorphism $\rho_\pi : W_K \rightarrow \widehat{G}(\overline{\mathbb{Q}}_l)$ is called to be *unramified* at $u \in X$, if $I_u \subset \mathrm{Ker} \rho_\pi$, where I_u is the inertia group. (Notice that the inertia group is defined up to a conjugacy). In this case, the conjugacy class $[\rho_\pi(\mathrm{Fr}_u)] \in \widehat{G}(\overline{\mathbb{Q}}_l)/\mathrm{conjugacy}$ is well-defined.

1.4. Known cases. $G = \mathrm{GL}_1$ (Class Field Theory), $G = \mathrm{GL}_2$ (Drinfeld [6], [7], [5]), $G = \mathrm{GL}_r$ (Lafforgue [14]).

1.5. Remarks. (a) The original conjecture of Langlands [15] goes into the opposite direction. Namely, Langlands conjectured the existence of an automorphic representation π (actually an L -packet of representations) associated to each continuous homomorphism $\rho_\pi : W_K \rightarrow \widehat{G}(\overline{\mathbb{Q}}_l)$, which is unramified almost everywhere.

(b) One expects that each $[\rho_\pi(\mathrm{Fr}_u)]$ is semi-simple. However, it is not known for groups other than GL_1 .

1.6. A weaker form of the Langlands conjecture. For every automorphic representation π of $G(\mathbb{A})$ and every algebraic representation ω of \widehat{G} , there exists a continuous representation $\rho_{\pi,\omega} := \omega \circ \rho_{\pi}$ of W_K over $\overline{\mathbb{Q}}_l$ (unramified almost everywhere) such that

$$(*) \quad \text{Tr}(\rho_{\pi,\omega}(\text{Fr}_u^m)) = \text{Tr}(\omega \circ c_u(\pi)^m) \text{ for almost all } u \text{ and all } m \geq 0.$$

1.7. The goal of this lecture is to describe a weaker form of Conjecture 1.6, which is proven in [9].

2. Formulation of the result

2.1. First Main Theorem. Let π be a cuspidal automorphic representation of $G(\mathbb{A})$ such that π_v is a cuspidal Deligne–Lusztig representation of $G(K_v)$ for some $v \in X$.

Then for every element $\omega \in \text{Rep}(\widehat{G})$ there exists a virtual representation $\rho_{\pi,\omega} \in \mathbb{Z}[\text{Rep}_{\overline{\mathbb{Q}}_l}(W_K)]$ satisfying (*). Moreover, $\rho_{\pi,\omega}$ belongs to $\mathbb{Z}[\text{Rep}_{\overline{\mathbb{Q}}_l}(\Gamma_K)]$, if the central character of π is of finite order.

2.2. Remark. More generally, our result applies when π_v belongs to a *cuspidal L-packet* of $G(K_v)$, that is, π_v belongs to a finite set $\Pi = (\tau_1, \dots, \tau_r)$ of smooth irreducible representations of $G(K_v)$ such that

- (i) each τ_i is cuspidal;
- (ii) each τ_i is induced from a compact mod center subgroup;
- (iii) the sum of the characters $\sum_{i=1}^r \chi(\tau_i)$ is stable on $G^{r,ss}(K_v)$ (the set of regular semi-simple elements).

Notice that it is conjectured in general and is known in most cases that assumption (i) implies assumption (ii) (see [11]).

2.3. Cuspidal Deligne–Lusztig representations. For a point x of the Bruhat–Tits building of $G(K_v)$, we denote by $G_x \subset G(K_v)$ the corresponding parahoric subgroup (see, for example, [16]). (In the case when G is semi-simple and simply connected, G_x is just the stabilizer $\text{Stab}_{G(K_v)}(x)$). Let $G_{x^+} \subset G_x$ be the pro-unipotent radical (that is, the maximal normal pro- p -subgroup of G_x), and let \overline{G}_x be the reductive group over \mathbb{F}_q such that $\overline{G}_x(\mathbb{F}_q) = G_x/G_{x^+}$.

By a *cuspidal Deligne–Lusztig representation* of $G(K_v)$ we mean the induced representation $\text{Ind}_{G_x \mathbb{Z}(G)(K_v)}^{G(K_v)}(\tau^0)$, where the restriction $\tau^0|_{G_x}$ is an inflation of an irreducible cuspidal Deligne–Lusztig representation of $\overline{G}_x(\mathbb{F}_q)$ (see [4]). (Notice that for such a τ^0 , the induced representation is irreducible and cuspidal).

2.4. Remark. Every cuspidal Deligne–Lusztig representation τ of $G(K_\nu)$ belongs to some cuspidal L -packet (at least if p is good for G , for example, if $p > 5$). Namely, there exists a finite set $\Pi = (\tau_1, \dots, \tau_r)$ of cuspidal Deligne–Lusztig representations $G(K_\nu)$ containing τ such that the sum $\sum_{i=1}^r \chi(\tau_i)$ is stable on $G^{r,ss}(K_\nu)$ (see [10]). Here $G^{r,ss} \subset G$ denotes the set of regular semi-simple elements of G .

2.5. Our strategy will be to construct a scheme \mathcal{X}_ω over K , equipped with an action of the group $G(\mathbb{A})$, and to “realize” $\rho_{\pi,\omega}$ as the “subquotient” of the virtual intersection cohomology with compact support $H_c^*(\mathcal{X}_\omega, \text{IC}(\overline{\mathbb{Q}}_l))$.

3. Moduli spaces of F -bundles

We fix an n -tuple $\overline{\omega} = (\omega_1, \dots, \omega_n) \in \text{Irr}(\widehat{G})^n$ such that $(\otimes_i \omega_i)|_{Z(\widehat{G})} = 1$. Following [17], we are going to associate to this tuple a certain Deligne–Mumford stack $\text{FBun}_{n,\overline{\omega}}$, which is locally of finite type over X^n .

3.1. Informal definition. $\text{FBun}_{n,\overline{\omega}}$ classifies a G -bundle \mathcal{G} on X , equipped with a “rational Frobenius-linear endomorphism” φ such that the “pole of φ ” is “bounded by $\overline{\omega}$ ”.

3.2. Formal definition. $\text{FBun}_{n,\overline{\omega}}$ classifies triples $(\mathcal{G}, \overline{x}, \varphi)$, where

- \mathcal{G} is a G -bundle on X ;
- \overline{x} is an n -tuple $(x_1, \dots, x_n) \in X^n$;
- φ is an isomorphism ${}^\tau \mathcal{G}|_{X \setminus \{x_1, \dots, x_n\}} \xrightarrow{\sim} \mathcal{G}|_{X \setminus \{x_1, \dots, x_n\}}$ of G -bundles such that ${}^\tau \mathcal{G} := \phi_q^* \mathcal{G}$, where ϕ_q is the arithmetic Frobenius automorphism, and

“the relative position $\text{rel}_{x_i}(\varphi({}^\tau \mathcal{G}), \mathcal{G})$ of $\varphi({}^\tau \mathcal{G})$ and \mathcal{G} at x_i is $\leq \omega_i$ ”.

3.3. Explanation. Explicitly, the last condition means the following: for every Weyl representation V_μ of G of the highest weight μ , we denote by \mathcal{E}_μ to be the corresponding vector bundle $G \backslash [G \times V_\mu]$ on X . Then φ gives rise to the isomorphism $\varphi_\mu : {}^\tau \mathcal{E}_\mu|_{X \setminus \{x_1, \dots, x_n\}} \xrightarrow{\sim} \mathcal{E}_\mu|_{X \setminus \{x_1, \dots, x_n\}}$ of vector bundles. Then the last condition of 3.2 means that

$$\varphi_\mu({}^\tau \mathcal{E}_\mu) \subset \mathcal{E}_\mu \left(\sum_{i=1}^n \langle \omega_i, \mu \rangle x_i \right) \text{ for all } \mu.$$

Here we identify $\omega_i \in \text{Irr}(\widehat{G})$ with the corresponding dominant weight of \widehat{G} (the highest weight), hence with the corresponding dominant coweight of G .

3.4. Examples. (a) Assume that ω_i is trivial for all i . Then $\text{FBun}_{n,\bar{\omega}}$ is essentially a product $\{\text{discrete stack } G(K)\backslash G(\mathbb{A})/G(\mathbb{O})\} \times X^n$.

(b) Assume that $G = \text{GL}_1$. In this case $\text{Irr}(\widehat{G}) = \mathbb{Z}$, hence $\bar{\omega}$ is an n -tuple $(k_1, \dots, k_n) \in \mathbb{Z}^n$ satisfying that $\sum_i k_i = 0$. Then we have a Cartesian square

$$\begin{array}{ccc} \text{FBun}_{n,\bar{\omega}} & \longrightarrow & \text{Pic } X \\ \downarrow & & \downarrow L \\ X^n & \xrightarrow{\mathcal{O}_{\bar{\omega}}} & \text{Pic}^0 X, \end{array}$$

where L is the Lang isogeny $\mathcal{L} \mapsto {}^t\mathcal{L} \otimes \mathcal{L}^{-1}$, while $\mathcal{O}_{\bar{\omega}}$ is the map $(x_1, \dots, x_n) \mapsto \mathcal{O}(\sum_i k_i x_i)$.

(c) Assume that $G = \text{GL}_r$, $n = 2$, ω_1 is the standard representation of $\widehat{G} = \text{GL}_r$, and ω_2 is the dual representation. In this case, $\text{FBun}_{n,\bar{\omega}}$ is the moduli space of F -sheaves (defined in [6]), considered by Drinfeld and Lafforgue.

3.5. Level structures. For each finite subscheme $D \subset X$ we denote by $\text{FBun}_{n,\bar{\omega},D}$ the moduli space of quadruples $(\mathcal{G}, \bar{x}, \varphi, \psi)$, where

- $(\mathcal{G}, \bar{x}, \varphi)$ belongs to $\text{FBun}_{n,\bar{\omega}} \times_{X^n} (X \setminus D)^n$, and
- and ψ is a trivialization $\psi: \mathcal{G}|_D \xrightarrow{\sim} G \times D$ commuting with φ (that is, $\psi \circ \varphi|_D = {}^t\psi$).

Then $\text{FBun}_{n,\bar{\omega},D}$ is a finite Galois covering of $\text{FBun}_{n,\bar{\omega}} \times_{X^n} (X \setminus D)^n$ with a Galois group $G(\mathcal{O}_D)$.

3.6. Basic virtual representation. Denote by $\mathcal{X}_{\bar{\omega}}$ the inverse limit of the $\text{FBun}_{n,\bar{\omega},D}$'s. In other words, $\mathcal{X}_{\bar{\omega}}$ classifies $(\mathcal{G}, \bar{x}, \varphi) \in \text{FBun}_{n,\bar{\omega}}$ together with “all level structures”. Then $\mathcal{X}_{\bar{\omega}}$ is a scheme over the generic point $\text{Spec } \mathbb{F}_q(X^n)$ of X^n equipped with an action of $G(\mathbb{A})$.

Moreover, $\mathcal{X}_{\bar{\omega}}$ is the inductive limit (with respect to open embeddings) of the projective limit (with respect to finite etale morphisms) of quasi-projective schemes. Therefore we can consider its virtual intersection cohomology with compact support

$$V_{\bar{\omega}} = \oplus_i (-1)^i V_{\bar{\omega}}^i := \oplus_i (-1)^i H_c^i(\mathcal{X}_{n,\bar{\omega}}, \text{IC}(\overline{\mathbb{Q}}_l)),$$

which is a smooth virtual representation of $G(\mathbb{A}) \times \Gamma_{\mathbb{F}_q(X^n)}$.

3.7. Example. In the example 3.4 (a) above, $\mathcal{X}_{\bar{\omega}}$ is essentially the discrete scheme $G(K)\backslash G(\mathbb{A})$, hence $V_{\bar{\omega}}$ is essentially the space of automorphic functions $\text{Func}(G(K)\backslash G(\mathbb{A}), \overline{\mathbb{Q}}_l)$. Thus $V_{\bar{\omega}}$ can be considered as a vast generalization of the space of automorphic functions.

4. The explicit form of the First Main theorem

4.1. Set-up. For simplicity, we will assume from now on that G is semi-simple and simply connected, hence \widehat{G} is adjoint. Then for every $\omega \in \text{Irr}(\widehat{G})$ we associate a scheme \mathcal{X}_ω over K , hence a smooth virtual representation V_ω of $G(\mathbb{A}) \times \Gamma_K$ (see 3.6).

We fix a cuspidal automorphic $\pi \in \text{Rep}(G(\mathbb{A}))$ of the form $\pi = \pi_\nu \otimes \pi^\nu$, where π^ν is an irreducible admissible representation of $G(\mathbb{A}^\nu)$ and π_ν belongs to a cuspidal L -packet $\Pi = (\tau_1, \dots, \tau_l)$ of Deligne–Lusztig representations of $G(K_\nu)$ (see 2.4).

4.2. Notation. (a) Set $m_{\text{cusp}}(\Pi \otimes \pi^\nu) := \sum_{\tau_i \in \Pi} m_{\text{cusp}}(\tau_i \otimes \pi^\nu)$, where

$$m_{\text{cusp}}(\tau_i \otimes \pi^\nu) := \dim \text{Hom}_{G(\mathbb{A})}(\tau_i \otimes \pi^\nu, \text{Cusp}(G(K) \backslash G(\mathbb{A})))$$

is the multiplicity of $\tau_i \otimes \pi^\nu$ in the cuspidal spectrum. By the assumption, $m_{\text{cusp}}(\pi_\nu \otimes \pi^\nu) \neq 0$, hence $m_{\text{cusp}}(\Pi \otimes \pi^\nu) \neq 0$.

(b) Similarly, we define a virtual representation $(V_\omega)_{\Pi \otimes \pi^\nu} := \oplus_{\tau_i \in \Pi} (V_\omega)_{\tau_i \otimes \pi^\nu}$ of Γ_K , where $(V_\omega)_{\tau_i \otimes \pi^\nu} := \text{Hom}_{G(\mathbb{A})}(\tau_i \otimes \pi^\nu, V_\omega^{SS})$ is the $(\tau_i \otimes \pi^\nu)$ -isotypical component of V_ω .

(c) Set $\rho_{\pi, \omega} := \frac{1}{m_{\text{cusp}}(\Pi \otimes \pi^\nu)} (V_\omega)_{\Pi \otimes \pi^\nu} \in \mathbb{Q}[\text{Rep}_{\overline{\mathbb{Q}}_l}(\Gamma_K)]$.

Our First Main Theorem has the following explicit form:

4.3. Second Main Theorem. The \mathbb{Q} -virtual representation $\rho_{\pi, \omega}$ belongs to $\mathbb{Z}[\text{Rep}_{\overline{\mathbb{Q}}_l}(\Gamma_K)]$ and satisfies $(*)$.

4.4. Remark. Our Second Main Theorem asserts that

$$\bigoplus_{\tau_i \in \Pi} (V_\omega)_{\tau_i \otimes \pi^\nu} \cong \bigoplus_{\tau_i \in \Pi} m_{\text{cusp}}(\tau_i \otimes \pi^\nu) \rho_{\pi, \omega}.$$

However, in general, $(V_\omega)_{\tau_i \otimes \pi^\nu}$ is not isomorphic to $m_{\text{cusp}}(\tau_i \otimes \pi^\nu) \rho_{\pi, \omega}$. For example, even in the case of SL_2 it might happens that $(V_\omega)_{\tau_i \otimes \pi^\nu} \neq 0$, while $m_{\text{cusp}}(\tau_i \otimes \pi^\nu) = 0$, and vice versa.

4.5. Conjecture. (a) $(V_\omega^i)_{\Pi \otimes \pi^\nu} = 0$ for each odd i , hence $\rho_{\pi, \omega}$ is a genuine representation and not just the virtual one.

(b) $(V_\omega^i)_{\Pi \otimes \pi^\nu} = 0$ for all $i \neq 0$, thus π satisfies the Ramanujan conjecture.

(c) There exists a homomorphism $\rho_\pi : \Gamma_K \rightarrow \widehat{G}(\overline{\mathbb{Q}}_l)$ such that $\rho_{\pi, \omega} = \omega \circ \rho_\pi$ for all ω .

5. Strategy of the proof of the Second Main Theorem

5.1. Reformulation. To show (*), we need to calculate the trace $Tr(\rho_{\pi,\omega}(\text{Fr}_u^m))$. For this it suffices to calculate $Tr(h \times \text{Fr}_u^m, V_\omega)$ for all $h \in \text{Hecke}(G(\mathbb{A}))$ of the form $h = h^u \otimes 1_{G(\mathcal{O}_v)}$ with $h^u \in \text{Hecke}(G(\mathbb{A}^u))$.

Problem. $\text{FBun}_{n,\omega}$ is not of finite type, hence V_ω is not an admissible representation of $G(\mathbb{A})$. Therefore the trace $\text{Tr}(h \times \text{Fr}_u^m, V_\omega)$ is not defined in general.

Solution. We set $(V_\omega)_\Pi = \oplus_{\tau_i \in \Pi} (V_\omega)_{\tau_i}$, where $(V_\omega)_{\tau_i} := \oplus_{\tau_i \in \Pi} \text{Hom}_{G(K_v)}(\tau_i, V_\omega^{SS})$ is the τ_i -isotypical component of V_ω .

Now we are ready to reformulate our Second Main Theorem.

5.2. Third Main Theorem.

- (a) $(V_\omega)_\Pi$ is an admissible representation of $G(\mathbb{A}^u)$;
- (b) For each $u \in X$, each $h^{v,u} \in \text{Hecke}(G(\mathbb{A}^{u,v}))$ and each m , we have an equality of traces

$$\text{Tr}(h^{v,u} \otimes 1_u \times \text{Fr}_u^m, (V_\omega)_\Pi) = \text{Tr}(h^{v,u} \otimes h_u(m, \omega) \otimes h_v, \text{Cusp}(G(K) \backslash G(\mathbb{A}))),$$

where h_v is an explicit linear combination of matrix coefficients of the τ_i 's, while $h_u(m, \omega) \in \mathcal{H}(G(K_u), G(\mathcal{O}_u))$ is the image under the Satake isomorphism of the function $[g \mapsto \text{Tr}(\omega(g^m))] \in \overline{\mathbb{Q}}_l[\widehat{T}/W]$.

5.3. It is standard that our Third Main Theorem implies that the \mathbb{Q} -virtual representation $\rho_{\pi,\omega}$ satisfies (*). It remains to show that $\rho_{\pi,\omega} \in \mathbb{Z}[\text{Rep}_{\overline{\mathbb{Q}}_l}(\Gamma_K)]$. For this we consider for each $n \in \mathbb{N}$ an n -tuple $\overline{\omega} = (\omega, \dots, \omega)$. Then it follows from the analog of the Third Main Theorem for $\overline{\omega}$ that

$$m_{\text{cusp}}(\Pi \otimes \pi^v)(\rho_{\pi,\omega} \boxtimes \dots \boxtimes \rho_{\pi,\omega}) \cong (V_{\overline{\omega}})_{\Pi \otimes \pi^v} \in \mathbb{Z}[\text{Rep}_{\overline{\mathbb{Q}}_l}((\Gamma_K)^n)]$$

for all $n \in \mathbb{N}$, which implies that $\rho_{\pi,\omega} \in \mathbb{Z}[\text{Rep}_{\overline{\mathbb{Q}}_l}(\Gamma_K)]$.

5.4. Sketch of the proof of the Third Main Theorem.

Step 1. Recall that $\tau_i = \text{Ind}_{H_i}^{G(K_v)}(\tau_i^0)$ for some finite dimensional representation (τ_i^0, W_i) of an open compact subgroup $H_i \subset G(K_v)$. Hence we can form an irreducible perverse sheaf \mathcal{L}_i on $H_i \backslash \mathcal{X}_\omega$ defined as the quotient under the diagonal action of $H_i \backslash [\text{IC}_\omega \otimes_{\overline{\mathbb{Q}}_l} W_i]$. By the adjointness of induction and restriction, the virtual representations $(V_\omega)_{\tau_i}$ and $\oplus_i (-1)^i H_c^i(H_i \backslash \mathcal{X}_{n,\overline{\omega}}, \mathcal{L}_i)$ of $G(\mathbb{A}^v) \times \Gamma_K$ are naturally isomorphic.

Step 2. Since τ_i is cuspidal, \mathcal{L}_i is supported on a quasi-compact open subscheme of $H_i \backslash \mathcal{X}_\omega$. (The proof of this assertion is very similar to the proof of the fact that

every cuspidal function is compactly supported). It follows that each $(V_\omega)_{\tau_i}$ and hence also $(V_\omega)_\Pi$ is admissible, completing the proof of (a).

Step 3. It is enough to prove the equality of traces for each m sufficiently large. Using Step 2 and a version of the Lefschetz Trace formula (which was proven in [18]) we get that $\text{Tr}(h^{v,u} \otimes 1_u \times \text{Fr}_u^m, (V_\omega)_\Pi)$ can be expressed in terms of “fixed points”.

Step 4. Since $(V_\omega)_\Pi$ is admissible, it will suffice to show the Third Main Theorem under the additional assumption that $h_{v'}$ is supported on $G^{r,ss}(K_{v'})$ for some $v' \neq v, u$.

Then using explicit calculations, similar to those of Kottwitz [12], the explicit description of IC_ω obtained in [17], and the fact that h_v is cuspidal, we see that the expression of Step 3 equals to

$$\sum O_{\gamma^{v,u}}(h^{v,u}) \cdot \text{TO}_{\delta_u}(f_u(1, \omega)) \cdot \text{WO}_{\gamma_v}(h_v),$$

where TO stands for twisted orbital integrals, and WO for weighted orbital integrals.

Step 5. Since $\sum_i \chi(\tau_i)|_{G(K_v^{r,ss})}$ is stable, we get that h_v is stable and cuspidal, hence the expression from Step 4 equals

$$\sum \text{SO}_{\gamma^{v,u}}(h^{v,u}) \cdot \text{STO}_{\delta_u}(f_u(1, \omega)) \cdot \text{SWO}_{\gamma_v}(h_v),$$

where SO , STO , and SWO means stable orbital integrals, stable twisted orbital integrals and stable weighted orbital integrals, respectively. Hence the latter expression is equal to

$$\sum \text{SO}_{\gamma^{v,u}}(h^{v,u}) \cdot \text{SO}_{\gamma_u}(f_u(m, \omega)) \cdot \text{SWO}_{\gamma_v}(h_v)$$

(by the fundamental lemma for stable base change [3],[13]), hence to

$$\sum O_{\gamma^{v,u}}(h^{v,u}) \cdot O_{\gamma_u}(f_u(m, \omega)) \cdot \text{WO}_{\gamma_v}(h_v)$$

(using again the fact that h_v is cuspidal and stable), hence finally to

$$\text{Tr}(h^{v,u} \otimes \text{BC}_n(\text{IC}_\omega) \otimes h_v, \text{Cusp}(G(K) \backslash G(\mathbb{A}))),$$

(using some simple form of the Arthur–Selberg trace formula).

References

[1] A. BOREL – “Automorphic L -functions”, Automorphic forms, representations and L -functions (Proc. Sympos. Pure Math., Oregon State Univ., Corvallis, Ore., 1977), Part 2, Proc. Sympos. Pure Math., XXXIII, Amer. Math. Soc., Providence, R.I., 1979, p. 27–61.

- [2] A. BOREL & H. JACQUET – “Automorphic forms and automorphic representations”, Automorphic forms, representations and L -functions (Proc. Sympos. Pure Math., Oregon State Univ., Corvallis, Ore., 1977), Part 1, Proc. Sympos. Pure Math., XXXIII, Amer. Math. Soc., Providence, R.I., 1979, With a supplement “On the notion of an automorphic representation” by R. P. Langlands, p. 189–207.
- [3] L. CLOZEL – “The fundamental lemma for stable base change”, *Duke Math. J.* **61** (1990), no. 1, p. 255–302.
- [4] P. DELIGNE & G. LUSZTIG – “Representations of reductive groups over finite fields”, *Ann. of Math. (2)* **103** (1976), no. 1, p. 103–161.
- [5] V. G. DRINFEL'D – “Cohomology of compactified moduli varieties of F -sheaves of rank 2”, *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)* **162** (1987), no. Avtomorfn. Funkts. i Teor. Chisel. III, p. 107–158, 189.
- [6] ———, “Moduli varieties of F -sheaves”, *Funktsional. Anal. i Prilozhen.* **21** (1987), no. 2, p. 23–41.
- [7] ———, “Proof of the Petersson conjecture for $GL(2)$ over a global field of characteristic p ”, *Funktsional. Anal. i Prilozhen.* **22** (1988), no. 1, p. 34–54, 96.
- [8] D. FLATH – “Decomposition of representations into tensor products”, Automorphic forms, representations and L -functions (Proc. Sympos. Pure Math., Oregon State Univ., Corvallis, Ore., 1977), Part 1, Proc. Sympos. Pure Math., XXXIII, Amer. Math. Soc., Providence, R.I., 1979, p. 179–183.
- [9] D. KAZHDAN & Y. VARSHAVSKY – “On the cohomology of the moduli spaces of F -bundles: stable cuspidal Deligne–Lusztig part”, in preparation.
- [10] ———, “Endoscopic decomposition of certain depth zero representations”, Studies in Lie theory, Progr. Math., vol. 243, Birkhäuser Boston, Boston, MA, 2006, p. 223–301.
- [11] J.-L. KIM – “Supercuspidal representations: an exhaustion theorem”, *J. Amer. Math. Soc.* **20** (2007), no. 2, p. 273–320 (electronic).
- [12] R. E. KOTTWITZ – “Shimura varieties and λ -adic representations”, Automorphic forms, Shimura varieties, and L -functions, Vol. I (Ann Arbor, MI, 1988), Perspect. Math., vol. 10, Academic Press, Boston, MA, 1990, p. 161–209.
- [13] J.-P. LABESSE – “Fonctions élémentaires et lemme fondamental pour le changement de base stable”, *Duke Math. J.* **61** (1990), no. 2, p. 519–530.
- [14] L. LAFFORGUE – “Chtoucas de Drinfeld et correspondance de Langlands”, *Invent. Math.* **147** (2002), no. 1, p. 1–241.
- [15] R. P. LANGLANDS – “Problems in the theory of automorphic forms”, Lectures in modern analysis and applications, III, Springer, Berlin, 1970, p. 18–61. Lecture Notes in Math., Vol. 170.
- [16] A. MOY & G. PRASAD – “Unrefined minimal K -types for p -adic groups”, *Invent. Math.* **116** (1994), no. 1–3, p. 393–408.
- [17] Y. VARSHAVSKY – “Moduli spaces of principal F -bundles”, *Selecta Math. (N.S.)* **10** (2004), no. 1, p. 131–166.
- [18] ———, “Lefschetz-Verdier trace formula and a generalization of a theorem of Fujiwara”, *Geom. Funct. Anal.* **17** (2007), no. 1, p. 271–319.

MAXIMAL SOBOLEV REGULARITY AT RADIAL POINTS

Ingo Witt

Mathematisches Institut, Georg-August Universität Göttingen
E-mail: iwitt@uni-math.gwdg.de

Abstract. We establish precise regularity results for solutions to pseudodifferential equations with real principal symbols near radial points, microlocally in the scale of H^s Sobolev spaces. There is the new phenomenon of *maximal Sobolev regularity*. We clarify this point by discussing how the two distinguished parametrices that one has near a radial point, where the pseudodifferential operator under study is of real principal type, extend into that radial point, and we provide an analytic formula for maximal Sobolev regularity in terms of the principal and subprincipal symbols.

1. Introduction

1.1. Statement of the main result. Let $P \in \Psi^m(X)$ be a classical pseudodifferential operator in a C^∞ manifold X . We shall assume that its principal symbol $p_m = \sigma^m(P)$ is real-valued and that P has simple characteristics, i.e., one has $dp_m \neq 0$ everywhere on $T^*X \setminus 0$. Such operators are well studied under the stronger assumption that P is of real principal type, in particular, one has $dp_m \not\parallel \alpha$, where $\alpha = \xi dx$ is the canonical one-form on $T^*X \setminus 0$. For instance, this class contains both elliptic and strictly hyperbolic differential operators. On the contrary, much less is known if the condition $dp_m \not\parallel \alpha$ is violated. This is despite the fact that such situations arise in applications as well, e.g., for differential operator of mixed elliptic-hyperbolic type. Characteristic points in $T^*X \setminus 0$ at which $dp_m \parallel \alpha$ holds are said to be radial for P . As turns out, near radial points the behavior of solutions u to the equation $Pu = f$

October 2008.

2000 Mathematics Subject Classification. Primary 35A17; Secondary 35D10, 58J40.

differs significantly from the behavior known for solutions to pseudodifferential equations of real principal type.

Radial points were previously studied by GUILLEMIN-SCHAEFFER [3] with regard to the propagation of C^∞ singularities, and by STERNIN [11], see also [9].

To state our result, note first that our arguments will be strictly microlocal, i.e., we shall work in the sheaf of microfunctions, neglecting C^∞ contributions, and Fourier integral operators, etc. will be defined only in open conic subsets of the cotangent bundle. We fix a radial point ζ^0 for P and make the assumption that no other ray near $\mathbb{R}_+\zeta^0$ consists of radial points. A further genericity assumption will be added below. Then the operator P is microlocally of real principal type near ζ^0 , with the direction $\mathbb{R}_+\zeta^0$ removed. Moreover, the characteristic set $p_m^{-1}(0)$ is connected near ζ^0 , and this remains true when the direction $\mathbb{R}_+\zeta^0$ is removed. According to DUISTERMAAT-HÖRMANDER [1], there are two distinguished parametrices \tilde{E}^\pm to P near ζ^0 without the direction $\mathbb{R}_+\zeta^0$, having the property that $\Delta \cup \Lambda^\pm$ are their respective wave front relations. Here, Δ denotes the diagonal in $(T^*X \setminus 0)^2$ near (ζ^0, ζ^0) , and Λ^\pm is the flow-out from $\Delta \cap (p_m \pi_j)^{-1}(0)$ under H_{p_m} in the positive (for Λ^+) and negative (for Λ^-) directions, respectively, with $\pi_j: (T^*X \setminus 0)^2 \rightarrow T^*X \setminus 0$ for $j = 1, 2$ being the projection onto the j th factor.

Remark. The parametrix construction of [1], see also [6], has been turned into a geometric (or symbolic) one by MELROSE-UHLMANN [8]. In fact, they characterized the kernels of these parametrices as so-called one-sided paired Lagrangian distributions. This construction was later improved by several people, see [4], [7].

We further set

$$(1.1) \quad s^0 = s_p^0(\zeta^0) = \frac{\Re(i\sigma_{\text{sub}}^{m-1}(P)(\zeta^0))}{\lambda} + \frac{m-1}{2},$$

where $\sigma_{\text{sub}}^{m-1}(P)(\zeta^0)$ is the subprincipal symbol of P evaluated at ζ^0 , $\lambda \neq 0$ is the factor appearing in the relation $H_p = \lambda R$ at ζ^0 , and m is the order of P .

Our main result states (upon an appropriate choice of orientation of the Hamilton flow of P):

Theorem 1.1. *Assume that no eigenvalue μ of $\mathfrak{A}_p^{m-1}(\zeta^0)$ satisfies $\Re\mu = \lambda/2$ (see Section 3.1 for the notation). Then the parametrices \tilde{E}^\pm extend into the direction $\mathbb{R}_+\zeta^0$ to provide parametrices E^\pm to P such that*

$$(1.2) \quad \text{WF}'(E^+) = (0 \times \mathbb{R}_+\zeta^0) \cup \text{WF}'(\tilde{E}^+), \quad \text{WF}'(E^-) = (\mathbb{R}_+\zeta^0 \times 0) \cup \text{WF}'(\tilde{E}^-).$$

Furthermore,

$$(1.3) \quad \begin{cases} E^+ : H^{s-m+1}(\zeta^0) \rightarrow H^s(\zeta^0), & \forall s > s^0, \\ E^- : H^{s-m+1}(\zeta^0) \rightarrow H^s(\zeta^0), & \forall s < s^0, \end{cases}$$

and these mapping properties are optimal.

Remark. In general, the parametrices E^\pm are not uniquely determined up to regularizing by condition (1.2). In fact, as shown in [3], there might be distributions $u \in H^{s^0-0}(\zeta^0) \setminus H^{s^0}(\zeta^0)$ such that $Pu \in C^\infty(\zeta^0)$, and even

$$\dim \{u \in H^{s^0-0}(\zeta^0) \mid Pu \in C^\infty(\zeta^0)\} / C^\infty(\zeta^0) = \infty$$

is possible. (If $u \in H^{s^0}(\zeta^0)$ and $Pu \in C^\infty(\zeta^0)$, then already $u \in C^\infty(\zeta^0)$.)

The gain of $m - 1$ derivatives in (1.3) is standard for parametrices to pseudodifferential operators of real principal type. However, near a radial point this remains valid only in a certain range of the Sobolev regularity s . Note that the relations $s > s^0$ and $s < s^0$ in (1.3) are *opposite to each other*. In particular, while $E^- f$ is defined for *all* right-hand sides f , but $E^- f$ even for $f \in C^\infty(\zeta^0)$ only belongs to $H^{s^0-0}(\zeta^0)$ and is in general not better, $E^+ f$ is defined only when $f \in H^{s^0+0}(\zeta^0)$, i.e., when f is sufficiently regular. This is further discussed in Section 2.2.

1.2. Interpretation of the result. The number $s^0 \in \mathbb{R}$ from (1.1) has to be interpreted as *maximal Sobolev regularity* for P at ζ^0 , for the following reasons:

- Given $f \in H^{s-m+1}(\zeta^0)$ for some $s > s^0$, there is one solution $u = E^+ f \in H^s(\zeta^0)$ to $Pu - f \in C^\infty(\zeta^0)$, and this solution u is essentially unique (i.e., unique modulo $C^\infty(\zeta^0)$),
- Any other solution u to $Pu - f \in C^\infty(\zeta^0)$ belongs to $H^{s^0-0}(\zeta^0)$, but does not belong to $H^{s^0}(\zeta^0)$.

1.3. Examples. Here we briefly discuss two examples.

1.3.1. The Euler operator. For the Euler operator

$$L = xD_x - \alpha, \quad x \in \mathbb{R},$$

where $\alpha \in \mathbb{C}$, one has two radial points $\zeta^{0,\pm} = (0, \pm 1) \in T^*\mathbb{R} \cong \mathbb{R}^2$. Solutions u to $Lu = 0$ are precisely the distributions that are homogeneous of degree $i\alpha$. There is one solution $(x \mp i0)^{i\alpha} \in C^\infty(\zeta^{0,\pm})$, while, for any solution u that is not a multiple of $(x \mp i0)^{i\alpha}$, one has $u \in H^{\Re(i\alpha)+1/2-0}(\zeta^{0,\pm})$, but $u \notin H^{\Re(i\alpha)+1/2}(\zeta^{0,\pm})$.

For instance, for the Heaviside function $H(x)$, one has $i\alpha = 0$ and one has $H(x) \in H_{\text{loc}}^{1/2-0}(\mathbb{R})$, while $H(x) \notin H^{1/2}(0)$. Similarly, for the Dirac distribution $\delta(x)$, one has $i\alpha = -1$ and $\delta(x) \in H^{-1/2-0}(\mathbb{R})$, while $\delta(x) \notin H^{-1/2}(0)$.

Remark. The two solutions $(x \mp i0)^{i\alpha}$ coincide when $i\alpha \in \mathbb{N}_0$, then yielding the C^∞ (polynomial) solution $u = x^{i\alpha}$ to $Lu = 0$.

In this example, maximal Sobolev regularity equals $s^0 = \Re(i\alpha) + 1/2$ for both radial points.

1.3.2. Mixed elliptic-hyperbolic equations. One motivation to study radial points comes from the theory of differential equations of mixed elliptic-hyperbolic type. For instance, let us consider a second-order differential operator

$$L = \partial_x^2 + k(x, y) \partial_y^2 + c(x, y) \partial_y$$

in a domain $\Omega \subseteq \mathbb{R}^2$, where $k \in C^\infty(\overline{\Omega}; \mathbb{R})$, $c \in C^\infty(\overline{\Omega}; \mathbb{C})$. Note that this operator L is elliptic when $k > 0$ and strictly hyperbolic when $k < 0$. Its principal symbol $\sigma^2(L)(x, y, \xi, \eta) = p(x, y, \xi, \eta) = -\xi^2 - k(x, y) \eta^2$ is real-valued. Assuming that L has only simple characteristics, i.e., one has $dk \neq 0$ whenever $k = 0$, then $\{k = 0\} \subset \overline{\Omega}$ is a C^∞ hypersurface, and the Hamiltonian vector field

$$H_p = -2\xi \frac{\partial}{\partial x} - 2k\eta \frac{\partial}{\partial y} + k_x \eta^2 \frac{\partial}{\partial \xi} + k_y \eta^2 \frac{\partial}{\partial \eta}$$

associated with p is radial at points where $k = k_x = 0$, $\xi = 0$. Maximal Sobolev regularity at such points is $s^0 = -\Re c(x, y) / k_y(x, y) + 3/2$. In particular, if one wants a solution u to $Lu = f$ to belong to $H^1(\Omega)$ (which is natural if one thinks of f as a function in $L^2(\Omega)$), one needs (in general at least) that $\Re c(x, y) / k_y(x, y) < 1/2$ holds at all points $(x, y) \in \Omega$, where $k = k_x = 0$, *irrespective of how regular the right-hand side f and possible boundary data are.* This is in strict contrast to the situation known to be valid for elliptic and strictly hyperbolic equations.

Acknowledgment. This is a preliminary report on joint work with MICHAEL RUZHANSKY of Imperial College London, U.K.

2. Facts from microlocal analysis

We recall some facts from microlocal analysis. For details, the reader is referred to [2], [6], [12].

2.1. Propagation of singularities. Let X, Y be C^∞ manifolds and $P: C_c^\infty(Y) \rightarrow \mathcal{D}'(X)$ be a sequentially continuous linear operator. By the Schwartz kernel theorem, such an operator possesses a distributional kernel $K_P \in \mathcal{D}'(X \times Y, 1_X \boxtimes \Omega_Y^1)$, with 1_X being the trivial line bundle over X and Ω_Y^1 being the density bundle over Y , such that

$$\langle Pu, v \rangle = \langle K_P, v \otimes u \rangle, \quad \forall v \in C_c^\infty(X, \Omega_X^1), u \in C_c^\infty(Y).$$

One formally writes

$$Pu(x) = \int_Y K_P(x, y) u(y).$$

Noting that $\text{WF}(K_P) \subseteq T^*(X \times Y) \setminus 0$ and identifying $T^*(X \times Y) \cong T^*X \times T^*Y$, then one introduces

$$\text{WF}'(P) = \{(\zeta, \zeta') \in T^*X \times T^*Y \mid (\zeta, -\zeta') \in \text{WF}(K_P)\}$$

as well as

$$\begin{aligned} \text{WF}_X(P) &= \{\zeta \in T^*X \setminus 0 \mid (\zeta, 0_y) \in \text{WF}(K_P) \text{ for some } y \in Y\}, \\ \text{WF}'_Y(P) &= \{\zeta' \in T^*Y \setminus 0 \mid (0_x, -\zeta') \in \text{WF}(K_P) \text{ for some } x \in X\}. \end{aligned}$$

Proposition 2.1. *Let $u \in \mathcal{D}'(Y)$ satisfy $\text{WF}(u) \cap \text{WF}'_Y(P) = \emptyset$ and let the canonical projection $\text{supp } K_P \cap (X \times \text{supp } u) \rightarrow X$ be proper. Then $Pu \in \mathcal{D}'(X)$ can uniquely be defined in such a manner that whenever $\{u^m\} \subset C_c^\infty(Y)$ is a sequence satisfying $\text{supp } u^m \subseteq Z$ for all m , where the projection $\text{supp } K_P \cap (X \times Z) \rightarrow X$ is proper, and $u^m \rightarrow u$ as $m \rightarrow \infty$ in $\mathcal{D}'(Y)$, then $Pu^m \rightarrow Pu$ as $m \rightarrow \infty$ in $\mathcal{D}'(X)$. Moreover,*

$$\text{WF}(Pu) \subseteq \text{WF}'(P) \circ \text{WF}(u) \cup \text{WF}_X(P).$$

Remark. If $\text{WF}_X(P) = \emptyset$ and $\text{WF}'_Y(P) = \emptyset$, then P extends to a continuous map $P: \mathcal{E}'(Y) \rightarrow \mathcal{D}'(X)$ satisfying $P: C_c^\infty(Y) \rightarrow C^\infty(X)$, and

$$\text{WF}(Pu) \subseteq \text{WF}'(P) \circ \text{WF}(u).$$

The latter is just a statement on the propagation of singularities.

In particular, this remark applies to operators of real principal type.

2.2. The wave front relation near a radial point. Let $P \in \Psi^m(X)$ be as above and $\zeta^0 \in T^*X \setminus 0$ be characteristic for P .

Definition 2.2. $\zeta^0 \in T^*X \setminus 0$ is called a *radial point* for P if dp and the canonical 1-form $\alpha = \xi dx$ are collinear at ζ^0 , i.e.,

$$dp + \lambda \xi dx = 0 \quad \text{at } \zeta^0$$

for some $\lambda \in \mathbb{R}$, $\lambda \neq 0$. An equivalent condition states that H_p and the *radial vector field* $R = \xi \partial / \partial \xi$ are collinear at ζ^0 , i.e., $H_{p_m} = \lambda \xi \frac{\partial}{\partial \xi}$ holds at ζ^0 .

Remark. We shall write $\lambda_p^{m-1}(\zeta^0) = \lambda$. With ζ^0 being radial, all points in the direction of ζ^0 are radial, and then

$$\lambda_p^{m-1}(v\zeta^0) = v^{m-1} \lambda_p^{m-1}(\zeta^0), \quad v > 0.$$

In conjunction with Proposition 2.1, an explanation for the apparently strange behavior of parametrices near a radial point ζ^0 is provided by the next result:

Lemma 2.3. *Upon an appropriate choice of orientation of the Hamilton flow of P , the closure of $WF'(\tilde{E}^\pm)$ in $T^*(X \times X) \setminus 0$ contains the directions $0_{x^0} \times \mathbb{R}_+ \zeta^0$ and $\mathbb{R}_+ \zeta^0 \times 0_{x^0}$, respectively, where x^0 is the image of ζ^0 under the canonical projection $T^*X \setminus 0 \rightarrow X$.*

Proof. For the sake of brevity, assume that $m = 1$, $\lambda = \lambda_p^0(\zeta^0) > 0$. Then, in local coordinates (x, ξ) near $\zeta^0 = (x^0, \xi^0)$, the Hamilton flow is given by

$$(2.1) \quad \dot{x} = \frac{\partial p_1}{\partial \xi}, \quad \dot{\xi} = -\frac{\partial p_1}{\partial x}.$$

In particular, the solution to (2.1) with $(x(0), \xi(0)) = (x^0, \xi^0)$ is $(x(t), \xi(t)) = (x^0, e^{\lambda t} \xi^0)$, and $(x(t), \xi(t)) \rightarrow (x^0, 0)$ as $t \rightarrow -\infty$. Hence, the result follows from the continuous dependence of the solutions to (2.1) on the initial conditions. \square

Therefore, in the notation of Section 2.1, one necessarily has $WF'_Y(E^+) \supseteq \mathbb{R}_+ \zeta^0$ (where now $Y = X$) and $WF_X(E^-) \supseteq \mathbb{R}_+ \zeta^0$ for any extensions of \tilde{E}^\pm as described in Theorem 1.1. As turns out, one actually has equality in both instances, more precisely, one has $WF'_Y(E^+) = \mathbb{R}_+ \zeta^0$, $WF_X(E^+) = \emptyset$ and $WF'_Y(E^-) = \emptyset$, $WF_X(E^-) = \mathbb{R}_+ \zeta^0$. This observation already indicates that it will not be possible to define $E^+ f$ for all $f \in \mathcal{E}'$, whereas, in general, $E^- f \notin C^\infty$ even for $f \in C_c^\infty$. This heuristics is made precise in Theorem 1.1.

3. Analytic properties of radial points

In this section, we prepare for the later reduction to normal forms and show that the number $s^0 = s_p^0(\zeta^0)$ as introduced in (1.1) behaves under multiplication of P by elliptic pseudodifferential operators as well as under conjugation of P by elliptic Fourier integral operators in a way that is consistent with its interpretation as maximal Sobolev regularity. One of the difficulties in dealing with radial points is that simple normal forms are only available in generic cases. This is related to the well-known fact that C^∞ vector fields cannot always be linearized near their stationary points.

3.1. Classification of radial points. We now study the question to what extent the operator P can be simplified near a radial point ζ^0 by conjugating it with an elliptic Fourier integral operator (e.g., think of a change of coordinates in X). As the covector α corresponds to the vector R under the identification of vectors and covectors by means of the symplectic form $\sigma = d\xi \wedge dx$, it is clear that the factor λ is an invariant of such a transformation. Another one is the *subprincipal symbol* of P at ζ^0 ,

$$\sigma_{\text{sub}}^{m-1}(P)(\zeta^0) = p_{m-1}(x^0, \xi^0) - \frac{1}{2i} \sum_{j=1}^n \frac{\partial^2 p_m}{\partial x_j \partial \xi_j}(x^0, \xi^0).$$

Here $p_m(x, \xi) + p_{m-1}(x, \xi) + \dots$ is the asymptotic expansion of the full symbol of P when written out in coordinates (x, ξ) in $T^*X \setminus 0$ associated with local coordinates x in X .

Lemma 3.1. *Let J be an elliptic Fourier integral operator with underlying canonic transformation $\chi: \Gamma \rightarrow \chi(\Gamma)$, where Γ is an open conic neighborhood of ζ^0 in $T^*X \setminus 0$. Then*

$$\sigma_{\text{sub}}^{m-1}(JPJ^{-1})(\chi(\zeta^0)) = \sigma_{\text{sub}}^{m-1}(P)(\zeta^0),$$

where as usual J^{-1} denotes a parametrrix to J .

Proof. This result which is well-known in case ζ^0 is a double characteristic point (i.e., when $\lambda = 0$) follows along the lines of the double characteristic case. See, e.g., [12, Chapter XV, §2]. \square

There is, however, a third invariant. To define it, we first look at the situation of a double characteristic point ζ^1 for P , i.e., a point $\zeta^1 \in T^*X \setminus 0$ at which $p_m = 0$, $dp_m = 0$. At such a point the fundamental matrix $\begin{pmatrix} \frac{\partial^2 p_m}{\partial x \partial \xi} & \frac{\partial^2 p_m}{\partial \xi^2} \\ -\frac{\partial^2 p_m}{\partial x^2} & -\frac{\partial^2 p_m}{\partial \xi \partial x} \end{pmatrix}$ (which is a symplectic version of the Hessian of p_m) is invariantly defined as the linearization

$$\delta H_{p_m}: T_{\zeta^1}(T^*X \setminus 0) \longrightarrow T_{\zeta^1}(T^*X \setminus 0)$$

of H_{p_m} at ζ^1 . This approach does not work in the case of a radial point ζ^0 , but we have the following substitute: We consider the vector field

$$H_{p_m} + q(\zeta)R$$

instead, where $q: T^*X \setminus 0 \rightarrow \mathbb{R}$ is homogeneous of degree $m - 1$ and $q(\zeta^0) = -\lambda$ (and arbitrary otherwise).

Proposition 3.2. *The linearization*

$$\delta(H_{p_m} + qR): T_{\zeta^0}(T^*X \setminus 0) \longrightarrow T_{\zeta^0}(T^*X \setminus 0)$$

leaves the kernel α^\perp of the canonical 1-form α as well as the direction \mathbb{R}_+R of the radial vector invariant. The induced map in $\widehat{E}_{\zeta^0} = \alpha^\perp / \mathbb{R}_+R$ is independent of q , and it is of the form

$$-\lambda/2 + \mathfrak{A}_p^{m-1}(\zeta^0)$$

where $\mathfrak{A}_p^{m-1}(\zeta^0) \in \text{sp}(\widehat{E}_{\zeta^0})$.

Proof. A straightforward computation. \square

Here, $\text{sp}(\widehat{E}_{\widehat{\zeta}^0})$ is the Lie algebra of infinitesimal linear symplectic transformation. (Hence, $-\lambda/2 + \mathfrak{A}_p^{m-1}(\zeta^0) \in \text{csp}(\widehat{E}_{\widehat{\zeta}^0})$, with conformal factor $-\lambda/2$.) To justify the last statement of Proposition 3.2, we need to note that the quotient $\alpha^\perp/\mathbb{R}_+R$ is canonically identified with the *contact hyperplane* $\widehat{E}_{\widehat{\zeta}^0} \subset T_{\widehat{\zeta}^0}(S^*X)$ for the canonical contact structure of the cosphere bundle S^*X , with $\widehat{\zeta}^0$ being the image of ζ^0 under the canonical projection $T^*X \setminus 0 \rightarrow S^*X$. In particular, $\widehat{E}_{\widehat{\zeta}^0}$ is a symplectic vector space.

Remark. In case $m = 1$, the Hamiltonian vector H_{p_m} is homogeneous of degree 0, so it induces a contact vector field \widehat{H}_{p_m} on S^*X which has a stationary point at $\widehat{\zeta}^0$. In this case, $-\lambda/2 + \mathfrak{A}_p^0(\zeta^0)$ is induced by the linearization of \widehat{H}_{p_m} at $\widehat{\zeta}^0$.

Note that the eigenvalues of $\mathfrak{A}_p^{m-1}(\zeta^0) \in \text{sp}(\widehat{E}_{\widehat{\zeta}^0})$ come in groups of either two, namely $\mu, -\mu$ if $\mu \in \mathbb{R} \cup i\mathbb{R}$, or four, namely $\mu, -\mu, \bar{\mu}, -\bar{\mu}$ if $\mu \notin \mathbb{R} \cup i\mathbb{R}$. We now make a genericity assumption which is of non-resonance type ruling out certain integral combinations of eigenvalues of $\mathfrak{A}_p^{m-1}(\zeta^0)$:

Definition/Condition 3.3. A radial point ζ^0 is called *generic* if the eigenvalues of $\mathfrak{A}_p^{m-1}(\zeta^0)$ are simple. Writing them as $\mu_1, -\mu_1, \dots, \mu_{n-1}, -\mu_{n-1}$, it is further required that the equation

$$\frac{m\lambda}{2} = m_1 \Re \mu_1 + \dots + m_{n-1} \Re \mu_{n-1}$$

admits no solution $(m, m_1, \dots, m_{n-1}) \in \mathbb{Z}^n$ with $m \neq 0$. For an explanation, see [3].

Proposition 3.4. Let $P, P' \in \Psi^m(X)$ and let ζ^0 and ζ'^0 be radial points for P and P' , respectively. If P near ζ^0 and P' near ζ'^0 are (microlocally) equivalent under conjugation by a zeroth-order elliptic Fourier integral operator, then

- (i) $\lambda_p^{m-1}(\zeta^0) = \lambda_{p'}^{m-1}(\zeta'^0)$,
- (ii) $\sigma_{\text{sub}}^{m-1}(P)(\zeta^0) = \sigma_{\text{sub}}^{m-1}(P')(\zeta'^0)$,
- (iii) There is a linear symplectic map $V: \widehat{E}_{\widehat{\zeta}^0} \rightarrow \widehat{E}_{\widehat{\zeta}'^0}$ such that

$$\mathfrak{A}_{p'}^{m-1}(\zeta'^0) = V \mathfrak{A}_p^{m-1}(\zeta^0) V^{-1}.$$

If ζ^0 and ζ'^0 are generic, then these conditions are also sufficient.

Proof. That (i), (iii) are necessary was shown in [3]. That (ii) is likewise necessary as well as sufficiency of all three conditions together in the generic case was added in [10]. □

As a corollary, we further have from [3]:

Proposition 3.5. (a) *By order reduction and conjugation by a zeroth-order elliptic Fourier integral operator, P can generically be put into a normal form near ζ^0 , where $m = 2$, $\zeta^0 = (0; e_n)$, and*

$$(3.1) \quad P = \langle A \partial_{x'}, \partial_{x'} \rangle + 2 \langle B x', \partial_{x'} \rangle \partial_{x_n} + \langle C x', x' \rangle \partial_{x_n}^2 + \lambda x_n \partial_{x_n}^2 + \alpha \partial_{x_n}.$$

Here $x = (x', x_n)$, $A, B, C \in M(n-1; \mathbb{R})$ are constant $(n-1) \times (n-1)$ matrices, $A = A^T$, $C = C^T$, $\lambda \in \mathbb{R}$, and $\alpha \in \mathbb{C}$.

(b) *If, in addition, $\mathfrak{A}_P^{m-1}(\zeta^0)$ has no purely imaginary eigenvalues, then a first-order normal form is available. See (4.1) below.*

Remark. (a) Assigning weight 1 to x' and weight 2 to x_n , the operator P in (3.1) becomes *homogeneous of degree 2*, including the first-order term. This observation simplifies the analysis of this operator P considerably.

(b) Not all of the normal forms listed in (3.1) are inequivalent.

3.2. Invariance properties of s^0 . We now check that formula (1.1) for $s_P^0(\zeta^0)$ is invariant under multiplication by elliptic pseudodifferential operators from both sides (with order shift when multiplying from the right). Invariance under conjugation by elliptic Fourier integral operators already follows from Proposition 3.4. As Fourier integral operators act continuously in the scale H^s of $(L^2$ -based) Sobolev spaces, this is consistent with the interpretation of s^0 as maximal Sobolev regularity.

Lemma 3.6. *Let $Q_j \in \Psi^{m_j}$ for $j = 0, 1$ be elliptic pseudodifferential operators with real principal symbols. Then*

$$s_{Q_1 P Q_0}^0(\zeta^0) = s_P^0(\zeta^0) + m_0.$$

Proof. Let q_0 be the principal symbol of Q_0 . Then taking $p_m(\zeta^0) = 0$ into account, a straightforward computation shows that

$$\sigma_{\text{sub}}^{m+m_0-1}(PQ_0) = \sigma_{\text{sub}}^{m-1}(P)q_0 + \frac{1}{2i} \{p_m, q_0\} \text{ at } \zeta^0,$$

where $\{, \}$ denotes the Poisson bracket. But

$$\{p_m, q_0\} = \frac{\partial p_m}{\partial \xi} \frac{\partial q_0}{\partial x} - \frac{\partial p_m}{\partial x} \frac{\partial q_0}{\partial \xi} = \lambda \xi \frac{\partial q_0}{\partial \xi} = \lambda m_0 q_0 \text{ at } \zeta^0$$

by assumption and Euler's relation, where $\lambda = \lambda_P^{m-1}(\zeta^0)$, so that

$$\begin{aligned} s_{PQ_0}^0 &= \frac{\Re \left(i \sigma_{\text{sub}}^{m+m_0-1}(PQ_0) \right)}{\lambda q_0} + \frac{m+m_0-1}{2} \\ &= \frac{\Re \left(i \sigma_{\text{sub}}^{m-1}(P)q_0 \right) + \lambda m_0 q_0 / 2}{\lambda q_0} + \frac{m+m_0-1}{2} = s_P^0 + m_0 \text{ at } \zeta^0. \end{aligned}$$

The proof of the relation $s_{Q_1 P}^0(\zeta^0) = s_P^0(\zeta^0)$ is similar. □

Putting everything together, we obtain:

Corollary 3.7. *The number $s_p^0(\zeta^0)$ transforms in a way that is consistent with its interpretation as maximal Sobolev regularity.*

4. Sketch of proof of the main theorem

4.1. Reduction to normal form. Here we sketch the proof of Theorem 1.1 only in a particular case, namely that $\Re\mu \neq 0$ holds for all eigenvalues μ of $\mathfrak{A}_p^{m-1}(\zeta^0)$, and that the radial point ζ^0 is generic. In this case, a first-order normal form for the operator P near ζ^0 is available. General generic radial points with second-order normal forms are treated by a different parametrix construction, which, in fact, is inspired by calculations in [3, §7], while non-generic radial points are eventually dealt with by a limiting procedure.

Remark. An exceptional case occurs when $\Re\mu = \lambda/2$ for some eigenvalue μ of $\mathfrak{A}_p^{m-1}(\zeta^0)$. Then, for instance in case $m = 1$, the topological properties of the flow of \widehat{H}_{p_1} near the stationary point $\widehat{\zeta}^0$ are unstable under small perturbations of P . See the remark following Proposition 3.2.

By order reduction and conjugating by an elliptic zeroth-order Fourier integral operator, we can now assume that $\zeta^0 = (0, e_n)$, $m = 1$, $\lambda = -1$, and $\mathfrak{A}_p^{m-1}(\zeta^0) = \begin{pmatrix} B & 0 \\ 0 & -B^T \end{pmatrix}$ for some $B \in M(n-1; \mathbb{R})$, where in addition $B + B^T > 0$. This means that $P = \langle Bx', D_{x'} \rangle + x_n D_{x_n} + \alpha$, where $\alpha \in \mathbb{C}$. We then write

$$(4.1) \quad P = \langle Ax, D_x \rangle + \alpha,$$

with $A = \begin{pmatrix} B & 0 \\ 0 & 1 \end{pmatrix} \in M(n; \mathbb{R})$ satisfying $A + A^T > 0$. For constructing the parametrix E^+ , we further assume that $s^0 < 0$, i.e., $\Re(i\alpha) - \text{tr}A/2 > 0$. This is achieved with the help of Lemma 3.6. In particular, $\Re(i\alpha) > 0$. We are then going to show that $E^+ : H^s \rightarrow H^s$ for all $s \geq 0$.

Remark. In order to complete the proof, one still needs to show that $E^+ : H^s(\zeta^0) \rightarrow H^s(\zeta^0)$ holds also for $s^0 < s < 0$. But having fixed such an s , one considers the operator $P' = Q_1 P Q_0$, where $Q_0 \in \Psi^{-s}$, $Q_1 \in \Psi^s$ are elliptic, and gets a parametrix $Q_1^{-1} E'^+ Q_0^{-1}$ for P' that is $H^s(\zeta^0)$ -continuous, with E'^+ being a parametrix for P' as constructed before. These two parametrices for P differ by an operator K with $\text{WF}'(K) \subseteq 0_0 \times \mathbb{R}\zeta^0$ and $\text{WF}'(PK) = \text{WF}'(KP) = \emptyset$. This provides enough information to conclude that the kernel of K is an isotropic distribution for the isotropic submanifold $0_0 \times (-\mathbb{R}\zeta^0) \subset T^*(X \times X)$ and, moreover, it belongs to $C^\infty(\zeta^0) \widehat{\otimes} H^{-s^0-0}(-\zeta^0)$. Thus, the difference $E^+ - Q_1^{-1} E'^+ Q_0^{-1}$ maps $H^{s^0+0}(\zeta^0)$ to $C^\infty(\zeta^0)$. Hence, E^+ is $H^s(\zeta^0)$ -continuous along with $Q_1^{-1} E'^+ Q_0^{-1}$.

4.2. Construction of E^+ . Let

$$E^+ f(x) = i \int_0^1 t^{i\alpha} f(t^A x) \frac{dt}{t}, \quad f \in C^\infty(\mathbb{R}^n).$$

For $\Re A > 0$, one has that $\lim_{t \rightarrow +0} f(t^A x) = f(0)$, and the integral on the right-hand side is well-defined.

Lemma 4.1. E^+ is a two-sided inverse to P as an operator in $C^\infty(\mathbb{R}^n)$.

Proof. Indeed, one has

$$\begin{aligned} PE^+ f(x) &= \int_0^1 t^{i\alpha} (\langle At^A x, (\partial_x f)(t^A x) \rangle + i\alpha f(t^A x)) \frac{dt}{t} \\ &= \int_0^1 \frac{d}{dt} (t^{i\alpha} f(t^A x)) dt = f(x). \end{aligned}$$

The proof that $E^+ P = \text{id}_{C^\infty(\mathbb{R}^n)}$ is similar. □

Lemma 4.2. The operator E^+ is $H^s(\mathbb{R}^n)$ -continuous for all $s \geq 0$.

Proof. A straightforward calculation shows that

$$\|f(t^A x)\|_s \leq C_s t^{-\text{tr}A/2} \|f\|_s, \quad t \in (0, 1],$$

for any $s \geq 0$ with some constant $C_s > 0$. One then gets

$$\|E^+ f\|_s \leq C_s \int_0^1 t^{\Re(i\alpha) - \text{tr}A/2 - 1} dt \|f\|_s = C'_s \|f\|_s$$

because of $\Re(i\alpha) - \text{tr}A/2 > 0$. □

Lemma 4.3. The operator E^+ coincides with \tilde{E}^+ away from radial directions. (Here coincidence is meant in the microlocal sense, i.e., up to regularizing.)

Proof. For E^+ is a parametrix to P , its wave front relation $\text{WF}'(E^+)$ is contained in $\Delta_{T^*X \setminus 0} \cup \{(x, \xi, t^A x, t^{-A^T} \xi) \mid \langle Ax, \xi \rangle = 0, t > 0\}$ away from radial directions. By uniqueness of \tilde{E}^+ up to regularizing, the statement of the lemma is equivalent to saying that in the latter relation one actually has $t \in (0, 1]$, and for that it is enough to show that $(x, \xi, t^A x, t^{-A^T} \xi) \notin \text{WF}'(E^+)$ for some $(t, x, \xi) \in \mathbb{R}_+ \times T^*\mathbb{R}^n \setminus 0$ satisfying $\langle Ax, \xi \rangle = 0$ and $t > 1$.

The kernel of E^+ equals

$$(4.2) \quad i \int_0^1 t^{i\alpha} \delta(y - t^A x) \frac{dt}{t},$$

i.e., it is the push-forward of $i t_+^{i\alpha - 1} H(1 - t) \otimes \delta(y - t^A x)$ under the projection $\mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n \times \mathbb{R}^n$, $(t, x, y) \mapsto (x, y)$. Now $\delta(y - t^A x)$ for fixed $t > 0$ has wave front

relation $\{(x, \xi, t^A x, t^{-A^T} \xi) \mid (x, \xi) \in \mathbb{R}^n \times (\mathbb{R}^n \setminus 0)\}$, and as integration in (4.2) is over $t \in (0, 1]$, the result follows. \square

4.3. Construction of E^- . There are two possibilities for obtaining E^- . One is by defining E^- as the adjoint to E^{*+} , where E^{*+} is constructed for the formally adjoint $P^* = \langle Ax, D_x \rangle + \bar{\alpha} - i \operatorname{tr} A$ to the operator P by using of the following obvious fact:

Lemma 4.4. *One has*

$$s_{p^*}^0(\zeta^0) = -s_p^0(\zeta^0).$$

Then, by duality, statements $E^{*+} : H^s(\zeta^0) \rightarrow H^s(\zeta^0)$ for $s > -s_p^0(\zeta^0)$ are translated to statements $E^- : H^s(\zeta^0) \rightarrow H^s(\zeta^0)$ for $s < s_p^0(\zeta^0)$, and vice versa.

The other possibility is by writing down the parametrix E^- explicitly. Now assuming $s^0 > 0$, i.e., $\Re(i\alpha) - \operatorname{tr} A/2 < 0$, this parametrix is of the form

$$E^- f(x) = -i \int_1^\infty t^{i\alpha} f(t^A x) \frac{dt}{t}, \quad f \in C_c^\infty(\mathbb{R}^n).$$

The relevant estimate becomes $\|f(t^A x)\|_s \leq C_s t^{-\operatorname{tr} A/2} \|f\|_s$ for all $t \geq 1$, where $s \leq 0$.

References

- [1] J. DUISTERMAAT & L. HÖRMANDER – “Fourier integral operators. II”, *Acta Math.* **128** (1972), p. 183–269.
- [2] Y. EGOROV – *Linear differential equations of principal type*, Nauka, Moscow, 1984, in Russian.
- [3] V. GUILLEMIN & D. SCHAEFFER – “On a certain class of Fuchsian partial differential operators”, *Duke Math. J.* **44** (1977), p. 157–199.
- [4] V. GUILLEMIN & G. UHLMANN – “Oscillatory integrals with singular symbols”, *Duke Math. J.* **48** (1981), p. 251–267.
- [5] L. HÖRMANDER – “Fourier integral operators. I”, *Acta Math.* **127** (1971), p. 79–183.
- [6] ———, *The analysis of linear partial differential equations IV*, Grundlehren Math. Wiss., vol. 275, Springer, Berlin, 1985.
- [7] M. JOSHI – “A symbolic construction of the forward fundamental solution of the wave operator”, *Comm. Partial Differential Equations* **23** (1998), p. 1349–1417.
- [8] R. MELROSE & G. UHLMANN – “Lagrangian intersection and the Cauchy problem”, *Comm. Pure Appl. Math.* **32** (1979), p. 483–519.
- [9] V. NAZAIKINSKII, V. SHATALOV & B. STERNIN – *Contact geometry and linear differential equations*, de Gruyter Exp. Math., vol. 6, de Gruyter, 1992.
- [10] M. RUZHANSKY & I. WITT – “Normal forms and regularity near radial points”, in preparation.
- [11] B. STERNIN – “Differential equations of subprincipal type”, *Math. USSR Sb.* **53** (1986), p. 37–67.
- [12] M. TAYLOR – *Pseudodifferential operators*, Princeton Math. Series, vol. 34, Princeton Univ. Press, Princeton, 1981.

ELLIPTIC GAMMA FUNCTION PROVIDES THE ČECH COCYCLE OF A GERBE

Chenchang Zhu

Mathematisches Institut, Georg-August-Universität Göttingen
E-mail: zhu@uni-math.gwdg.de

Abstract. G. Felder and A. Varchenko discovered certain modular formulas for elliptic gamma functions. These identities are generalized to an infinite set of identities for elliptic gamma functions associated to pairs of planes in 3-dimensional space in [2]. There we also use the language of stacks and gerbes to give a natural framework for a systematic description of these identities and their domain of validity. In this note I summarize the work in [2] with an emphasize on the Čech open covers.

1. Introduction

The elliptic gamma function [6] is a function of three complex variables obeying

$$\Gamma(z + \sigma, \tau, \sigma) = \theta_0(z, \tau) \Gamma(z, \tau, \sigma),$$
$$\theta_0(z, \tau) = \prod_{j=0}^{\infty} (1 - e^{2\pi i((j+1)\tau - z)})(1 - e^{2\pi i(j\tau + z)}).$$

In [3] three-term relations for Γ involving $ISL_3(\mathbb{Z}) = SL_2(\mathbb{Z}) \ltimes \mathbb{Z}^3$ were discovered, generalizing the modular properties of theta functions under $ISL_2(\mathbb{Z}) = SL_2(\mathbb{Z}) \ltimes \mathbb{Z}^2$.

In [2], we show that these identities are a special case of a set of three-term relations for a family of gamma functions $\Gamma_{a,b}$, which are interpreted geometrically as giving a meromorphic section of a hermitian gerbe on the *universal triptic curve*.

November 2008.

Supported by the Courant Center Higher Order Structures in Mathematics at the University Göttingen.

The result generalizes the fact that the theta function θ_0 is a section of a hermitian line bundle on the universal elliptic curve. We call this gerbe **gamma gerbe**.

2. Enlarged gamma function family

We describe gamma gerbe by the enlarged gamma function family. For a, b linearly independent in the set Λ_{prim} of primitive vectors (namely not multiples of other vectors) in the lattice $\Lambda = \mathbb{Z}^3$, there is a unique primitive $\gamma \in \Lambda_{prim}^\vee$ in the dual lattice such that $\det(a, b, \cdot) = s\gamma$ for some $s > 0$. For $w \in \mathbb{C}, x \in \Lambda \otimes \mathbb{C} = \mathbb{C}^3$ for which the products converge we define

$$\Gamma_{a,b}(w, x) := \frac{\prod_{\delta \in C_{+-}/\mathbb{Z}\gamma} (1 - e^{-2\pi i(\delta(x)-w)/\gamma(x)})}{\prod_{\delta \in C_{-+}/\mathbb{Z}\gamma} (1 - e^{2\pi i(\delta(x)-w)/\gamma(x)})},$$

where $C_{+-} = C_{+-}(a, b) = \{\delta \in \Lambda^\vee \mid \delta(a) > 0, \delta(b) \leq 0\}$ and $\mathbb{Z}\gamma$ acts on it by translation. We set similarly $C_{-+}(a, b) = C_{-+}(b, a)$. We define $\Gamma_{a,\pm a} = 1$. The function $\Gamma_{a,b}$ is meromorphic on $\mathbb{C} \times (U_a^+ \cap U_b^+)$, where U_a^+ 's are carefully chosen as

$$U_a^+ = \{x \in \mathbb{C}^3 \mid \text{Im}(\alpha(x)\overline{\beta(x)}) > 0\}$$

for any oriented basis α, β of the plane $\delta(a) = 0$. It is easy to check that U_a^+ is independent of the choice of α and β .

For linearly independent $a, b \in \Lambda_{prim}$, $\Gamma_{a,b}$ is a finite product of ordinary elliptic gamma functions:

$$(1) \quad \Gamma_{a,b}(w, x) = \prod_{\delta \in F/\mathbb{Z}\gamma} \Gamma\left(\frac{w + \delta(x)}{\gamma(x)}, \frac{\alpha(x)}{\gamma(x)}, \frac{\beta(x)}{\gamma(x)}\right),$$

for any $\alpha, \beta \in \Lambda^\vee$ satisfying $\alpha(b) = \beta(a) = 0$ and $\alpha(a) > 0, \beta(b) > 0$, $F = \{\delta \in \Lambda^\vee \mid 0 \leq \delta(a) < \alpha(a), 0 \leq \delta(b) < \beta(b)\}$.

3. The geometry of the universal triptic curve

Let X be the total space of the line bundle $O(1) \rightarrow (\mathbb{C}P^2 - \mathbb{R}P^2)$. Geometrically we think of $O(1)$ as the dual bundle to the tautological line bundle and of $\mathbb{C}P^2$ as the projectivization of $\Lambda_{\mathbb{C}} = \Lambda \otimes_{\mathbb{Z}} \mathbb{C}$, where $\Lambda = \mathbb{Z}^3$ is a free abelian group of rank 3 equipped with a volume form $\det : \wedge^3 \Lambda \rightarrow \mathbb{Z}$. The group $\text{Aut}(\Lambda) \cong SL_3(\mathbb{Z})$ of linear transformations of $\Lambda_{\mathbb{C}}$ mapping Λ to itself and preserving the volume form, acts naturally on X . The dual lattice $\Lambda^\vee = \text{Hom}(\Lambda, \mathbb{Z}) \cong \mathbb{Z}^3 \subset V_{\mathbb{C}}^*$ acts on $O(1)$ fiberwise by translation and we get an action of $\text{Aut}(\Lambda) \ltimes \Lambda^\vee \cong ISL_3(\mathbb{Z})$. More explicitly, this

group acts linearly on $\mathbb{C} \times V_{\mathbb{C}}$ via $(g, \mu)(w, x) = (w - \alpha(x), gx)$, and this action induces an action on $X = (\mathbb{C} \times (V_{\mathbb{C}} - \mathbb{C} \cdot V_{\mathbb{R}})) / \mathbb{C}^{\times}$. We will later learn that $\mathcal{X} = [X / ISL_3(\mathbb{Z})]$ is the universal triptic curve. In this section we first study the geometry of this stack.

The complex manifold X has a natural $ISL_3(\mathbb{Z})$ -equivariant open covering $\mathcal{V} = (V_a)_{a \in \Lambda_p}$ labeled by the set of primitive vectors Λ_p in Λ : for $a \in \Lambda_p$, $V_a = \{(w, x) \mid x \in U_a^+ \} / \mathbb{C}^{\times}$. Our extended elliptic gamma functions $\Gamma_{a,b}$ will give some Čech cocycle with respect to this $ISL_3(\mathbb{Z})$ -equivariant open covering $\mathcal{V} = (V_a)_{a \in \Lambda_p}$. We first prove some properties of V_a 's.

Proposition 3.1. *For any $a_1, \dots, a_p \in \Lambda_p$, $\cap_k V_{a_k}^+$ is either empty or contractible.*

Proof. Let $\tilde{U}_a^+ = U_a^+ / \mathbb{C}^{\times}$. The V_a are trivial line bundles over \tilde{U}_a^+ so it is sufficient to prove the claim for \tilde{U}_a^+ .

We divide the proof into two cases:

Case 1: Suppose there are three linearly independent elements in $\{a_k\}$. Since a $SL_3(\mathbb{Q})$ transformation of $\mathbb{R}^3 \supset \Lambda_p$ will not change the topology of $\cap_k \tilde{U}_{a_k}^+$, and any three linearly independent rational vectors can be transformed to the basis vectors under $SL_3(\mathbb{Q})$, we might as well assume that they are e_1, e_2, e_3 . Then $\cap_i \tilde{U}_{e_i}^+ = \{(x_1, x_2, x_3) \mid \text{Im}(x_1 \bar{x}_2) > 0, \text{Im}(x_2 \bar{x}_3) > 0, \text{Im}(x_3 \bar{x}_1) > 0\}$. We may assume that $x_3 = 1$, $x_1 = r_1 e^{-i\phi}$ and $x_2 = r_2 e^{i\psi}$ with $r_1, r_2 > 0$. Then the above set is exactly the set where

$$(2) \quad 0 < \phi < \pi, \quad 0 < \psi < \pi, \quad 0 < 2\pi - \phi - \psi < \pi,$$

and $r_1, r_2 \in \mathbb{R}^+$. This gives us the constraint on (ϕ, ψ) that it is inside a triangle Δ bounded by three lines defined by the above linear equations in the \mathbb{R}^2 and no constraint on r_1 and r_2 .

Any other vector a_k can be written as $a_k = s_k e_1 + t_k e_2 + u_k e_3$ with $s_k, t_k, u_k \in \mathbb{Q}$. Assume at first that $u_k \neq 0$. Then a basis in $H(a_k)$ can be chosen as $\alpha = u_k e_1 - s_k e_3$ and $\beta = u_k e_2 - t_k e_3$. Then $\tilde{U}_{a_k}^+$ gives us the constraint that

$$\text{Im}((u_k x_1 - s_k x_3)(u_k \bar{x}_2 - t_k \bar{x}_3)) > \text{or} < 0,$$

depending on the orientation of α and β . This inequality is equivalent to

$$(3) \quad \sin(\psi) s_k \rho_1 + \sin(\phi) t_k \rho_2 - \sin(\psi + \phi) u_k > \text{or} < 0,$$

where $\rho_i = r_i^{-1}$. Notice the symmetry of this inequality. In fact, we will arrive at the same inequality if we assume $s_k \neq 0$ or $t_k \neq 0$. Therefore for a fixed value of (ϕ_0, ψ_0) , the restriction of ρ_i 's is given by a series of linear equations on \mathbb{R}^2 . Hence ρ_i 's have to be in a certain polygon P in \mathbb{R}^2 .

With the condition (2), we observe that $\left(\frac{\sin \phi \sin(\phi_0 + \psi_0)}{\sin \phi_0 \sin(\phi + \psi)} \rho_1, \frac{\sin \psi \sin(\phi_0 + \psi_0)}{\sin \psi_0 \sin(\phi + \psi)} \rho_2\right)$ satisfies (3) at point (ϕ, ψ) as long as (ρ_1, ρ_2) satisfies (3) at the point (ϕ_0, ψ_0) .

Then it is easy to see that the following map $\cap_k \tilde{U}_{a_k}^+ \rightarrow \Delta \times P$ is a homeomorphism:

$$[(\rho_1^{-1} e^{-i\phi}, \rho_2^{-1} e^{i\psi}, 1)] \mapsto (\phi, \psi, \frac{\sin \phi \sin(\phi_0 + \psi_0)}{\sin \phi_0 \sin(\phi + \psi)} \rho_1, \frac{\sin \psi \sin(\phi_0 + \psi_0)}{\sin \psi_0 \sin(\phi + \psi)} \rho_2).$$

Case 2: Suppose that all a_k 's lie in the same plane. Since \tilde{U}_a^+ is homeomorphic to $\tilde{U}_{e_1}^+$ which is $\mathbb{C} \times H_+$ where H_+ is the upper half plane, the claim is trivial in the case that all a_k lie on the same line. By [2, Lemma 3.9] the intersection is empty if all of a_k 's do not lie on the same side of some plane. So after a $SL_3(\mathbb{Q})$ transformation, we can assume $a_1 = e_1$ and $a_2 = e_2$ and everything else lies in between, namely $a_k = s_k e_1 + t_k e_2$ with $s_k, t_k \in \mathbb{Q}^+$. $\tilde{U}_{e_1}^+ \cap \tilde{U}_{e_2}^+$ consists of points such that $\text{Im}(x_1 \bar{x}_3) < 0$ and $\text{Im}(x_2 \bar{x}_3) > 0$. After normalizing x_3 to 1, $\tilde{U}_{e_1}^+ \cap \tilde{U}_{e_2}^+$ is simply $H_+ \times H_-$. On the other hand, an oriented basis of the plane $H(a_k) := \{\alpha(a) = 0\} \subset \Lambda^\vee$ can be chosen as $s_k e_2 - t_k e_1$ and e_3 . So $\text{Im}(s_k x_2 - t_k x_1) \bar{x}_3 > 0$ always holds, namely $\tilde{U}_{a_k} \subset \tilde{U}_{e_1}^+ \cap \tilde{U}_{e_2}^+$. Therefore, in this case, $\cap \tilde{U}_{a_k} = \tilde{U}_{e_1}^+ \cap \tilde{U}_{e_2}^+ = H_+ \times H_-$ is contractible. \square

Proposition 3.2. *For any $a_1, \dots, a_p \in \Lambda_p, \cap_k V_{a_k}^+$ is either empty or Stein.*

Proof. Submanifolds of \mathbb{C}^n given by $|f_j(z)| < 1$, for f_1, \dots, f_k holomorphic are *domains of holomorphy* [5, Theorem 2.5.13]. Domains of holomorphy are Stein manifolds. The open sets V_a are isomorphic as complex manifolds to $H \times \mathbb{C} \times \mathbb{C}$, where H is the upper half plane. Hence V_a is a domain of holomorphy with $k = 1, f_1(z) = e^{iz_1}$. By [5, Cor. 2.5.7] finite intersections of domains of holomorphy are domains of holomorphy again. \square

Analytic coherent sheaves (such as \mathcal{O}) on Stein manifolds have vanishing cohomology in positive degree [5, Theorem 7.4.3]. If we have a contractible Stein manifold we reach the same conclusion for \mathcal{O}^\times with the long exact sequence. Hence our covering \mathcal{V} is a good covering in the sense that $H^{>0}(\cap_k V_{a_k}^+, \mathcal{O}^\times) = 0$. This implies that the Čech cohomology with respect to \mathcal{V}

$$\check{H}_\mathcal{V}^*(\mathcal{X}, \mathcal{O}^\times) = H^*(\mathcal{X}, \mathcal{O}^\times),$$

calculate the sheaf cohomology of \mathcal{X} .

4. Main theorems

The functions $\Gamma_{a,b}$ satisfy cocycle conditions generalizing the three-term relations of [3]:

$$(4) \quad \Gamma_{a,b}(w, x)\Gamma_{b,a}(w, x) = 1, \quad x \in U_a^+ \cap U_b^+,$$

$$\Gamma_{a,b}(w, x)\Gamma_{b,c}(w, x)\Gamma_{c,a}(w, x) = \exp\left(-\frac{\pi i}{3}P_{a,b,c}(w, x)\right), \quad x \in U_a^+ \cap U_b^+ \cap U_c^+,$$

where $P_{a,b,c}(w, x) \in \mathbb{Q}(x)[w]$ can be explicitly described in terms of the Bernoulli polynomial $B_{3,3}$, see [2]. Moreover the gamma functions obey cocycle identities related to the action of the group $ISL_3(\mathbb{Z}) = SL_3(\mathbb{Z}) \ltimes \mathbb{Z}^3$. Fix a framing of Λ_{prim} , namely for each $a \in \Lambda_{prim}$ a choice of oriented basis $(\alpha_1, \alpha_2, \alpha_3)$ of $\Lambda^\vee \otimes \mathbb{R}$ such that $\alpha_1(a) = 1, \alpha_2(a) = \alpha_3(a) = 0$. Let

$$\Delta_a((g, \mu); w, x) = \prod_{j=0}^{\mu(g^{-1}a)-1} \theta_0\left(\frac{w + j\alpha_1(x)}{\alpha_3(x)}, \frac{\alpha_2(x)}{\alpha_3(x)}\right),$$

where $(g, \mu) \in ISL_3(\mathbb{Z}) = SL_3(\mathbb{Z}) \ltimes \mathbb{Z}^3$. Then we have

$$(5) \quad \frac{\Gamma_{a,b}(w + \mu(g^{-1}x), x)}{\Gamma_{a,b}(w, x)} = e^{\pi i P_{a,b}((g,\mu);w,x)} \frac{\Delta_a((g, \mu); w, x)}{\Delta_b((g, \mu); w, x)},$$

$$(6) \quad \Delta_a(\hat{g}\hat{h}; w, x) = e^{2\pi i P_a(\hat{g}, \hat{h}; w, x)} \Delta_a(\hat{g}; w, x)\Delta_a(\hat{h}; w - \mu(x), g x),$$

where $\hat{g} = (g, \mu), \hat{h} = (h, \nu)$ and $P_{a,b}, P_a$ are again in $\mathbb{Q}(x)[w]$.

By equations (4) (5) (6) we have:

Theorem 4.1. *There is an $ISL_3(\mathbb{Z})$ -equivariant Čech 2-cocycle*

$$(\phi_{a,b,c}, \phi_{a,b}, \phi_a) = (e^{-\frac{2\pi i}{3!}P_{a,b,c}(w,x)}, e^{-\frac{2\pi i}{2!}P_{a,b}((g,\mu);w,x)}, e^{-2\pi i P_a((g,\mu),(h,\nu);w,x)}),$$

in $C^2_{ISL_3(\mathbb{Z})}(\mathcal{V}, \mathcal{O}^\times)$. The image of ϕ in the equivariant Čech complex with values in the sheaf \mathcal{M}^\times of invertible meromorphic sections is the coboundary of the equivariant cochain $(\Gamma_{a,b}, \Delta_a) \in C^1_{ISL_3(\mathbb{Z})}(\mathcal{V}, \mathcal{M}^\times)$.

The **gamma gerbe** \mathcal{G} is the holomorphic equivariant gerbe on X corresponding to ϕ . Equivalently, it is a holomorphic gerbe on the stack $\mathcal{X} = [X/ISL_3(\mathbb{Z})]$.

More geometrically, if we view gerbes over stacks as central extensions of groupoids, then \mathcal{G} is presented by a groupoid $R \rightrightarrows U_0$ fitting in the central extension of groupoids over U_0 :

$$1 \rightarrow \mathbb{C}^\times \times U_0 \rightarrow R \rightarrow U_1 \rightarrow 1,$$

where $U_0 = \sqcup V_a$, $U_1 = U_0 \times_X \times (ISL_3(\mathbb{Z}) \times X) \times_X U_0$, $R = \sqcup L_{a,b} \otimes L_b(g)^{-1}$ with $L_{a,b}$, $L_b(g)$ the \mathbb{C}^\times -bundles with transition functions $\phi_{a,b,c} \phi_{a,b,d}^{-1}$ (on $(V_a \cap V_b) \cap V_c \cap V_d$) and $\phi_{b,b'}(g, -) \phi_{b,b''}^{-1}(g, -)$ (on $V_b \cap V_{b'} \cap V_{b''}$) respectively. Notice that $U_1 = \cup W_{g,a,g^{-1}b}$ where $W_{g,a,g^{-1}b} = \{(g, y) | y \in V_a, g^{-1}y \in V_{g^{-1}b}\}$. Then $\Gamma_{a,b} \Delta_b^{-1}$ provides a meromorphic groupoid homomorphism $U_1 \rightarrow R$, hence Γ 's and Δ 's can be viewed as a meromorphic section of \mathcal{G} . A *hermitian structure of a gerbe* in this language is simply a hermitian structure of the complex line bundle associated to the central extension. There is a *unique* compatible connective on a hermitian gerbe and its curvature represents the Dixmier-Douady class of the gerbe.

Theorem 4.2. *Using the notation in (1), there is a hermitian structure $h_{a,b} h_b^{-1}$ on \mathcal{G} with*

$$h_{a,b}(w, x) = \prod_{\delta \in F/\mathbb{Z}\gamma} h_3 \left(\frac{w + \delta(x)}{\gamma(x)}, \frac{\alpha(x)}{\gamma(x)}, \frac{\beta(x)}{\gamma(x)} \right)$$

and

$$h_a((g, \mu); w, x) = \prod_{j=0}^{\mu(g^{-1}a)-1} h_2 \left(\frac{w + j\alpha_1(x)}{\alpha_3(x)}, \frac{\alpha_2(x)}{\alpha_3(x)} \right),$$

where h_n are defined by Bernoulli polynomials:

$$h_n(z, \tau_1, \dots, \tau_{n-1}) = \exp \left(-(4\pi/n!) B_{n-1,n}(\zeta, t_1, \dots, t_{n-1}) \right),$$

with $\zeta = \text{Im } z$, $t_j = \text{Im } \tau_j$.

Moreover, as with line bundles, we can construct the gamma gerbe \mathcal{G} via (pseudo)-divisors. A *triptic curve* \mathcal{E} is a holomorphic stack of the form $[\mathbb{C}/\iota(\mathbb{Z}^3)]$ with $\iota: \mathbb{Z}^3 \rightarrow \mathbb{C}$ a map of rank 2 over \mathbb{R} . An *orientation* of a triptic curve \mathcal{E} is given by a choice of a generator of $H^3(\mathcal{E}, \mathbb{Z}) \cong \mathbb{Z}$. Then the stack $\mathcal{T}r := [(\mathbb{C}P^2 - \mathbb{R}P^2)/SL_3(\mathbb{Z})]$ is the moduli space of oriented triptic curves. The stack $\mathcal{X} = [X/ISL_3(\mathbb{Z})]$ is the total space of the universal family of triptic curves over $\mathcal{T}r$. Given an étale map $U \rightarrow \mathcal{E}$, let $Z_U = 0 \times_{\mathcal{E}} U$. Z_U is naturally a discrete subset of a principal oriented \mathbb{R} -bundle on U . Then a *pseudodivisor* on U is a function $D: Z_U \rightarrow \mathbb{Z}$ such that if $\lim y_n = +\infty$ (resp. $-\infty$) for a sequence y_n in Z_U with compact support in U then $\lim D(y_n) = 1$ (resp. -1). The notion of positive/negative infinity is derived from the orientation of the fibres of the \mathbb{R} -bundle. We can globalize this to \mathcal{X} , namely for an étale map $U \rightarrow \mathcal{X}$, a *pseudodivisor* on U is a function $D: \mathcal{T}r \times_{\mathcal{X}} U \rightarrow \mathbb{Z}$ such that for every point $q \rightarrow \mathcal{T}r$ with corresponding fibre $\mathcal{E} = q \times_{\mathcal{T}r} \mathcal{X}$, the restriction $q \times_{\mathcal{X}} U$ is a pseudodivisor on $U \times_{\mathcal{X}} \mathcal{E}$. Then for two such D_i 's, the pushforward $p_*(D_1 - D_2)$ is a

divisor on U , hence can be used to twist a line bundle L to $L(p_*(D_1 - D_2))$, where $p: \mathcal{T}r \times_{\mathcal{X}} U \rightarrow U$. Using the categorical description of gerbes in [1], we then have

Theorem 4.3. *The gamma gerbe \mathcal{G} is a gerbe over \mathcal{X} made up by the following data: for U with an étale open map $U \rightarrow \mathcal{X}$,*

$$\begin{aligned} \text{Obj}(\mathcal{G}_U) &= \{(L, D) \mid L \text{ is a line bundle on } U \text{ and } D \text{ a pseudodivisor}\}, \\ \text{Mor}(\mathcal{G}_U)((L_1, D_1) \rightarrow (L_2, D_2)) &= \Gamma^\times(U, (L_1^* \otimes L_2)(p_*(D_2 - D_1))), \end{aligned}$$

the invertible holomorphic sections.

We also have the following theorems calculating various cohomology groups and Dixmier-Douady classes of the gamma gerbes or its restriction.

Theorem 4.4. *Let $\mathcal{E}_r = \mathbb{C}/\iota(\mathbb{Z}^r)$, where $x_j = \iota(e_j)$, the images of the standard basis vectors, are assumed to be \mathbb{Q} -linearly independent and to span \mathbb{C} over \mathbb{R} . Then*

$$H^{i \leq r-2}(\mathcal{E}_r, \mathcal{O}^\times) = \wedge^i(\mathbb{C}^r / (x_1, \dots, x_r)\mathbb{C}) / \wedge^i(\mathbb{Z}^r), \quad H^{r-1}(\mathcal{E}_r, \mathcal{O}^\times) = \mathcal{E}_r \times \mathbb{Z},$$

and $H^{\geq r}(\mathcal{E}_r, \mathcal{O}^\times) = 0$. In particular, for generic triptic curves \mathcal{E} , the groups classifying holomorphic and topological gerbes on \mathcal{E} are

$$H^2(\mathcal{E}, \mathcal{O}^\times) = \mathcal{E} \times \mathbb{Z} \quad \text{and} \quad H^3(\mathcal{E}, \mathbb{Z}) = \mathbb{Z},$$

respectively.

Theorem 4.5. *The Dixmier-Douady class $c(\mathcal{G}|_{\mathcal{E}})$ of the restriction of the gamma gerbe to \mathcal{E} is a generator of $H^3(\mathcal{E}, \mathbb{Z}) = \mathbb{Z}$.*

Theorem 4.6. *$H^3(\mathcal{X}, \mathbb{Z})$ fits in a short exact sequence*

$$0 \rightarrow \mathbb{Z} \rightarrow H^3(\mathcal{X}, \mathbb{Z}) / \text{torsion} \rightarrow H^3(\mathbb{Z}^3, \mathbb{Z}) \cong \mathbb{Z} \rightarrow 0.$$

The image of the Dixmier-Douady class $c(\mathcal{G}) \in H^3(\mathcal{X}, \mathbb{Z})$ of the gamma gerbe is a generator of $H^3(\mathbb{Z}^3, \mathbb{Z})$.

There should exist non-abelian versions of this story in the context of q-deformed conformal field theory [4].

References

- [1] J.-L. BRYLINSKI – *Loop spaces, characteristic classes and geometric quantization*, Progress in Mathematics, vol. 107, Birkhäuser Boston Inc., Boston, MA, 1993.
- [2] G. FELDER, A. HENRIQUES, C. A. ROSSI & C. ZHU – “A gerbe for the elliptic gamma function”, *Duke Math. J.* **141** (2008), no. 1, p. 1–74.
- [3] G. FELDER & A. VARCHENKO – “The elliptic gamma function and $SL(3, \mathbf{Z}) \times \mathbf{Z}^3$ ”, *Adv. Math.* **156** (2000), no. 1, p. 44–76.
- [4] ———, “ q -deformed KZB heat equation: completeness, modular properties and $SL(3, \mathbf{Z})$ ”, *Adv. Math.* **171** (2002), no. 2, p. 228–275.
- [5] L. HÖRMANDER – *An introduction to complex analysis in several variables*, third ed., North-Holland Mathematical Library, vol. 7, North-Holland Publishing Co., Amsterdam, 1990.
- [6] S. N. M. RUIJSENAARS – “First order analytic difference equations and integrable quantum systems”, *J. Math. Phys.* **38** (1997), no. 2, p. 1069–1146.

THE AFFINE MACDONALD'S FORMULA

David Kazhdan

Einstein Institute of Mathematics, The Hebrew University of Jerusalem
E-mail: kazhdan@math.huji.ac.il

Abstract. In a paper with A. Braverman we defined the affine spherical Hecke algebra. It is natural then to ask about the spherical eigenfunctionals. In the finite-dimensional case they are described by MacDonal. In this note we will talk about their affine analogue.

1. The affine Macdonald's formula

Topic: Spherical functions on Hecke algebras.

Let G be a group, H a subgroup, $\rho : G \longrightarrow \text{Aut}(V)$ an irreducible representation of G such that $\dim V^H = \dim((V^\vee)^H) = 1$. Fix $v \in V^H$, $v^\vee \in (V^\vee)^H$, $\langle v^\vee, v \rangle = 1$. Consider the matrix coefficient $\varphi_\rho : G \longrightarrow \mathbb{C}$, $\varphi_\rho(g) = \langle v^\vee, gv \rangle$. Then φ is $H \times H$ -invariant on G .

Example 1. $G = H \times H$, $H \xrightarrow{\Delta} H \times H$. Any irreducible representation ρ of G has the form $\rho = \pi_1 \otimes \pi_2$ where π_1, π_2 are irreducible representations of H ; $\dim V^H = 1$, $\dim((V^\vee)^H) = 1$ iff $\pi_2 = \pi_1^\vee$. Here $\rho_{\pi \otimes \pi^\vee}(h_1, h_2) := \text{Tr}_\pi(h_1 h_2^{-1})$.

Example 2. Let G be a semisimple reductive group, $H \subset G$ the maximal compact subgroup. Then for any irreducible representation $\rho : G \longrightarrow \text{Aut}(V)$ we have $\dim V^H \leq 1$. The representation ρ is of class I iff $\dim V^H = 1$. The $f = \varphi_\rho$ is called the spherical function of ρ . If G is a complex group then φ_ρ is given as a regular function on G given by an implicit formula similar to Weyl's character formula.

If G is an arbitrary split reductive group then the formula is more complicated. Consider the case where G is a group over a p -adic field.

Why should we have $\dim V^H \leq 1$? Back to the general case. Assume $H = K$ be compact, G local compact. We define $\mathcal{H} = \mathcal{H}(G, K)$ to be the space of finite measures, $K \times K$ invariant on G with compact support. Then $m : G \times G \longrightarrow G$ induces an algebra structure on \mathcal{H} , $\mu_1 \mu_2 := m_*(\mu_1 \square \mu_2)$ (\square – direct product of measures). If ρ is a representation then for any $\mu \in \mathcal{H}$ we define $\rho(\mu) \in \text{Aut}(V)$ by

$$\rho(\mu) := \int \rho(g) d\mu(g)|_{V^K}.$$

Lemma 3.

- a) If ρ is irreducible, then the representation $\mu \longrightarrow \rho(\mu)$ of \mathcal{H} on V_ρ^K is irreducible.
- b) The algebra \mathcal{H} is commutative iff for any irreducible $\rho : G \longrightarrow \text{Aut}(V)$ we have $\dim(V^K) \leq 1$.

How do we see that \mathcal{H} is commutative?

$G = \text{GL}_n(\mathbb{R})$, $K = \text{SO}_n(\mathbb{R})$. Consider the map $\tau : \mathcal{H} \longrightarrow \mathcal{H}$ induced by the anti-involution $g \longmapsto {}^t g$; τ is an anti-involution of \mathcal{H} . On the other hand since ${}^t g \in KgK$ it follows that $\text{as } \tau = \text{id}$.

Assume now G is a split \mathbb{Q}_p -group, e.g. $\text{GL}_n(\mathbb{Q}_p)$, $\text{Sp}_{2n}(\mathbb{Q}_p)$, $\text{E}_8(\mathbb{Q}_p)$, $K = G(\mathbb{Z}_p) \subset G$. The same argument as before shows that \mathcal{H} is commutative.

Theorem 4 (Satake). We have a canonical map

$$\mathcal{H}(G, K) \cong \mathbb{C}({}^L G) {}^L G.$$

We give a construction for the case $G = \text{GL}_n(\mathbb{Q}_p)$. The Chevalley theorem says that the restriction

$$\mathbb{C}(\text{GL}_n(\mathbb{C})) \longrightarrow \mathbb{C}(T_{\mathbb{C}}), \quad T = \begin{pmatrix} * & & & 0 \\ & \ddots & & \\ & & \ddots & \\ 0 & & & * \end{pmatrix}$$

induces an isomorphism invariant under conjugacy

$$\mathbb{C}(\text{GL}_n(\mathbb{C}))^{\text{GL}_n(\mathbb{C})} \cong \mathbb{C}(T_{\mathbb{C}})^{\Sigma_n}$$

So to construct \mathcal{H} we have to construct a homomorphism $\kappa_t : \mathcal{H} \longrightarrow \mathbb{C}$ for each $t \in T$ so that $\kappa_t = \kappa_{tw}$ for all $w \in \Sigma_n$. Let

$$B = \begin{pmatrix} b_{11} & & * \\ & \ddots & \\ 0 & & b_{nn} \end{pmatrix} \subset G.$$

To $t \in T$ consider $\lambda_t : B \longrightarrow \mathbb{C}^\times$, $\lambda_t(b) = t_1^{v(b_{11})} \cdots t_n^{v(b_{nn})}$ and since $G = KB$ we have $\dim V_t^K = 1$. We obtain a homomorphism $\kappa_t : \mathcal{H} \longrightarrow \text{End}(V_t^K)$. One can check that $\kappa_t = \kappa_{tw}$ for $w \in \Sigma_n$ and $\kappa : \mathcal{H} \longrightarrow \mathbb{C}({}^L G)^{{}^L G}$.

The space $\mathbb{C}({}^L G)^{{}^L G}$ has a canonical basis consisting of characters

$$\chi_\mu, \quad \mu \in X_+({}^L G) = X_+(G)$$

consisting of characters of irreducible representations of the group ${}^L G$. On the other hand the space $\mathcal{H}(G, K)$ has a basis of characters of functions

$$\delta_\lambda, \quad \lambda \in K \backslash G / K = (T(F) / T(\mathcal{O}))_W = X_+(G).$$

So we have a decomposition $\kappa^{-1}(\chi_\mu) = \sum_\lambda c_\mu^\lambda \delta_\lambda$. What do the c_μ^λ look like?

- $c_\mu^\lambda = c_\mu^\lambda(p)$ are polynomials in p .
- $c_\mu^\lambda(1) =$ multiplicity of the weight $\lambda \in X^*(T)$ in the representation on V_μ of ${}^L G$.

Let $\tau : \text{SL}_2 \longrightarrow {}^L G$ be the principal SL_2 . Then the restriction of $T = \tau \left(\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \right)$ defines a representation on V_μ ($\tau \left(\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \right)$ is regular unipotent). On the graded space

$$\bar{V}_\mu = \bigoplus_{i \geq 0} \bar{V}_\mu(i)$$

T^\vee preserves the filtration and $c_\mu^\lambda(p) = \sum_{i \geq 0} p^i \cdot$ (multiplicity of λ in $\bar{V}_\mu(i)$).

We are interested in the extension to Kac-Moody algebras. If G is a reductive group over a field F we can determine an affine Kac-Moody algebra \tilde{G} such that

$$0 \longrightarrow F^\times \longrightarrow \tilde{G} \longrightarrow G' \longrightarrow 0$$

$$G' = F^\times \rtimes G[t, t^{-1}].$$

If $F = \mathbb{Q}_p$ Braverman and Kazhdan define the corresponding Hecke algebra $\mathcal{H}(\tilde{G}(\mathbb{Q}_p), \tilde{G}(\mathbb{Z}_p))$ and construct the Satake isomorphism

$$\mathcal{H} \longrightarrow \{\text{Integral representations of } \tilde{G}^\vee(\mathbb{C})\}.$$

HERMANN WEYL AND THE EARLY HISTORY OF GAUGE THEORIES

Norbert Straumann

Institute for Theoretical Physics, University of Zürich
E-mail: norbert@physik.unizh.ch

Abstract. One of the major developments of twentieth century physics has been the gradual recognition that a common feature of the known fundamental interactions is their gauge structure. In this talk the early history of gauge theory is reviewed, emphasizing especially Weyl's seminal contributions of 1918 and 1929.

1. Introduction

The history of gauge theories begins with General Relativity, which can be regarded as a non-Abelian gauge theory of a special type. To a large extent the other gauge theories emerged in a slow and complicated process gradually from General Relativity. Their common geometrical structure – best expressed in terms of connections on fiber bundles – is now widely recognized.

It all began with H. Weyl [1], who made in 1918 the first attempt to extend General Relativity in order to describe gravitation and electromagnetism within a unifying geometrical framework. This brilliant proposal contains the germs of all *mathematical* aspects of non-Abelian gauge theory. The word “gauge” (german: “Eich-”) transformation appeared for the first time in this paper, but in the everyday meaning of change of length or change of calibration.

Einstein admired Weyl's theory as “a coup of genius of the first rate,” but immediately realized that it was physically untenable. After a long discussion Weyl finally

admitted that his attempt was a failure as a physical theory. (For a discussion of the intense Einstein–Weyl correspondence, see [2].) It paved, however, the way for the correct understanding of gauge invariance. Weyl himself reinterpreted in 1929 his original theory after the advent of quantum theory in a grand paper [3]. Weyl's reinterpretation of his earlier speculative proposal had actually been suggested before by London [12], Fock [16], Klein [35], and others arrived at the principle of gauge invariance in the framework of wave mechanics along completely different lines. It was, however, Weyl who emphasized the role of gauge invariance as a *constructive principle* from which electromagnetism can be derived. This point of view became very fruitful for our present understanding of fundamental interactions. (For a more extensive discussion, see [4].)

2. Weyl's Attempt to Unify Gravitation and Electromagnetism

On the 1st of March 1918 Weyl writes in a letter to Einstein ([5], Vol. 8B, Doc. 472): “These days I succeeded, as I believe, to derive electricity and gravitation from a common source ...” Einstein's prompt reaction by postcard indicates already a physical objection which he explained in detail shortly afterwards. Before we come to this we have to describe Weyl's theory of 1918.

2.1. Weyl's Generalization of Riemannian Geometry. Weyl's starting point was purely mathematical. He felt a certain uneasiness about Riemannian geometry, as is clearly expressed by the following sentences early in his paper:

But in Riemannian geometry described above there is contained a last element of geometry “at a distance” (ferngeometrisches Element) – with no good reason, as far as I can see; it is due only to the accidental development of Riemannian geometry from Euclidean geometry. The metric allows the two magnitudes of two vectors to be compared, not only at the same point, but at any arbitrarily separated points. *A true infinitesimal geometry should, however, recognize only a principle for transferring the magnitude of a vector to an infinitesimally close point* and then, on transfer to an arbitrary distant point, the integrability of the magnitude of a vector is no more to be expected than the integrability of its direction.

After these remarks Weyl turns to physical speculation and continues as follows:

On the removal of this inconsistency there appears a geometry that, surprisingly, when applied to the world, *explains not only the gravitational phenomena but also the electrical*. According to the resultant theory both spring from the same source, indeed *in general one cannot separate gravitation and electromagnetism in a unique manner*. In this theory all physical quantities have a world geometrical meaning; the action appears from the beginning as a pure number. It leads to an essentially unique universal law; it even allows us to understand in a certain sense why the world is four-dimensional.

In brief, Weyl's geometry can be described as follows. First, the spacetime manifold M is equipped with a *conformal structure*, that is, with a class $[g]$ of conformally equivalent Lorentz metrics g (and not a definite metric as in General Relativity). This corresponds to the requirement that it should only be possible to compare lengths at one and the same world point. Second, it is assumed, as in Riemannian geometry, that there is an affine (linear) torsion-free connection which defines a covariant derivative ∇ , and respects the conformal structure. Differentially this means that for any $g \in [g]$ the covariant derivative ∇g should be proportional to g :

$$(1) \quad \nabla g = -2A \otimes g \quad (\nabla_\lambda g_{\mu\nu} = -2A_\lambda g_{\mu\nu}),$$

where $A = A_\mu dx^\mu$ is a differential 1-form.

Alternatively, the second condition can be formulated in bundle theoretical language as follows. For a conformal manifold $(M, [g])$ we can introduce the bundle of *conformal frames*, which are linear frames consisting of pairwise orthogonal vectors of equal length (relative to $g \in [g]$). The set $W(M)$ of conformal frames on M can be regarded in an obvious manner as the total space of a principle fibre bundle, whose structure group G is the group consisting of all positive multiples of homogeneous Lorentz transformations, that is, $G \cong O(1, 3) \times \mathbb{R}_+$. This *conformal (Weyl) bundle* is a reduction of the bundle of linear frames $L(M)$. A *Weyl connection* is a torsion-free connection on $W(M)$. (As such it has a unique extension to $L(M)$.) It is easy to see that the corresponding covariant derivative satisfies (1). (For details, see [9].)

Consider now a curve $\gamma: [0, 1] \rightarrow M$ and a parallel-transported vector field X along γ . If l is the length of X , measured with a representative $g \in [g]$, we obtain from (1) the following relation between $l(p)$ for the initial point $p = \gamma(0)$ and $l(q)$ for

the end point $q = \gamma(1)$:

$$(2) \quad l(q) = \exp\left(-\int_{\gamma} A\right) l(p).$$

Thus, the ratio of lengths in q and p (measured with $g \in [g]$) depends in general on the connecting path γ (see Figure 1). The length is only independent of γ if the

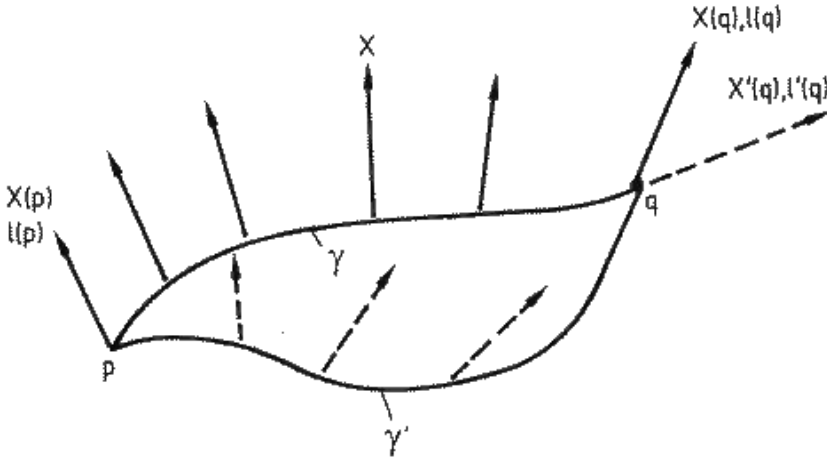


FIGURE 1. Path dependence of parallel transport of length

exterior differential of A ,

$$(3) \quad F = dA \quad (F_{\mu\nu} = \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu}),$$

vanishes.

The compatibility requirement (1) leads to the following expression for the Christoffel symbols in Weyl's geometry:

$$(4) \quad \Gamma^{\mu}_{\nu\lambda} = \frac{1}{2}g^{\mu\sigma}(g_{\lambda\sigma,\nu} + g_{\sigma\nu,\lambda} - g_{\nu\lambda,\sigma}) + g^{\mu\sigma}(g_{\lambda\sigma}A_{\nu} + g_{\sigma\nu}A_{\lambda} - g_{\nu\lambda}A_{\sigma}).$$

The second A -dependent term is a characteristic new piece in Weyl's geometry which has to be added to the Christoffel symbols of Riemannian geometry.

Until now we have chosen a fixed, but arbitrary metric in the conformal class $[g]$. This corresponds to a choice of calibration (or gauge). Passing to another calibration with metric \bar{g} , related to g by

$$(5) \quad \bar{g} = e^{2\lambda}g,$$

the potential A in (1) will also change to \bar{A} , say. Since the covariant derivative has an absolute meaning, \bar{A} can easily be worked out: On the one hand we have by definition

$$(6) \quad \nabla \bar{g} = -2\bar{A} \otimes \bar{g},$$

and on the other hand we find for the left side with (1)

$$(7) \quad \nabla \bar{g} = \nabla(e^{2\lambda} g) = 2d\lambda \otimes \bar{g} + e^{2\lambda} \nabla g = 2d\lambda \otimes \bar{g} - 2A \otimes \bar{g}.$$

Thus

$$(8) \quad \bar{A} = A - d\lambda \quad (\bar{A}_\mu = A_\mu - \partial_\mu \lambda).$$

This shows that a change of calibration of the metric induces a “*gauge transformation*” for A :

$$(9) \quad g \rightarrow e^{2\lambda} g, \quad A \rightarrow A - d\lambda.$$

Only gauge classes have an absolute meaning. (The Weyl connection is, however, gauge invariant. This is conceptually clear, but can also be verified by direct calculation from (4).) Note that the integral in 3 is gauge invariant.

2.2. Electromagnetism and Gravitation. Turning to physics, Weyl assumes that his “purely infinitesimal geometry” describes the structure of spacetime and consequently he requires that physical laws should satisfy a double-invariance: 1. They must be invariant with respect to arbitrary smooth coordinate transformations. 2. They must be *gauge invariant*, that is, invariant with respect to substitutions (7) for an arbitrary smooth function λ .

Nothing is more natural to Weyl than identifying A_μ with the vector potential and $F_{\mu\nu}$ in (3) with the field strength of electromagnetism. In the absence of electromagnetic fields ($F_{\mu\nu} = 0$) the scale factor $\exp(-\int_\gamma A)$ in (2) for length transport becomes path independent (integrable) and one can find a gauge such that A_μ vanishes for simply connected spacetime regions. In this special case one is in the same situation as in General Relativity.

Weyl proceeds to find an action which is generally invariant as well as gauge invariant and which would give the coupled field equations for g and A . We do not want to enter into this, except for the following remark. In his first paper [1] Weyl proposes what we call nowadays the Yang–Mills action

$$(10) \quad S(g, A) = -\frac{1}{4} \int \text{Tr}(\Omega \wedge * \Omega).$$

Here Ω denotes the curvature form and $*\Omega$ its Hodge dual.⁽¹⁾ Note that the latter is gauge invariant, that is, independent of the choice of $g \in [g]$. In Weyl's geometry the curvature form splits as $\Omega = \hat{\Omega} + F$, where $\hat{\Omega}$ is the metric piece [9]. Correspondingly, the action also splits,

$$(11) \quad \text{Tr}(\Omega \wedge *\Omega) = \text{Tr}(\hat{\Omega} \wedge *\hat{\Omega}) + \text{Tr}(F \wedge *F).$$

The second term is just the Maxwell action. Weyl's theory thus contains formally all aspects of a non-Abelian gauge theory.

Weyl emphasizes, of course, that the Einstein–Hilbert action is not gauge invariant. Later work by Pauli [10] and by Weyl himself [7, 1] led soon to the conclusion that the action (10) could not be the correct one, and other possibilities were investigated (see the later editions of Weyl's classic treatise [7]).

Independent of the precise form of the action Weyl shows that in his theory gauge invariance implies the *conservation of electric charge* in much the same way as general coordinate invariance leads to the conservation of energy and momentum.⁽²⁾ This beautiful connection pleased him particularly: "... [it] seems to me to be the strongest general argument in favour of the present theory – insofar as it is permissible to talk of justification in the context of pure speculation." The invariance principles imply five "Bianchi type" identities. Correspondingly, the five conservation laws follow in two independent ways from the coupled field equations and may be "termed the eliminants" of the latter. These structural connections hold also in modern gauge theories.

2.3. Einstein's Objection and Reactions of Other Physicists. After this sketch of Weyl's theory we come to Einstein's striking counterargument which he first communicated to Weyl by postcard. The problem is that if the idea of a nonintegrable length connection (scale factor) is correct, then the behavior of clocks would depend on their history. Consider two identical atomic clocks in adjacent world points and bring them along different world trajectories which meet again in adjacent world points. According to (2) their frequencies would then generally differ. This is in clear contradiction with empirical evidence, in particular with the existence of stable atomic spectra. Einstein therefore concludes (see [5], Vol. 8B, Doc. 507):

⁽¹⁾In local coordinates, the integrand in (10) is $R_{\alpha\beta\gamma\delta}R^{\alpha\beta\gamma\delta}\sqrt{-g}dx^0 \wedge \dots \wedge dx^3$, where $R_{\alpha\beta\gamma\delta}$ is the curvature tensor of the Weyl connection; this is the expression used by Weyl.

⁽²⁾We adopt here the somewhat naive interpretation of energy-momentum conservation for generally invariant theories of the older literature.

... (if) one drops the connection of the ds to the measurement of distance and time, then relativity loses all its empirical basis.

Nernst shared Einstein's objection and demanded on behalf of the Berlin Academy that it should be printed in a short amendment to Weyl's article. Weyl had to accept this. One of us has described the intense and instructive subsequent correspondence between Weyl and Einstein elsewhere [2] (see also Vol. 8B of [5]). As an example, let us quote from one of the last letters of Weyl to Einstein ([5], Vol. 8B, Doc. 669):

This [insistence] irritates me of course, because experience has proven that one can rely on your intuition; so unconvincing as your counterarguments seem to me, as I have to admit ...

By the way, you should not believe that I was driven to introduce the linear differential form in addition to the quadratic one by physical reasons. I wanted, just to the contrary, to get rid of this "methodological inconsistency (*Inkonsequenz*)" which has been a bone of contention to me already much earlier. And then, to my surprise, I realized that it looked as if it might explain electricity. You clap your hands above your head and shout: But physics is not made this way! (Weyl to Einstein 10. 12. 1918).

Weyl's reply to Einstein's criticism was, generally speaking, this: The real behavior of measuring rods and clocks (atoms and atomic systems) in arbitrary electromagnetic and gravitational fields can be deduced only from a dynamical theory of matter.

Not all leading physicists reacted negatively. Einstein transmitted a very positive first reaction by Planck, and Sommerfeld wrote enthusiastically to Weyl that there was "... hardly doubt that you are on the correct path and not on the wrong one."

In his encyclopedia article on relativity [11] Pauli gave a lucid and precise presentation of Weyl's theory, but commented on Weyl's point of view very critically. At the end he states:

... In summary one may say that Weyl's theory has not yet contributed to getting closer to the solution of the problem of matter.

Also Eddington's reaction was at first very positive but he soon changed his mind and denied the physical relevance of Weyl's geometry.

The situation was later appropriately summarized by F. London in his 1927 paper [12] as follows:

In the face of such elementary experimental evidence, it must have been an unusually strong metaphysical conviction that prevented Weyl from abandoning the idea that Nature would have to make use of the beautiful geometrical possibility that was offered. He stuck to his conviction and evaded discussion of the above-mentioned contradictions through a rather unclear re-interpretation of the concept of “real state,” which, however, robbed his theory of its immediate physical meaning and attraction.

In this remarkable paper, London suggested a reinterpretation of Weyl’s principle of gauge invariance within the new quantum mechanics: The role of the metric is taken over by the wave function, and the rescaling of the metric has to be replaced by a phase change of the wave function.

In this context an astonishing early paper by Schrödinger [13] has to be mentioned, which also used Weyl’s “World Geometry” and is related to Schrödinger’s later invention of wave mechanics. This relation was discovered by V. Raman and P. Forman [14]. (See also the discussion by C. N. Yang in [15].)

Even earlier than London, V. Fock [16] arrived along a completely different line at the principle of gauge invariance in the framework of wave mechanics. His approach was similar to the one by O. Klein [35].

The contributions by Schrödinger [13], London [12] and Fock [16] are commented in [8], where also English translations of the original papers can be found. Here, we concentrate on Weyl’s seminal paper “Electron and Gravitation.”

3. Weyl’s 1929 Classic: “Electron and Gravitation”

Shortly before his death late in 1955, Weyl wrote for his *Selecta* [17] a postscript to his early attempt in 1918 to construct a “unified field theory.” There he expressed his deep attachment to the gauge idea and adds (p. 192):

Later the quantum-theory introduced the Schrödinger–Dirac potential ψ of the electron–positron field; it carried with it an experimentally based principle of gauge invariance which guaranteed the conservation of charge, and connected the ψ with the electromagnetic potentials A_μ in the same way that my speculative theory had connected the gravitational potentials $g_{\mu\nu}$ with the A_μ , and measured the A_μ in known atomic, rather than unknown cosmological units. I have no doubt but that the correct context

for the principle of gauge invariance is here and not, as I believed in 1918, in the intertwining of electromagnetism and gravity.

This re-interpretation was developed by Weyl in one of the great papers of this century [3]. Weyl's classic does not only give a very clear formulation of the gauge principle, but contains, in addition, several other important concepts and results – in particular his two-component spinor theory.

The modern version of the gauge principle is already spelled out in the introduction:

The Dirac field-equations for ψ together with the Maxwell equations for the four potentials f_p of the electromagnetic field have an invariance property which is formally similar to the one which I called gauge invariance in my 1918 theory of gravitation and electromagnetism; the equations remain invariant when one makes the simultaneous substitutions

$$\psi \text{ by } e^{i\lambda}\psi \quad \text{and} \quad f_p \text{ by } f_p - \frac{\partial\lambda}{\partial x^p},$$

where λ is understood to be an arbitrary function of position in four-space. Here the factor $\frac{e}{c\hbar}$, where $-e$ is the charge of the electron, c is the speed of light, and $\frac{\hbar}{2\pi}$ is the quantum of action, has been absorbed in f_p . The connection of this "gauge invariance" to the conservation of electric charge remains untouched. But a fundamental difference, which is important to obtain agreement with observation, is that the exponent of the factor multiplying ψ is not real but pure imaginary. ψ now plays the role that Einstein's ds played before. It seems to me that this new principle of gauge invariance, which follows not from speculation but from experiment, tells us that the electromagnetic field is a necessary accompanying phenomenon, not of gravitation, but of the material wave field represented by ψ . Since gauge invariance involves an arbitrary function λ it has the character of "general" relativity and can naturally only be understood in that context.

We shall soon enter into Weyl's justification which is, not surprisingly, strongly associated with General Relativity. Before this we have to describe his incorporation of the Dirac theory into General Relativity which he achieved with the help of the tetrad formalism.

One of the reasons for adapting the Dirac theory of the spinning electron to gravitation had to do with Einstein's recent unified theory which invoked a distant parallelism with torsion. E. Wigner [18] and others had noticed a connection between this theory and the spin theory of the electron. Weyl did not like this and wanted to dispense with teleparallelism. In the introduction he says:

I prefer not to believe in distant parallelism for a number of reasons. First my mathematical intuition objects to accepting such an artificial geometry; I find it difficult to understand the force that would keep the local tetrads at different points and in rotated positions in a rigid relationship. There are, I believe, two important physical reasons as well. The loosening of the rigid relationship between the tetrads at different points converts the gauge factor $e^{i\lambda}$, which remains arbitrary with respect to ψ , from a constant to an arbitrary function of space-time. In other words, only through loosening the rigidity does the established gauge invariance become understandable.

This thought is carried out in detail after Weyl has set up his two-component theory in special relativity, including a discussion of P and T invariance. He emphasizes thereby that the two-component theory excludes a linear implementation of parity and remarks: "It is only the fact that the left-right symmetry actually appears in Nature that forces us to introduce a second pair of ψ -components." To Weyl the mass problem is thus not relevant for this.⁽³⁾ Indeed he says: "Mass, however, is a gravitational effect; thus there is hope of finding a substitute in the theory of gravitation that would produce the required corrections."

3.1. Tetrad Formalism. In order to incorporate his two-component spinors into General Relativity, Weyl was forced to make use of local tetrads (Vierbeine). In Section 2 of his paper he develops the tetrad formalism in a systematic manner. This was presumably independent work, since he does not give any reference to other authors. It was, however, mainly E. Cartan who demonstrated with his work [20] the usefulness of locally defined orthonormal bases – also called moving frames – for the study of Riemannian geometry.

⁽³⁾At the time it was thought by Weyl, and indeed by all physicists, that the two-component theory requires a zero mass. In 1957, after the discovery of parity nonconservation, it was found that the two-component theory could be consistent with a finite mass. See K. M. Case [19].

In the tetrad formalism the metric is described by an arbitrary basis of orthonormal vector fields $\{e_\alpha(x); \alpha = 0, 1, 2, 3\}$. If $\{e^\alpha(x)\}$ denotes the dual basis of 1-forms, the metric is given by

$$(12) \quad g(x) = \eta_{\mu\nu} e^\mu(x) \otimes e^\nu(x), \quad (\eta_{\mu\nu}) = \text{diag}(1, -1, -1, -1).$$

Weyl emphasizes, of course, that only a class of such local tetrads is determined by the metric: the metric is not changed if the tetrad fields are subject to spacetime-dependent Lorentz transformations:

$$(13) \quad e^\alpha(x) \rightarrow \Lambda^\alpha_\beta(x) e^\beta(x).$$

With respect to a tetrad, the connection forms $\omega = (\omega^\alpha_\beta)$ have values in the Lie algebra of the homogeneous Lorentz group:

$$(14) \quad \omega_{\alpha\beta} + \omega_{\beta\alpha} = 0.$$

(Indices are raised and lowered with $\eta^{\alpha\beta}$ and $\eta_{\alpha\beta}$, respectively.) They are determined (in terms of the tetrad) by the first structure equation of Cartan:

$$(15) \quad de^\alpha + \omega^\alpha_\beta \wedge e^\beta = 0.$$

(For a textbook derivation see, for example, [21], especially Sections 2.6 and 8.5.) Under local Lorentz transformations (13) the connection forms transform in the same way as the gauge potential of a non-Abelian gauge theory:

$$(16) \quad \omega(x) \rightarrow \Lambda(x)\omega(x)\Lambda^{-1}(x) - d\Lambda(x)\Lambda^{-1}(x).$$

The curvature forms $\Omega = (\Omega^\mu_\nu)$ are obtained from ω in exactly the same way as the Yang–Mills field strength from the gauge potential:

$$(17) \quad \Omega = d\omega + \omega \wedge \omega$$

(second structure equation).

For a vector field V , with components V^α relative to $\{e_\alpha\}$, the covariant derivative DV is given by

$$(18) \quad DV^\alpha = dV^\alpha + \omega^\alpha_\beta V^\beta.$$

Weyl generalizes this in a unique manner to spinor fields ψ :

$$(19) \quad D\psi = d\psi + \frac{1}{4}\omega_{\alpha\beta}\sigma^{\alpha\beta}\psi.$$

Here, the $\sigma^{\alpha\beta}$ describe infinitesimal Lorentz transformations (in the representation of ψ). For a Dirac field these are the familiar matrices

$$(20) \quad \sigma^{\alpha\beta} = \frac{1}{2}[\gamma^\alpha, \gamma^\beta].$$

(For two-component Weyl fields one has similar expressions in terms of the Pauli matrices.)

With these tools the action principle for the coupled Einstein–Dirac system can be set up. In the massless case the Lagrangian is

$$(21) \quad \mathcal{L} = \frac{1}{16\pi G} R - i\bar{\psi}\gamma^\mu D_\mu\psi,$$

where the first term is just the Einstein–Hilbert Lagrangian (which is linear in Ω). Weyl discusses, of course, immediately the consequences of the following two symmetries:

- (i) local Lorentz invariance,
- (ii) general coordinate invariance.

3.2. The New Form of the Gauge Principle. All this is a kind of a preparation for the final section of Weyl’s paper, which has the title “electric field.” Weyl says:

We come now to the critical part of the theory. In my opinion the origin and necessity for the electromagnetic field is in the following. The components $\psi_1 \psi_2$ are, in fact, not uniquely determined by the tetrad but only to the extent that they can still be multiplied by an arbitrary “gauge factor” $e^{i\lambda}$. The transformation of the ψ induced by a rotation of the tetrad is determined only up to such a factor. In special relativity one must regard this gauge factor as a constant because here we have only a single point-independent tetrad. Not so in General Relativity; every point has its own tetrad and hence its own arbitrary gauge factor; because by the removal of the rigid connection between tetrads at different points the gauge factor necessarily becomes an arbitrary function of position.

In this manner Weyl arrives at the gauge principle in its modern form and emphasizes: “From the arbitrariness of the gauge factor in ψ appears the necessity of introducing the electromagnetic potential.” The first term $d\psi$ in (19) has now to be replaced by the covariant gauge derivative $(d - iA)\psi$ and the nonintegrable scale

factor (1) of the old theory is now replaced by a phase factor:

$$\exp\left(-\int_{\gamma} A\right) \rightarrow \exp\left(-i \int_{\gamma} A\right),$$

which corresponds to the replacement of the original gauge group \mathbb{R} by the compact group $U(1)$. Accordingly, the original Gedankenexperiment of Einstein translates now to the Aharonov–Bohm effect, as was first pointed out by C. N. Yang in [22]. The close connection between gauge invariance and conservation of charge is again uncovered. Current conservation follows, as in the original theory, in two independent ways: it is a consequence both of the field equations for matter plus gauge invariance and of the field equations for the electromagnetic field plus gauge invariance. This corresponds to an identity in the coupled system of field equations which has to exist as a result of gauge invariance. All this is nowadays familiar to students of physics and does not need to be explained in more detail.

Much of Weyl's paper penetrated also into his classic book "The Theory of Groups and Quantum Mechanics" [23]. There he mentions also the transformation of his early gauge theoretic ideas: "This principle of gauge invariance is quite analogous to that previously set up by the author, on speculative grounds, in order to arrive at a unified theory of gravitation and electricity. But I now believe that this gauge invariance does not tie together electricity and gravitation, but rather electricity and matter."

When Pauli saw the full version of Weyl's paper he became more friendly and wrote [24]:

In contrast to the nasty things I said, the essential part of my last letter has since been overtaken, particularly by your paper in *Z. f. Physik*. For this reason I have afterward even regretted that I wrote to you. After studying your paper I believe that I have really understood what you wanted to do (this was not the case in respect of the little note in the *Proc. Nat. Acad.*). First let me emphasize that side of the matter concerning which I am in full agreement with you: your incorporation of spinor theory into gravitational theory. I am as dissatisfied as you are with distant parallelism and your proposal to let the tetrads rotate independently at different space points is a true solution.

In brackets Pauli adds:

Here I must admit your ability in Physics. Your earlier theory with $g'_{ik} = \lambda g_{ik}$ was pure mathematics and unphysical. Einstein was justified in criticizing and scolding. Now the hour of your revenge has arrived.

Then he remarks in connection with the mass problem:

Your method is valid even for the massive [Dirac] case. I thereby come to the other side of the matter, namely the unsolved difficulties of the Dirac theory (two signs of m_0) and the question of the two-component theory. In my opinion these problems will not be solved by gravitation . . . the gravitational effects will always be much too small.

Many years later, Weyl summarized this early tortuous history of gauge theory in an instructive letter [25] to the Swiss writer and Einstein biographer C. Seelig, which we reproduce in English translation.

The first attempt to develop a unified field theory of gravitation and electromagnetism dates to my first attempt in 1918, in which I added the principle of gauge invariance to that of coordinate invariance. I myself have long since abandoned this theory in favour of its correct interpretation: gauge invariance as a principle that connects electromagnetism not with gravitation but with the wave field of the electron. – Einstein was against it [the original theory] from the beginning, and this led to many discussions. I thought that I could answer his concrete objections. In the end he said “Well, Weyl, let us leave it at that! In such a speculative manner, without any guiding physical principle, one cannot make Physics.” Today one could say that in this respect we have exchanged our points of view. Einstein believes that in this field [Gravitation and Electromagnetism] the gap between ideas and experience is so wide that only the path of mathematical speculation, whose consequences must, of course, be developed and confronted with experiment, has a chance of success. Meanwhile my own confidence in pure speculation has diminished, and I see a need for a closer connection with experiments in quantum physics, since in my opinion it is not sufficient to unify Electromagnetism

and Gravity. The wave fields of the electron and whatever other irreducible elementary particles may appear must also be included.

Independently of Weyl, V. Fock [26] also incorporated the Dirac equation into General Relativity by using the same method. On the other hand, H. Tetrode [27], E. Schrödinger [28] and V. Bargmann [29] reached this goal by starting with space-time dependent γ -matrices, satisfying $\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}$. A somewhat later work by L. Infeld and B. L. van der Waerden [30] is based on spinor analysis.

4. Concluding Remarks

Gauge invariance became a serious problem when Heisenberg and Pauli began to work on a relativistically invariant Quantum Electrodynamics that eventually resulted in two important papers “On the Quantum Dynamics of Wave Fields” [32], [33]. Straightforward application of the canonical formalism led, already for the free electromagnetic field, to nonsensical results. Jordan and Pauli on the other hand, proceeded to show how to quantize the theory of the *free field* case by dealing only with the field strengths $F_{\mu\nu}(x)$. For these they found commutation relations at different space-time points in terms of the now famous invariant Jordan–Pauli distribution that are manifestly Lorentz invariant.

The difficulties concerned with applying the canonical formalism to the electromagnetic field continued to plague Heisenberg and Pauli for quite some time. By mid-1928 both were very pessimistic, and Heisenberg began to work on ferromagnetism.⁽⁴⁾ In the fall of 1928 Heisenberg discovered a way to bypass the difficulties. He added the term $-\frac{1}{2}\varepsilon(\partial_\mu A^\mu)^2$ to the Lagrangian, in which case the component π_0 of the canonical momenta

$$\pi_\mu = \frac{\partial L}{\partial_0 A_\mu}$$

does no more vanish identically ($\pi_0 = -\varepsilon\partial_\mu A^\mu$). The standard canonical quantization scheme can then be applied. At the end of all calculations one could then take the limit $\varepsilon \rightarrow 0$.

⁽⁴⁾Pauli turned to literature. In a letter of 18 February 1929 he wrote from Zürich to Oskar Klein: “For my proper amusement I then made a short sketch of a utopian novel which was supposed to have the title ‘Gulivers journey to Urania’ and was intended as a political satire in the style of Swift against present-day democracy. [...] Caught in such dreams, suddenly in January, news from Heisenberg reached me that he is able, with the aid of a trick . . . to get rid of the formal difficulties that stood against the execution of our quantum electrodynamics.” [6]

In their second paper, Heisenberg and Pauli stressed that the Lorentz condition cannot be imposed as an operator identity but only as a supplementary condition selecting admissible states. This discussion was strongly influenced by a paper of Fermi from May 1929.

For this and the further main developments during the early period of quantum field theory, we refer to Chapter 1 of [34].

As in Weyl's work General Relativity also played a crucial role in Pauli's discovery of non-Abelian gauge theories. (See Pauli's letters to Pais and Yang in Vol. 4 of [6].) He arrived at all basic equations through dimensional reduction of a generalization of Kaluza–Klein theory, in which the internal space becomes a two-sphere. (For a description in modern language, see [4].)

In contrast, in the work of Yang and Mills General Relativity played no role. In an interview Yang said on this in 1991:

“It happened that one semester [around 1970] I was teaching General Relativity, and I noticed that the formula in gauge theory for the field strength and the formula in Riemannian geometry for the Riemann tensor are not just similar – they are, in fact, the same if one makes the right identification of symbols! It is hard to describe the thrill I felt at understanding this point.”

The developments after 1958 consisted in the gradual recognition that – contrary to phenomenological appearances – Yang–Mills gauge theory could describe weak and strong interactions. This important step was again very difficult, with many hurdles to overcome.

References

- [1] H. Weyl, *Gravitation und Elektrizität*. Sitzungsber. Akademie der Wissenschaften Berlin, 465–480 (1918). Siehe auch die *Gesammelten Abhandlungen*. 6 Vols. Ed. K. Chadrasekharan, Springer-Verlag (an English translation is given in [8]).
- [2] N. Straumann, *Zum Ursprung der Eichtheorien bei Hermann Weyl*. Physikalische Blätter **43** (11), 414–421 (1987).
- [3] H. Weyl, *Elektron und Gravitation*. Z. Phys. **56**, 330 (1929).
- [4] L. O’Raifeartaigh and N. Straumann, *Gauge Theory: Historical Origins and Some Modern Developments*. Rev. Mod. Phys. **72**, 16–23 (2000).
- [5] *The Collected Papers of Albert Einstein*, Vols. 1–10 (Princeton University Press, 1987). See also: [<http://www.einstein.caltech.edu/>].
- [6] W. Pauli, *Wissenschaftlicher Briefwechsel mit Bohr, Einstein, Heisenberg u.a.* Vol. 1–4, edited by K. von Meyenn (Springer-Verlag, New York).

- [7] H. Weyl, *Space · Time · Matter*. Translated from the 4th German Edition. London: Methuen 1922. *Raum · Zeit · Materie*, 8. Auflage, Springer-Verlag (1993).
- [8] L. O’Raifeartaigh, *The Dawning of Gauge Theory*. Princeton University Press (1997).
- [9] J. Audretsch, F. Gähler and N. Straumann, *Comm.Math.Phys.* **95**, 41 (1984).
- [10] W. Pauli, *Zur Theorie der Gravitation und der Elektrizität von H. Weyl*. *Physikalische Zeitschrift* **20**, 457–467 (1919).
- [11] W. Pauli, *Relativitätstheorie*. Encyklopädie der Mathematischen Wissenschaften 5.3, Leipzig: Teubner, 539–775 (1921). W. Pauli, *Theory of Relativity*. Pergamon Press, New York (1958).
- [12] F. London, *Quantenmechanische Deutung der Theorie von Weyl*. *Z. Phys.* **42**, 375 (1927).
- [13] E. Schrödinger, *Z. Phys.* **12**, 13 (1922).
- [14] V. Raman and P. Forman, *Hist. Studies. Phys. Sci.* **1**, 291 (1969).
- [15] E. Schrödinger, *Centenary Celebration of a Polymath*, ed. C. Kilmister, Cambridge University Press (1987).
- [16] V. Fock, *Z. Phys.* **39**, 226 (1926).
- [17] H. Weyl, *Selecta*. Birkhäuser-Verlag 1956.
- [18] E. Wigner *Z. Phys.* **53**, 592 (1929).
- [19] C. M. Case, *Phys. Rev.* **107**, 307 (1957).
- [20] E. Cartan, *Leçons sur la Géométrie des Espaces de Riemann*, Gauthier-Villars, Paris, 1928; 2nd edn., 1946.
- [21] N. Straumann, *General Relativity, With Applications to Astrophysics*, Texts and Monographs in Physics, Springer Verlag, 2004.
- [22] C. N. Yang, *Hermann Weyl’s Contribution to Physics*. In: *Hermann Weyl*, Edited by K. Chandrasekharan, Springer-Verlag 1980.
- [23] H. Weyl, *Gruppentheorie und Quantenmechanik*. Wissenschaftliche Buchgesellschaft, Darmstadt 1981 (Nachdruck der 2. Aufl., Leipzig 1931). Engl. translation: “Group Theory and Quantum Mechanics,” Dover, New York, 1950.
- [24] [6], p. 518.
- [25] In Carl Seelig: *Albert Einstein*. Europa Verlag Zürich 1960, p. 274.
- [26] V. Fock, *Z. Phys.* **57**, 261 (1929).
- [27] H. Tetrode, *Z. Phys.* **50**, 336 (1928).
- [28] E. Schrödinger, *Sitzungsber. Preuss. Akad. Wiss.*, **105** (1932).
- [29] V. Bargmann, *Sitzungsber. Preuss. Akad. Wiss.*, **346** (1932).
- [30] L. Infeld and B. L. van der Waerden, *Sitzungsber. Preuss. Akad. Wiss.*, **380** and **474** (1932).
- [31] J. D. Jackson and L. B. Okun, *Historical Roots of Gauge Invariance*. *Rev. Mod. Phys.* **73**, 663–680 (2001).
- [32] W. Heisenberg and W. Pauli. *Zur Quantenelektrodynamik der Wellenfelder. I*. *Zeitschrift für Physik*, **56**, 1–61 (1929).
- [33] W. Heisenberg and W. Pauli. *Zur Quantenelektrodynamik der Wellenfelder. II*. *Zeitschrift für Physik*, **59**, 168–190 (1930).
- [34] S. S. Schweber, *Quantum Electrodynamics and the Men Who Made It: Dyson, Feynman, Schwinger, and Tomonaga*. Princeton University Press (1994), Princeton, N.J.
- [35] O. Klein, *Z. Phys.* **37**, 895 (1926); for an English translation see [8].

These proceedings contain most of the contributions to the Göttingen–Jerusalem Conference 2008 on „Symmetries in Algebra and Number Theory“ including three addresses given at the conference opening, and two contributions to the Satellite Conference „On the Legacy of Hermann Weyl“. The contributions are survey articles or report on recent work by the authors, for example new results on the famous Leopoldt conjecture.



ISBN: 978-3-940344-96-0

Universitätsverlag Göttingen