GAFA, Geom. funct. anal.
Special Volume – GAFA2000, 705 – 741
1016-443X/00/S20705-37 $ 1.50+0.20/0

GAFA Geometric And Functional Analysis

PERSPECTIVES ON THE ANALYTIC THEORY OF L-FUNCTIONS

H. IWANIEC AND P. SARNAK

Contents:

1 Introduction and Background

To the general mathematician L-functions might appear to be an esoteric and special topic in number theory. We hope that the discussion below will convince the reader otherwise. Time and again it has turned out that the crux of a problem lies in the theory of these functions. At some level it is not entirely clear to us why L-functions should enter decisively, though in hindsight one can give reasons. Our plan is to introduce L-functions and describe the central problems connected with them. We give a sample (this is certainly not meant to be a survey) of results towards these conjectures as well as some problems that can be resolved by finessing these conjectures. We also mention briefly some of the successful present-day tools and the role they might play in the big picture.

An L-function is a type of generating function formed out of local data associated with either an arithmetic-geometric object (such as an abelian variety defined over a number field) or with an automorphic form (it is expected that the latter set contains the former one, Shimura-Taniyama for

Supported in part by NSF Grants DMS-98-01642 and DMS-99-70386 as well as by grants to the Institute for Advanced Study from the Ambrose Monell Foundation, the Hansmann Membership and the Veblen Fund.

special cases and Langlands in general). Fix a number field K (i.e. a finite algebraic extension of $\mathbb{Q}$), the reader will not lose too much by restricting to $\mathbb{Q}$. An L-function takes the form of a product of degree $m \geq 1$ over all primes p of K

$$L(s) = \prod_p L_p(s) , \tag{1}$$

where the local factors are

$$L_p(s) = \prod_{j=1}^{m} \left(1 - \alpha_j(p)\,(Np)^{-s}\right)^{-1} \tag{2}$$

for suitable complex numbers $\alpha_j(p)$ and where Np is the norm of p. As a function of s this product converges absolutely for $\Re(s) > 1$ (see below) and we can multiply out to get the series

$$L(s) = \sum_{a \neq 0} c(a) N(a)^{-s} , \tag{3}$$

the sum being over integral ideals.

We give some concrete examples all being for $K = \mathbb{Q}$.

(1) The Riemann zeta function ($m = 1$)

$$\zeta(s) = \prod_p \left(1 - p^{-s}\right)^{-1} = \sum_{n=1}^{\infty} n^{-s} . \tag{4}$$

(2) Dirichlet L-functions ($m = 1$)

$$L(s,\chi) = \prod_p \left(1 - \chi(p)p^{-s}\right)^{-1} = \sum_{n=1}^{\infty} \chi(n) n^{-s} , \tag{5}$$

where χ is a character of the group of primitive residue classes $a(\mathrm{mod} q)$ (more precisely a multiplicative function on $\mathbb{Z}$ which is periodic of period q). The minimal period q is called the conductor of χ.

(3) For $m = 2$ we give the example of the L-functions of elliptic curves defined over $\mathbb{Q}$. Let E be such a nonsingular curve given by the equation

$$E: \quad y^2 = x^3 + ax + b , \tag{6}$$

$a, b \in \mathbb{Q}$. For a prime p at which reducing E modulo p yields a nonsingular curve over $\mathbb{F}_p$ (the field with p-elements), one defines the local factor $L_p(s,E)$ as follows: Let $N_E(p)$ be the number of solutions of (6) with x, y in $\mathbb{F}_p$ (not counting the point at infinity) and let $a_E(p) = p - N_E(p)$. Define

$$L_p(s,E) = \left(1 - \frac{a_E(p)}{\sqrt{p}}\, p^{-s} + p^{-2s}\right)^{-1} . \tag{7}$$

Note that this is not the standard algebraists normalization but it is very convenient for analytic purposes. The L-function $L(s,E)$ is defined by

$$L(s,E) = \prod_p L_p(s,E) , \tag{8}$$

where at primes p for which E has singular reduction (there being finitely many of these) special care must be taken in defining the local factor $L_p(s,E)$.

(4) Again we take $m = 2$ and give an example of a holomorphic modular form and its L-function. Let $\mathbb{H}$ be the upper half plane. For $z \in \mathbb{H}$ and $m \geq 1$ set

$$F(z) = \sum_{\mu \in \mathbb{Z}[\sqrt{-1}]} \mu^{4m} e^{2\pi i N(\mu) z} , \tag{9}$$

where $N(\mu) = \mu\bar{\mu}$. It turns out that $F(z)$ is a holomorphic modular form of weight $k = 4m+1$ for the subgroup

$$\Gamma_0(4) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z}); 4|c \right\}$$

of the modular group $SL_2(\mathbb{Z})$. That is to say it transforms appropriately under $z \to \frac{az+b}{cz+d}$ for $\gamma = \left(\begin{smallmatrix} a b \\ c d \end{smallmatrix}\right) \in \Gamma_0(4)$, with nebentypus character $\left(\frac{-4}{d}\right)$. Its L-function is

$$L(s,F) = \sum_{t=1}^{\infty} \left(\sum_{\substack{\mu=\mu_1+i\mu_2 \\ \mu_1 \geq 0, \mu_2 > 0 \\ N(\mu)=t}} \left(\frac{\mu}{|\mu|}\right)^{4m} \right) t^{-s} = \prod_p L_p(s,F) . \tag{10}$$

For $p \equiv 3(4)$

$$L_p(s,F) = (1 - p^{-2s})^{-1} \tag{11}$$

while for $p \equiv 1(4)$

$$L_p(s,F) = (1 - c(p)p^{-s} + p^{-2s})^{-1} , \tag{12}$$

where

$$c(p) = \frac{1}{4} \sum_{N(\mu)=p} \left(\frac{\mu}{|\mu|}\right)^{4m} . \tag{13}$$

Alternatively we have

$$L(s,F) = \prod_p \left(1 - \lambda(p)(Np)^{-s}\right)^{-1} , \tag{14}$$

where p runs over the prime ideals of $\mathbb{Q}\left(\sqrt{-1}\right)$ and λ is the "Grossencharacter" given by $\lambda((\alpha)) = (\alpha/|\alpha|)^{4m}$. The form F is special among the modular forms for $\Gamma_0(4)$ in having an expression (14) in terms of a quadratic extension. It is an example of a "CM" modular form.

(5) An example of a non-CM holomorphic modular form is the popular
$$\Delta(z) = q \prod_{n=0}^{\infty} (1-q^n)^{24} := \sum_{n=1}^{\infty} \tau(n) q^n , \tag{15}$$
where $q = e^{2\pi i z}$ is a holomorphic cusp form of weight 12 for $SL(2,\mathbb{Z})$ (that is $\Delta\left(\frac{az+b}{cz+d}\right) = (cz+d)^{12}\Delta(z)$ for $\binom{ab}{cd} \in SL(2,\mathbb{Z})$). Its L-function is
$$L(s,\Delta) = \sum_{n=1}^{\infty} \frac{\tau(n)}{n^{11/2}} n^{-s} = \prod_p \left(1 - \frac{\tau(p)}{p^{11/2}} p^{-s} + p^{-2s}\right)^{-1} . \tag{16}$$
That it factors into an "Euler product" as indicated is a consequence of Δ being a Hecke eigenform (see below).

(6) Our last example is a Maass cusp form for $GL(2,\mathbb{Z})$ on $\mathbb{H}$. That is a real-analytic function $\phi(z)$ on $\mathbb{H}$ satisfying
   (a) $\phi(\gamma z) = \phi(z)$ for $\gamma \in SL(2,\mathbb{Z}) = \Gamma$
   (b) $\phi(-\bar{z}) = \phi(z)$
   (c) $\Delta\phi + \lambda\phi = 0$, $\lambda > \frac{1}{4}$
   (Δ being the Laplacian for the hyperbolic metric)
   (d) ϕ is square integrable on $\Gamma\backslash\mathbb{H}$ (in this particular example this property is equivalent to being a *cusp* form).

These ϕ's are far less tangible than the previous examples. Indeed that there are any such ϕ's is far from obvious. The only proof of their existence is through the trace formula (indeed this demonstration was part of Selberg's original motivation for developing the trace formula). Our present understanding is that these elusive forms exist in abundance only for Γ's as above, such as congruence subgroups of $SL_2(\mathbb{Z})$ [PhS], [Wo]. The Hecke operators T_n defined by
$$(T_n\phi)(z) = \frac{1}{\sqrt{n}} \sum_{ad=n} \sum_{b (\mathrm{mod} d)} \phi\left(\frac{az+b}{d}\right)$$
act on these eigenspaces, they commute with each other and are self-adjoint on $L^2(\Gamma\backslash\mathbb{H})$. We may therefore simultaneously diagonalize and assume that
$$T_n\phi = \lambda_\phi(n)\phi, \text{ for all } n \geq 1 . \tag{17}$$
For such Maass eigenforms we have corresponding degree two L-functions:
$$L(s,\phi) = \sum_{n=1}^{\infty} \lambda_\phi(n) n^{-s} = \prod_p L_p(s,\phi) , \tag{18}$$

where

$$L_p(s,\phi) = (1-\alpha_{1,\phi}(p)p^{-s})^{-1}(1-\alpha_{2,\phi}(p)p^{-s})^{-1} \tag{19}$$

and

$$\alpha_{1,\phi}(p)\,\alpha_{2,\phi}(p) = 1, \qquad \alpha_{1,\phi}(p)+\alpha_{2,\phi}(p) = \lambda_\phi(p)\ . \tag{20}$$

That completes our list of examples of L-functions.

We now dive in with the general modern definition of an automorphic cusp form and its L-function. We consider only the group GL_m since it is expected from general conjectures of Langlands [La] that all L-functions are products of these standard L-functions (we emphasize that other groups and even the exceptional groups play an important role in understanding these L-functions (see below)). Let $\mathbb{A}_K$ be the adele ring of K, that is the restricted product, $\Pi_v K_v$ over the completions of K. An automorphic cusp form F on $GL_m(\mathbb{A}_K)$ is an irreducible representation of $GL_m(\mathbb{A}_K)$ occurring in $L_0^2(GL_m(K)\backslash GL_m(\mathbb{A}_K))$ under the right regular representation of $GL_m(\mathbb{A}_K)$. Here we are assuming that all functions transform under the action of the center by a unitary central (idele class) character and the subscript zero refers to the cuspidal subspace [GelP]. The relation to the classical description of modular forms is that there are special functions in such an irreducible representation which are classical modular forms. Now an F as above is of the form $\otimes_v F_v$ where v ranges over all places of K (finite and archimedean) and F_v is a unitary (in fact generic [JS]) representation of $GL_m(K_v)$. Using suitable parameters of the local representation of F_v (Satake parameters [Sa2] if F_v is unramified, which is the case for all but a finite number of parameters, and Langlands parameters in general) one defines the numbers $\alpha_{F,j}(v)$, $j=1,\dots,m$ in (2) above. In particular, this yields the definition of the local factors $L_v(s,F) = L(s,F_v)$ for v finite. At the archimedean places there are similar parameters for the local representations of $GL_m(\mathbb{R})$ or $GL_m(\mathbb{C})$. The local L-factors take the form

$$L(s,F_v) = \prod_{j=1}^{m} \Gamma_v\big(s-\mu_{j,F}(v)\big)\ , \tag{21}$$

where

$$\Gamma_v(s) = \begin{cases} \pi^{-s/2}\,\Gamma\left(\frac{s}{2}\right), & \text{if } K_v \sim \mathbb{R} \\ (2\pi)^{-s}\,\Gamma(s), & \text{if } K_v \sim \mathbb{C}\ . \end{cases} \tag{22}$$

The standard global L-function of F is then

$$L(s,F) = \prod_{v \text{ finite}} L(s,F_v). \tag{23}$$

The product converges absolutely for $\Re(s) > 1$. Moreover, the analogue of Riemann's analytic continuation and functional equation are known in this generality (Hecke [Hec], Godement-Jacquet [GoJ] and see also Tamagawa [T]). Define the completed function

$$\Lambda(s, F) = \Bigg(\prod_{v \text{ archim.}} L(s, F_v) \Bigg) \cdot L(s, F) . \tag{24}$$

Then $\Lambda(s, F)$ extends to an entire function (except in the case $m = 1$ and F is the trivial representation when Λ has poles at $s = 0$ and $s = 1$) and satisfies the functional equation

$$\Lambda\big(1 - s, \widetilde{F}\big) = \bar{\epsilon}_F \, N_F^{s-\frac{1}{2}} \Lambda(s, F) . \tag{25}$$

Here $N_F \geq 1$ is an integer called the conductor of F, ϵ_F is the root number (which has modulus 1) and $\tilde{F}$ is the representation contragredient to F.

The cuspidal spectrum of $L^2(GL_m(K)\backslash GL_m(\mathbb{A}_K))$ is discrete so that the set of standard L-functions is countable. In the form that we have described them, these L-functions are not unrelated to each other (see for example (10) and (14) above). It is known (see [AC] that each $L(s, F)$ for K is a product of $L(s, F')$ for $\mathbb{Q}$ (with $m' = md$, $d = \deg(K/\mathbb{Q})$). For many purposes it is convenient to think of $L(s, F)$ over K rather than of larger degree over $\mathbb{Q}$. The cuspidal standard $L(s, F)$ over $\mathbb{Q}$ are all independent of each other and they form the basic building blocks for all L-functions.

One can form more general L-functions from these basic ones, that is the tensor powers. Very special cases of these are known to have analytic continuations and functional equations. There are at present two methods to attack this problem of continuation both depend on the analytic properties of Eisenstein series. These are the Rankin-Selberg method, see [Bu], and the Langlands-Shahidi method [Sh1]. For example given F and F' automorphic cuspidal on GL_m and $GL_{m'}$ respectively, then $L(s, F \otimes F')$ is an L-function of degree mm' whose local factor at a place v of K at which both F and F' are unramified is:

$$L(s, F_v \otimes F_v') = \prod_{j=1}^{m} \prod_{k=1}^{m'} \big(1 - \alpha_{F,j}\, \alpha_{F',k}\, N(v)^{-s}\big)^{-1} . \tag{26}$$

The precise analytic continuations and functional equations for these are known [JPS]. The function $L(s, F \otimes \tilde{F})$ has non-negative coefficients in its expansion (3). This together with its analytic properties (i.e. pole at $s = 1$) imply that the product (23) converges absolutely in $\Re(s) > 1$.

Another special case that is known [G] is the degree 8 triple product of GL_2 forms. Let F, G and H be three cusp forms on GL_2/K. At a place v

where F, G and H are unramified define the local L-function of degree 8 by

$$L(s, F_v \otimes G_v \otimes H_v) = \prod_{\substack{\epsilon_j \in \{1,2\} \\ j=1,2,3}} \left(1 - \alpha_{F,\epsilon_1}(v)\, \alpha_{G,\epsilon_2}(v)\, \alpha_{H,\epsilon_3}(v) N(v)^{-s}\right)^{-1}. \quad (27)$$

Set

$$L(s, F \otimes G \otimes H) = \prod_v L(s, F_v \otimes G_v \otimes H_v). \quad (28)$$

Then $L(s, F \otimes G \otimes H)$ has an analytic continuation and functional equation $s \to 1 - s$.

Some special cases of the symmetric power L-functions of GL_2 forms are known to have analytic continuations and functional equations. For $n \geq 1$ and F on GL_2/K define the local factor of the n-th symmetric power (at an unramified place) by

$$L(s, \mathrm{sym}^n F_v) = \prod_{j=0}^{n} \left(1 - \left(\alpha_{F,1}(v)\right)^j \left(\alpha_{F,2}(v)\right)^{n-j} N(v)^{-s}\right)^{-1}. \quad (29)$$

The global n-th symmetric power L-function $L(s, \mathrm{sym}^n F)$ is defined to be the product of these local factors. For $n = 1$ this is just the standard L-function. For $n = 2$ the analytic properties were established by Shimura [Shi2]. Recently, Kim and Shahidi [KiSh] established the expected analytic properties for $n = 3$.[1] Their proof uses at one point the unitary dual of the exceptional group G_2!

The above discussion has indicated why L-functions of automorphic forms enjoy certain analytic properties. For the examples (3) of L-functions of elliptic curves over $\mathbb{Q}$ this follows from the spectacular progress by Wiles [Wi] and [TaW] which asserts that "elliptic curves over $\mathbb{Q}$ are modular." This implies that $L(s, E)$ is an $L(s, F)$ for a suitable holomorphic weight 2 cusp form F on a congruence subgroup of $SL_2(\mathbb{Z})$. Indeed, the construction of automorphic forms (and hence of L-functions) from arithmetic-geometric settings is one of the major thrusts of modern number theory. Our interest here is beyond this and at the same time much older. That is, we are given an automorphic form and its L-function and we investigate its properties (beyond just analyticity) and their applications.

[1] Added in proof: Recently they have also established this for $n = 4$.

2 Fundamental Conjectures

We turn to some of the basic problems which until their resolution are expected to be a focus of the subject. The first is a well known generalization of Riemann's Conjecture.

A) **Grand Riemann Hypothesis** (GRH)
The zeros ρ_F of any $\Lambda(s, F)$ have real part equal to $\frac{1}{2}$.

Comments:

(A1) Crisp, falsifiable and far reaching this conjecture is the epitome of what a good conjecture should be. Moreover, it has many striking consequences (some described below). One of its powers lies in that it ensures uniform (up to square root of the number of terms - like random numbers) cancellations in sums over $c_F(a)$ or $c_F(p)$ (as in (2) and (3)). It is in this form that one often uses it in applications to problems in which the local data in $c_F(p)$ is being used to analyze something global and *visa versa*. In practice, GRH is often used as a working hypothesis (and an apparently very reliable one at that) in that one proceeds by using it, and in this way many results are established under GRH. However, there have been sufficiently powerful advances in the theory that in a number of cases one can dispense with GRH and the desired result is established unconditionally.

(A2) The true strength of GRH lies in the statement for the general $L(s, F)$, or at least for some infinite family of L-functions such as Dirichlet L-functions $L(s, \chi)$. For example, the case of $\zeta(s)$ itself has few consequences (it is of course directly equivalent to the size of the remainder term in the Prime Number Theorem). For a recent description and discussion of RH, see Bombieri [B2]. There is no $L(s, F)$ for which GRH is known. For families such as $L(s, \chi)$, there are results, "density theorems", which assert that almost all their zeros lie near $\Re(s) = \frac{1}{2}$. These can often be used as a substitute for GRH (see section 4 below).

(A3) For $\zeta(s)$, $L(s, \chi)$ and some $GL_2/\mathbb{Q}$ L-functions extensive numerical experimentations have confirmed GRH in impressive ranges. This is important supporting evidence for the truth of GRH. The function field analogues (see section 3) are known to be true and this is further strong evidence in favor of GRH.

The direct results that have been established towards GRH are modest. The method of Hadamard and de la Vallée Poussin for $\zeta(s)$ (in their proof of the Prime Number Theorem) can be used together with the analytic

›roperties of $L(s,F)$ and $L(s,F\otimes\tilde{F})$ to show that $L(1+it,F)\neq 0$ for $\in\mathbb{R}$, cf. [R]. The lower bounds for $|L(1+it,F)|$ that one obtains this vay are all roughly of the same quality[2] except for one major lacuna. That s the case of $L(1,\chi)$, χ quadratic over $\mathbb{Q}$ (this being the first instance ›f nonvanishing of L-functions and is due to Dirichlet in his proof of the nfinitude of primes in arithmetic progressions). For this case instead of the ower bound of $(\log q)^{-1}$ for $L(1,\chi_q)$ the best known *effective* lower bound s $L(1,\chi_q)\gg\frac{\log q}{\sqrt{q}}$, (if q is prime and slightly weaker in general [Gol], [GrZ]). The last is an excellent example of the use of GL_2 L-functions (in particular L-functions of elliptic curves of high rank) to give information about GL_1 L-functions. *Ineffectively* Siegel [Si2], following Landau, established the lower bound; given $\epsilon>0$ there is $C_\epsilon>0$ such that for any $q>1$

$$L(1,\chi_q)\geq C_\epsilon q^{-\epsilon}\,. \tag{30}$$

GRH implies the so-called Lindelöff Hypothesis which, if true, is a very useful bound for L-functions on the critical line.

Precisely for the purpose of estimating $L(\frac{1}{2}+it,F)$ we introduce the quantity (the "analytic conductor")

$$C(F,t)=N_F\prod_{j=1}^{m}\prod_{v\,\text{archim.}}\left(1+|\mu_{j,F}(v)+it|^{d(v)}\right)\,, \tag{31}$$

where for v archimedean

$$d(v)=1\quad\text{if}\quad K_v=\mathbb{R}\quad\text{and}\quad d(v)=2\quad\text{if}\quad K_v=\mathbb{C}\,. \tag{32}$$

Fix m and K. Let $d=\deg(K/\mathbb{Q})$. The Lindelöff Hypothesis asserts that for any $\epsilon>0$,

$$L\left(\tfrac{1}{2}+it,F\right)\underset{\epsilon}{\ll}\left(C(F,t)\right)^{\epsilon}\,. \tag{33}$$

It follows from the functional equation for $\Lambda(s,F)$ and the convexity bounds of Phragmen-Lindelöff that for $\epsilon>0$

$$L\left(\tfrac{1}{2}+it,F\right)\underset{\epsilon}{\ll}\left(C(F,t)\right)^{\frac{1}{4}+\epsilon}\,. \tag{34}$$

Because of its many applications we single out the following problem as a basic one.

B) **Subconvexity Problem**

For m and K fixed to show there is $\delta>0$ such that

$$L\left(\tfrac{1}{2}+it,F\right)\ll\left(C(F,t)\right)^{\frac{1}{4}-\delta}.$$

[2] In the special case of $\zeta(s)$ some improvements have been given using far reaching methods of I.M. Vinogradov.

Actually in applications we usually have some subfamily (i.e. only one of the parameters t, N_F or $\| F \|_{\text{archim.}}$ varies) and we seek subconvexity estimates uniformly for the subfamily. This problem (B) is solved in a number of cases and we discuss this and some of their applications in section 6.

Next we discuss the generalized Ramanujan Conjecture. It is the local analogue of GRH and is a spectral problem concerning the local representations F_v of $GL_m(K_v)$ of the global automorphic cuspidal representation F. It asserts that for a place v at which F_v is unramified, F_v should be tempered (see [Sa1]). Equivalently this can be stated in terms of $L(s, F_v)$ as follows:

C) **Generalized Ramanujan Conjecture** (GRC)

Let F be an automorphic cuspidal representation of $GL_m(\mathbb{A}_k)$ which is unramified at a place v. Then for v finite $|\alpha_{j,F}(v)| = 1$ while for v archimedean, $\Re(\mu_{j,F}(v)) = 0$.

Comments:

(C1) Again this is a clean and far reaching conjecture. It is in the background in many applications of the spectral theory of automorphic forms to problems in analytic number theory.

(C2) The original problem of Ramanujan was concerned with the case $F = \Delta(z)$ (see (15) above). In this case GRC is equivalent to the original Ramanujan Conjecture:
$$|\tau(p)| \leq 2p^{\frac{11}{2}} . \tag{35}$$
For this case and more generally the case of holomorphic cusp forms of even integral weight for congruence subgroups of $SL(2, \mathbb{Z})$ the conjecture was established by Deligne [D2].

(C3) For $K = \mathbb{Q}$ and $\mathbb{Q}_v = \mathbb{R}$, GRC is equivalent to the Selberg Eigenvalue Conjecture; that for any N
$$\lambda_1\big(\Gamma(N)\backslash\mathbb{H}\big) \geq \tfrac{1}{4} . \tag{36}$$
Here $\Gamma(N)$ is the principal congruence subgroup of $\Gamma(1) = SL(2, \mathbb{Z})$, that is $\Gamma(N) = \{\gamma \in \Gamma(1); \gamma \equiv \left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right) \operatorname{mod} \mathrm{N}\}$, and λ_1 is the smallest eigenvalue of the Laplacian on the cuspidal space $L_0^2(\Gamma(N)\backslash\mathbb{H})$.

There are nontrivial and very useful general bounds towards GRC. Firstly, there are purely local bounds which use only that F_v is unitary and generic [JS] (these properties of F_v follow from F being a cusp form). These bounds are
$$\Big| \log_{N(v)} \big|\alpha_{j,F}(v)\big| \Big| < \tfrac{1}{2}, \text{ for } v \text{ finite} \tag{37}$$

$$\left|\Re\left(\mu_{j;F}(v)\right)\right| < \tfrac{1}{2}, \quad \text{for } v \text{ archimedean} . \tag{38}$$

This can be viewed as the analogue of the convexity bound for L-functions. For this problem a "subconvex" bound is known in general [LuRS]. For F on $GL_m(\mathbb{A}_K)$ cuspidal

$$\left|\log_{N(v)} \alpha_{j,F}(v)\right| \le \tfrac{1}{2} - \tfrac{1}{m^2+1} \ , \text{ for } v \text{ finite} \tag{39}$$

$$\left|\Re\left(\mu_{j,F}(v)\right)\right| \le \tfrac{1}{2} - \tfrac{1}{m^2+1} \ , \text{ for } v \text{ archimedean} . \tag{40}$$

The proof of this is global relying on the analytic properties of Rankin-Selberg L-functions as well as a technique of persistence of zeros for families of L-functions (in this case twists by ray class characters) and also a positivity argument. This theme of families will recur often in what follows. Combining the above bounds for $m = 3$ together with the Gelbart-Jacquet [GeJ] symmetric square lift from GL_2 to GL_3 yields an improved bound[3] for GL_2:

$$\left|\log_{N(v)} \alpha_{j,F}(v)\right| \le \tfrac{1}{5} \ , \text{ for } v \text{ finite} \tag{41}$$

$$\left|\Re\left(\mu_{j,F}(v)\right)\right| \le \tfrac{1}{5} \ , \text{ for } v \text{ archimedean} . \tag{42}$$

Remarkably the last at finite places can be proven by a quite different method, see Shahidi [Sh2] who uses exceptional groups. For $K = \mathbb{Q}$ and $Q_\infty \cong \mathbb{R}$ (42) yields $\lambda_1(\Gamma(N)\backslash\mathbb{H}) \ge \frac{21}{100}$ for the Selberg problem (36) above. This being greater than $\frac{3}{16}$ has significant corollaries (see section 7). The use of the family of twists of L-functions by Dirichlet characters for the purpose of obtaining estimates towards the GRC for Maass forms for $GL_2/\mathbb{Q}$, at finite places p, was introduced in [DuI1]. In that case it was used to exploit the extra functional equations afforded by the family as well as to overcome the lack of positivity for the coefficients of the symmetric square L-function. Their method may be used to give a slight improvement of (41) in the case $K = \mathbb{Q}$.

It was observed early on (Tate, Langlands, Serre) that the expected analytic properties (i.e. meromorphic continuation and location of poles) of the symmetric power L-functions imply GRC as well as conjectures about the distribution of the "angles" $\{\alpha_{F,1}(v), \dots, \alpha_{F,m}(\nu)\}$ as $N(v) \to \infty$ (the so-called Sato-Tate Conjectures). For F on $GL_2/\mathbb{Q}$ and $n \ge 1$ consider

$$G_n(s) = L\left(s, \operatorname{sym}^n F \otimes \widetilde{\operatorname{sym}^n F}\right) , \tag{43}$$

[3] We have just learned that for this case, Kim and Shahidi have established the improved bound replacing $\frac{1}{5}$ by $\frac{5}{34}$ in (41) and (42). Added in proof: Their methods when combined with the method of twisting leads for the case of $K = \mathbb{Q}$ to bounds with $\frac{1}{5}$ replaced by $\frac{7}{64}$ (see [KiS])

where $L(s, \mathrm{sym}^n F)$ is given following (29). The series for $G_n(s)$ is of the form

$$G_n(s) = \sum_{m=1}^{\infty} b(m) m^{-s} \tag{44}$$

with $b(m) \geq 0$. If as expected $G_n(s)$ is analytic for $\Re(s) > 1$ (it certainly has a pole at $s = 1$) then the positivity of the coefficients easily implies that

$$b(m) \ll_\epsilon m^{1+\epsilon}, \text{ for any } \quad \epsilon > 0\ . \tag{45}$$

Now examining the coefficients of (43) we have for p a prime at which F in unramified and $e \geq 1$

$$\left| \sum_{j=0}^{n} \left[\left(\alpha_{F,1}(p)\right)^j \left(\alpha_{F,2}(p)\right)^{n-j} \right]^e \right|^2 \leq e\, b(p^e)\ . \tag{46}$$

Hence, combining (45) and (46), the fact that $|\alpha_{F,1}(p)\alpha_{F_2}(p)| = 1$ and letting $e \to \infty$ and $\epsilon \to 0$ we conclude that

$$\max\left\{ \left|\alpha_{F,1}(p)\right|, \left|\alpha_{F,2}(p)\right| \right\} \leq p^{\frac{1}{2n}}\ . \tag{47}$$

Thus the knowledge that $G_n(s)$ is analytic for $\Re(s) > 1$ for all n yields GRC for F. The GRC for $p = \infty$, i.e. the Selberg Conjecture would also follow from similar considerations. The behavior of the distribution of the angles $\{\alpha_{F,1}(p),\ \alpha_{F,2}(p)\}$ requires a little more, that is the analytic properties of $G_n(s)$ up to and including $\Re(s) = 1$.

The last fundamental conjecture that we mention is the Birch and Swinnerton-Dyer Conjecture. This conjecture was discovered experimentally (i.e. through numerical experimentation) in looking for elliptic curve analogues of the Siegel Mass Formula (see section 6) for quadratic forms.

D) **Birch and Swinnerton-Dyer Conjecture** (BSC)

Let $E/\mathbb{Q}$ be an elliptic curve and $L(s,E)$ its L-function. Then the order of vanishing of $L(s,E)$ at $s = \frac{1}{2}$ is equal to the rank of the group of $\mathbb{Q}$-rational points on E.

Comments:

(D1) Again this qualifies as an excellent and perhaps somewhat unexpected conjecture at the time. It contains highly nontrivial local to global information. Recall that the L-function $L(s,E)$ is defined from local data while the rank of the group of rational points is one of the most interesting global geometric invariants of $E(\mathbb{Q})$.

(D2) The point $s = \frac{1}{2}$ is the only explicitly known point at which any $\Lambda(s, F)$ vanishes. Vanishing at $s = \frac{1}{2}$ could happen simply because of the sign of the functional equation (if $F = \tilde{F}$ and $\epsilon_F = -1$), but as in the case of $\Lambda(s, E)$, it could vanish to order greater than 1 for deeper arithmetical reasons.

As with the last conjecture there are substantial results towards the BSC for elliptic curves over $\mathbb{Q}$. The works of Coates-Wiles [CoW] for CM elliptic curves and Kolyvagin-Lugachev [KoL] and Gross-Zagier [GrZ] in general imply essentially that the Conjecture is true if the order of vanishing at $s = \frac{1}{2}$ is at most 1. It should be noted that the only general method to construct rational points on a given E is Heegner's construction [Hee] (we are not asking to find elliptic curves containing a given rational point). It is unclear what role these play when $L(s, E)$ vanishes to order ≥ 2.

3 Function Field Analogues

As was mentioned in section 2 the function field analogue of GRH is known. There is a lot to be learned from this algebro-geometric analogue and it has led to many insights for L-functions over number fields. As in the number field case the Riemann Hypothesis in the function field has striking implications. In particular it yields optimal bounds for exponential and character sums over finite fields and these are a basic tool in many of the results mentioned already as well as ones mentioned below. In fact, the special cases of the GRC that have been established make use of GRH in the function field. So the function field is an important part of our story and we review this analogue briefly.

The starting point is to replace the field K by a finite extension k of the field $\mathbb{F}_q(t)$, $\mathbb{F}_q$ being the field with q elements. We define, following Artin, the zeta function of k

$$\zeta_k(T) := \prod_v \left(1 - T^{\deg(\nu)}\right)^{-1}, \tag{48}$$

the product being over all the places (i.e. primes) of k and $\deg(v)$ is the corresponding local extension degree. The field k may be realized as the field of functions of a nonsingular projective curve C over $\mathbb{F}_q$. This allows one to give an alternate useful expression for $\zeta_k(T)$. If N_n is the number of points on C defined over $\mathbb{F}_{q^n}$, then

$$\zeta_k(T) = \zeta(T, C/\mathbb{F}_q) = \exp\left(\sum_{n=1}^{\infty} \frac{N_n T^n}{n}\right). \tag{49}$$

Using this and the Riemann-Roch theorem for the curve C over $\bar{\mathbb{F}}_q$ one can show the analogue of the analytic continuation and functional equation of $\zeta(s)$, for $\zeta_k(T)$. That is,

$$\zeta_k(T) = \frac{P(T, C/\mathbb{F}_q)}{(1-T)(1-qT)}, \tag{50}$$

where P is an integral polynomial of degree $2g$ with g the genus of C and P satisfies a functional equation relating its values at T to $1/qT$. The analogue of GRH is the statement that all the zeros of P be on the circle $|T| = 1/\sqrt{q}$. This was established by Weil [We1]. There are a number of ideas that go into his proof (he gave two quite different proofs). The numbers N_n can be realized as the number of points on $C(\bar{\mathbb{F}}_q)$ which are fixed by the n^{th} power of the Frobenius morphism (raising coordinates to the power q). This suggests the use of a Lefschetz trace formula to linearize this counting. To achieve this Weil passes to the Jacobian X of C. For ℓ prime to q and $\nu \geq 1$ the corresponding Frobenius endomorphism α acts on the ℓ^ν division points of X, giving rise to an ℓ-adic matrix realization of α. Its eigenvalues are shown to be the inverses of the zeros of $P(T)$. This gives an important spectral interpretation of the zeros. The proof that the zeros are on the circle $1/\sqrt{q}$ requires a further elaborate analysis of α in the endomorphism ring of X and in particular the use of the positivity of Rosati involutions.

The definition of the zeta functions $\zeta(T, V/\mathbb{F}_q)$ for smooth projective varieties V over $\mathbb{F}_q$ was given by Weil. He put forth conjectures about the rationality, functional equations and analogues of the Riemann Hypothesis for these zetas. The first was proven by Dwork. A different proof was given by Grothendieck who also established the other analytic properties (i.e. functional equations, location of poles) by using his ℓ-adic cohomology theory. In particular, Grothendieck gives a spectral interpretation of $\zeta(T, V/\mathbb{F}_q)$ in terms of the characteristic polynomial of the induced linear action of Frobenius on the cohomology groups of the variety.

The proof of the analogue of GRH is due to Deligne [D2]. An important methodological difference in his proof being that the zeros are not shown to have a given absolute value ("purity") in one step and with one variety V. For example, if V is a smooth hypersurface in $\mathbb{P}^{2n}$ then he places $\zeta(T, V)$ in a family V_t, $t \in U$ a parameter space. The arithmetic fundamental group $\pi_1(U)$ has representations via monodromy in the various cohomology groups $H^i(\bar{V}_0, \mathbb{Q}_\ell)$, where V_0 is a fixed base point (in this hypersurface example only the middle dimensional cohomology group contains nontrivial information). In this way one may realize $\zeta(T, V/\mathbb{F}_q)$ as a local factor of an L-function associated with the monodromy representation above. One can furthermore examine various tensor powers of this representation. Also these new L-functions have known analytic properties (or at least one can locate their poles using invariant theory for the representations of the monodromy groups). One is now very much in the position that one is in deriving the local GRC from the global analytic properties of the symmetric power L-functions (see section 2). In fact, similar positivity arguments with arbitrarily high dimensional representations of the monodromy groups yield in the limit that the zeros of $\zeta(T, V/\mathbb{F}_q)$ are all on the circle $|T| = q^{-n+1/2}$. So the family, its symmetry and positivity are the key ingredients in the known proof of the GRH for varieties over finite fields.

The solution by Deligne of these Weil Conjectures allowed him to solve the special cases of GRC mentioned in section 2. The reduction itself is deep and is due to Eichler [E] and Igusa [I] in the special case of weight 2 and Ihara [Ih] and Deligne [D1] in general.

We end this section by mentioning the function field analogue of automorphic forms F on GL_m. Replacing, as we did at the start of this section, K by k we may consider the space $L^2(GL_m(k)\backslash GL_m(\mathbb{A}_k))$ and its cuspidal subspace. In a recent paper Lafforgue [L] has completed the program started by Drinfeld of (amongst other things) establishing the GRC for these automorphic cusp forms. A key ingredient of course is Deligne's proof of the Weil Conjectures above. There are many other crucial ingredients such as the trace formula [A] and the converse theorem [CogP].

4 Dirichlet L-Functions $GL(1)/\mathbb{Q}$

The work of Linnik [Li1] marked the beginning of a series of developments which give in some sense GRH (for Dirichlet L-functions) on average. This is not just an exercise but is a powerful tool which produces results not covered by GRH. In many applications of GRH one has, say sums of sums over primes in different arithmetic progressions, and GRH would give approximations for each sum. Since one is averaging over different progressions it is just as useful in such situations to know that the approximation offered by GRH is correct on average. The many developments during the period 1950-1970 mentioned above are based on a penetrating study of the orthogonality of Dirichlet characters (to different moduli!) and culminated in the Bombieri-Vinogradov Theorem.

For $(a,q)=1$, let

$$\psi(x;q,a) = \sum_{\substack{p\le x\\ p\equiv a(q)}} \log p\ , \tag{51}$$

the sum being over primes. According to GRH

$$\psi(x;q,a) = \frac{x}{\varphi(q)} + O\big(x^{1/2}(\log x)^2\big)\ , \tag{52}$$

where $\varphi(q)$ is the number of residue classes (modulo q) prime to q (this equivalence is essentially due to Riemann).

The Bombieri-Vinogradov Theorem (in a slightly stronger form by Bombieri [B1]) asserts that for $A>0$ there is $B>0$ such that

$$\sum_{q\le Q} \max_{(a,q)=1} \left|\psi(x;q,a) - \frac{x}{\varphi(q)}\right| \ll \frac{x}{(\log x)^A}\ ,$$

where $Q = x^{1/2}/(\log x)^B$. So this comes close to (52) on average in this range. By the way (53) is closely related to statements giving nontrivial bounds for the number of zeros $\rho=\beta+i\gamma$ of $L(s,\chi)$ of conductor $q\le Q$ and with $\beta\ge\sigma$ $(\sigma>\frac12)$, $|\gamma|\le T$, known as Density Theorems.

More recent results [BFI] use much more sophisticated tools including bounds for exponential sums over finite fields as well as $GL_2/\mathbb{Q}$ spectral theory in the form connected with sums of Kloosterman sums (see section 7). Their results concern primes in progressions to moduli beyond $\sqrt{x}$ and cannot be derived from GRH. For example it is shown (among stronger, but more complicated results) that for any $a\ne 0$, $A>0$ there is $B>0$ such that

$$\sum_{\substack{(a,q)=1\\ q\le\sqrt{x}(\log x)^A}} \left|\psi(x;q,a) - \frac{\psi(x)}{\varphi(q)}\right| \ll \frac{x}{(\log x)^3}(\log\log x)^B\ . \tag{53}$$

We turn to Problem B of section 2 which concerns the size of $L(s,\chi)$ on the critical line. The first developments are much older. Weyl's method [Wey] of shifting the argument and repeated squaring in estimating sums $\sum_n e(\alpha f(n))$, where $e(z)=e^{2\pi iz}$ and f is a polynomial, led to the subconvexity estimate (the convexity exponent here is $1/4$)

$$\zeta\big(\tfrac12+it\big) \ll \big(|t|+1\big)^{\frac16} \tag{54}$$

for the Riemann zeta function (the same can be done this way for $L(s,\chi)$ in the t-aspect). There have been many improvements of the exponent $\frac16$, but our emphasis here is on subconvexity.

For the case of $L(s,\chi)$ (s fixed with real part equal to one half) in the conductor q of χ aspect, there is the result of Burgess [Bur]. It gives the subconvexity estimates (again the convexity exponent is $\frac{1}{4}$)

$$L(s,\chi) \underset{\epsilon}{\ll} q^{\frac{3}{16}+\epsilon} . \tag{55}$$

Burgess proceeds by estimating the sums

$$S = \sum_{N<n\leq N+H} \chi(n) \tag{56}$$

for N and H of certain sizes. He obtains nontrivial bounds by summing S and its shifts to large (even) powers which allows him to make use of bounds for complete character sums which in turn rely on Weil's GRH in the function field for curves of suitably large genus. Interestingly there are ranges in (56) where Burgess obtains nontrivial bounds and for which the GRH for $L(s,\chi)$ yields nothing nontrivial.

7

In the next section we discuss a recent improvement of (56) for χ quadratic.

5 Special Values

The question as to whether an L-function $L(s,F)$ vanishes at a special point on the critical line has arisen in various contexts and is apparently a fundamental one (note that such a question is not addressed by GRH). It arises in the problem of examining the instability of the elusive Maass cusp forms (see section 1). For this problem the L-functions in question are $L(s,\phi\otimes Q)$, ϕ a Maass form and Q a holomorphic cusp form of weight 4 (all this for $GL_2/\mathbb{Q}$). The special points being $s=\frac{1}{2}\pm ir$, where the Laplace eigenvalue of ϕ is $\frac{1}{4}+r^2$. The other special point that arises is $s=\frac{1}{2}$ for self-dual forms F (i.e. $F=\tilde{F}$). This point is the central symmetry point for the functional equation of $L(s,F)$. In the case that $L(s,F)$ is the L-function of an elliptic curve (or an abelian variety) then the vanishing at $s=\frac{1}{2}$ is related to rank of the group of rational points, this being presented in D) of section 2. Note that if F is self dual then $L(s,F)$ is real for s real and since $L(s,F)\to 1$ as $s\to\infty$, it follows that if we admit GRH then

$$L\left(\tfrac{1}{2},F\right) \geq 0 . \tag{57}$$

In the simplest case, that is F being a quadratic Dirichlet character (57) is not known (in fact one can show that if $L\left(\frac{1}{2},\chi\right)\geq 0$ for χ quadratic then one can eliminate in part the Landau-Siegel lacuna mentioned in section 2). So it is quite striking that for $PGL(2)/K$ cusp forms F (these are self-dual) one can show that

$$L\left(\tfrac{1}{2},F\otimes\chi\right) \geq 0 \tag{58}$$

for any quadratic ray class character χ. This final version is due to Guo [Gu], it completed a series of developments beginning with Waldspurger [W]. The theta function treatments of (58) (Guo proceeds differently using the relative trace formula) proceed by expressing $L\left(\frac{1}{2}, F \otimes \chi\right)$ as a sum of squares of the χ-th Fourier coefficient of a form of half-integral weight which corresponds to F as in Shimura [Shi1], [Shi3]. We will exploit this in the next section.

An application of (59) for the case $K = \mathbb{Q}$ and ϕ a Maass cusp form was given recently in [ConrI]. By incorporating (for χ quadratic) $\int_{-\infty}^{\infty} |L(\frac{1}{2} + it, \chi)|^6 \, e^{-t^2} dt$ as part of a family involving $L^3\left(\frac{1}{2}, \phi \otimes \chi\right)$ it is shown that for s fixed with $\Re(s) = \frac{1}{2}$

$$L(s, \chi) \ll q^{\frac{1}{6}+\epsilon} . \tag{59}$$

This gives the first improvement over (58) and is another pleasing example of the use of GL_2 theory to understand GL_1 L-functions. It highlights our point of view that L- functions be considered as a whole and especially in families. We add that the proof of the above appeals to bounds for exponential sums in two variables over finite fields and in particular to Deligne's estimates which follow from the general GRH for varieties over finite fields.

Another case where (58) has been established is the following [HK]. Let F_1, F_2, F_3 be three forms on $PGL(2)/K$ and $L(s, F_1 \otimes F_2 \otimes F_3)$ the L-function (28) which has degree eight. Then

$$L\left(\tfrac{1}{2}, F_1 \otimes F_2 \otimes F_3\right) \geq 0 . \tag{60}$$

As in the previous example the proof of (60) involves expressing the special value as a sum of squares of "periods" of $F_1F_2F_3$. For analytic applications one needs an entirely explicit relation between these special values and periods. In his thesis [Wa] has proved such an explicit relation for forms over $\mathbb{Q}$. For example, for Maass forms of full level (i.e. for forms on $SL(2, \mathbb{Z})$) as in (6) of section 1 he shows that

$$\frac{\Lambda\left(\frac{1}{2}, \phi_1 \otimes \phi_2 \otimes \phi_3\right)}{\prod_{j=1}^{3} \Lambda(1, \mathrm{sym}^2 \phi_j)} = \frac{\pi^4}{216} \left| \int_{SL(2,\mathbb{Z})\backslash\mathbb{H}} \phi_1(z)\phi_2(z)\phi_3(z) \, \frac{dxdy}{y^2} \right|^2 , \tag{61}$$

where Λ is the completed L-function

$$\Lambda(s, \phi_1 \otimes \pi_2 \otimes \phi_3) = \pi^{-4s} \prod_{\epsilon_j = \pm 1} \Gamma\left(\frac{s + \epsilon_1 r_1 + \epsilon_2 r_2 + \epsilon_3 r_3}{2}\right) L(s, \phi_1 \otimes \phi_2 \otimes \phi_3)$$

and ϕ_1, ϕ_2, ϕ_3 are normalized to have L^2-norm equal to one on $SL(2,\mathbb{Z})\backslash\mathbb{H}$. We will exploit this beautiful formula in the next section.

Of special interest in applications is to know how often a family of L-functions vanish at a special point. The technique of mollification (championed by Selberg in his proof that a positive proportion of the zeros of $\zeta(s)$ lie on $\Re(s) = \frac{1}{2}$) has been successfully developed in the context of special values of GL_2 forms (at least over $\mathbb{Q}$) in [IwS] and [KowMV2]. We mention some results in this direction. Let N be squarefree and fix $k \geq 2$. Let $H_k^*(N)$ denote the set of holomorphic newforms F of weight k for $\Gamma_0(N)$, that is on the modular curve $X_0(N)$. The sign ϵ_F of the functional equation for $L(s,F)$ is ± 1. When N is large (which is our interest here) roughly one half of the forms have each sign, the total number being $|H_k^*(N)| \sim \frac{k-1}{12}\varphi(N)$. If $\epsilon_F = -1$ then $L\left(\frac{1}{2}, F\right) = 0$ and we are interested in $L'\left(\frac{1}{2}, F\right)$. It is shown [IwS], [KowMV2] that

$$\varliminf_{N\to\infty} \frac{\#\{F \in H_k^*(N); \epsilon_F = 1, L\left(\frac{1}{2}, F\right) \geq (\log N)^{-2}\}}{\#\{F \in H_k^*(N); \epsilon_f = 1\}} \geq \tfrac{1}{2}, \quad (62)$$

$$\varliminf_{N\to\infty} \frac{\#\{F \in H_k^*(N); \epsilon_F = -1, L'\left(\frac{1}{2}, F\right) \neq 0\}}{\#\{F \in H_k^*(N); \epsilon_F = -1\}} \geq \tfrac{7}{8}, \quad (63)$$

$$\varlimsup_{N\to\infty} \frac{1}{|H_k^*(N)|} \sum_{F\in H_k^*(N)} \operatorname{ord}_{s=\frac{1}{2}} L(s,F) \leq 1.2. \quad (64)$$

We expect that the constants $\frac{1}{2}$ and $\frac{7}{8}$ in (62) and (63) can be replaced by 1 while 1.2 in (64) can be replaced by $\frac{1}{2}$. It is tantalizing that an improvement in (62) of the $\frac{1}{2}$ to any $c > \frac{1}{2}$, would resolve the Landau-Siegel lacuna (section 2). The proof of this implication [IwS] exploits the positivity (58).

Next (62) and (63) together with the results [KoL], [GrZ] towards BSC mentioned in D of section 2 imply results on the ranks of the Mordell-Weil groups of the Jacobian varieties $J_0(N) = JAC(X_0(N))/\mathbb{Q}$. Precisely (62) yields a quotient ("winding quotient" [M]) $M_0(N)$ of $J_0(N)$ over $\mathbb{Q}$, which has only finitely many rational points and has dimension which is asymptotically at least $\frac{1}{4}\dim J_0(N)$. Moreover, (63) implies that for large N the rank of $J_0(N)$ is asymptotically at least $\frac{7}{16}\dim J_0(N)$. Finally, (64) together with BSC imply that rank $J_0(N) \leq 1.2\dim J_0(N)$, for N large.

The application of nonvanishing to spectral deformation theory also concerns $X_0(N) = \Gamma_0(N)\backslash\mathbb{H}$ (for N fixed). Let ϕ_j (with eigenvalue λ_j) be an orthonormal basis of Maass Hecke cusp forms and let Q be a fixed holomorphic cusp form of weight $k \geq 1$. Recently Luo [Lu] using the mollification methods as above has shown that

$$\varliminf_{\lambda \to \infty} \frac{\#\{\lambda_j \le \lambda; L\left(\frac{1}{2} + ir_j, Q \otimes \phi_j\right) \neq 0\}}{\#\{\lambda_j \le \lambda\}} > 0 . \tag{65}$$

This has striking applications to the question of nonexistence of Maass cusp forms for the general quotient $\Gamma\backslash\mathbb{H}$, Γ in the deformation (Teichmuller) space of $\Gamma_0(N)$, see [PhS] and [Wo].

6 Subconvexity and Equidistribution

Up to now the discussion has centered around L-functions only. In this section we give two examples of applications to problems which at first sight appear to have nothing to do with L-functions. First we describe some results on the subconvexity problem B of section 2 for Euler products of degree at least two. In these cases the coefficients of the L-functions are arithmetical and inexplicit so that the methods of Weyl and Burgess don't apply. Instead sophisticated new methods are needed (see section 7).

For $L(s, F)$ with F a cusp form on $GL_2/\mathbb{Q}$, subconvexity has been established in all the parameter aspects (in s aspect in [Goo] and [Me] while in the other parameters in the series of papers by Duke Friedlander and Iwaniec). We concentrate on the twisting by Dirichlet characters. For a fixed F on $GL(2)/\mathbb{Q}$ a cuspidal eigenform (i.e. a holomorphic or Maass form on $\Gamma_0(N)\backslash\mathbb{H}$) and χ a (primitive) Dirichlet character, the following subconvexity estimate (s is fixed with $\Re(s) = \frac{1}{2}$) was established in [DuFI]:

$$L(s, F \otimes \chi) \ll q^{\frac{5}{12}+\epsilon} . \tag{66}$$

Here q is the conductor of χ and the convexity bound is $q^{1/2}$.

The methods used to deal with F on $GL(2)/\mathbb{Q}$ run into a number of difficulties (not the least of which are the units) for number fields. Recently the authors of [CogPS] have resolved these difficulties. Let K be a totally real extension of $\mathbb{Q}$. Fix a holomorphic Hilbert modular cusp form of even integral weight (i.e. a form on GL_2/K). Let χ range over primitive ray class characters of conductor $\mathcal{Q}$ (we have in mind $N(\mathcal{Q}) \to \infty$). Then for s fixed with $\Re(s) = \frac{1}{2}$

$$L(s, F \otimes \chi) \ll N(\mathcal{Q})^{\frac{49}{100}+\epsilon} . \tag{67}$$

Again the conductor of $F \otimes \chi$ is $N(\mathcal{Q})^2$ so that the convexity bound for (67) is $N(\mathcal{Q})^{1/2}$.

Some progress has also been made for Euler products (over $\mathbb{Q}$) of higher degree. Fix a holomorphic or Maass cusp form G for $\Gamma_0(N)\backslash\mathbb{H}$ (so N is fixed) and let F vary over the holomorphic newforms for $\Gamma_0(N)$ of weight k.

Then for s fixed with $\Re(s) = \frac{1}{2}$, the Rankin-Selberg L-functions $L(s, F \otimes G)$ satisfy the subconvexity estimate [S2] (in the k-aspect)

$$L(s, F \otimes G) \ll k^{\frac{27}{28}+\epsilon} \tag{68}$$

(here the "analytic" conductor is k^4 so that the convexity bound is k).

Also, for Rankin-Selberg L-functions, but in the level aspect, [KowMV1] have established a subconvexity estimate. Precisely, fix G and let F vary over holomorphic newforms of the same weight as G, but of level $N \to \infty$. Then for s fixed with $\Re(s) = \frac{1}{2}$,

$$L(s, F \otimes G) \ll N^{\frac{1}{2}-\frac{1}{96}+\epsilon} \tag{69}$$

(the convexity bound being $N^{\frac{1}{2}}$).

We turn to the applications. The first is to Hilbert's 11-th problem: which integers are integrally represented by a given quadratic form over a number field? The case of binary quadratic forms is equivalent to the theory of relative quadratic extensions and their class groups and Hilbert class fields. For forms in four or more variables the situation is quite different and has been understood for some time. The case of three variables has remained open and we describe below the essential part of its resolution.

Fix the number field K. The problem of which integers ν in K are represented by the genus of a given integral quadratic form $f(x_1, x_2, \ldots x_n)$ is answered completely by Siegel's mass formula [Si4] (which gives the number of solutions in terms of local data, via the product of local masses). So if there is one class in the genus of f the formula resolves the representation problem for f. If $n \geq 3$ and f is indefinite at an archimedean place v of K then Kneser's [Kn] results on the class numbers and weights of the spinor genus of f show that we are more or less in the one class in the genus situation. So we restrict to the difficult case when $f(x_1, \ldots, x_n)$ is definite over a totally real field K. For four or more variables one can proceed either by using analytic methods of Hilbert modular forms and in particular the bounds towards GRC for GL_2/K (see (39) for weight two holomorphic cusp forms) or by using algebraic methods ([HsKK], [C]), to prove the following: There is C_f (depending on f effectively) such that if $\nu \in \mathcal{O}_K$ is (totally) positive and $N(\nu) \geq C_f$, then ν is primitively represented by f iff it is primitively represented locally at every completion v (the local conditions are satisfied for all but finitely many primes and are easily checked).

For f a form in three variables the problem is much more difficult and is resolved (at least for squarefree ν) by the estimates (66) and (67). The connection is as follows: using the relation between the special value $L\left(\frac{1}{2}, F \otimes \chi\right)$, $\chi^2 = 1$ and the "χ-th" Fourier coefficient of half-integral cusp forms ([W], [Shi3]) mentioned in the last section, one finds that the bound (66) for $\mathbb{Q}$ and (67) in general, give nontrivial bounds for the *squarefree* Fourier coefficients of half-integral weight holomorphic cusp forms. Here and in other such problems, the convexity bound for the L-function corresponds exactly to the "trivial" bound for the Fourier coefficients. Moreover, the Lindelöff Hypothesis in the quadratic twisting χ aspect, for $L\left(\frac{1}{2}, F \otimes \chi\right)$, is equivalent (or determines) the *half-integral weight* GRC. Put another way, a nontrivial bound for the squarefree Fourier coefficients of a half-integral weight cusp form is equivalent to a subconvexity bound for $L\left(\frac{1}{2}, F \otimes \chi\right)$ while GRH for $L(s, F \otimes \chi)$ (via Lindelöff) implies the optimal bound for these coefficients. In the case $K = \mathbb{Q}$ a nontrivial bound for the Fourier coefficients of such forms was derived earlier in [Iw1] by a different method. For the case of Hilbert modular forms the passage via (67) gives the first bounds towards the Ramanujan Conjectures of half-integral weight.

We return to the form $f(x_1, x_2, x_3)$. Its theta function $\theta_f(z)$ is a Hilbert modular form of weight $\frac{3}{2}$ whose coefficients give the number of representations by f. Write $\theta = E + C$ where E is an Eisenstein series and C a cusp form. The ν-th coefficient of the Eisenstein series (a linear combination of standard Eisenstein series) depends only on the genus of f and is a product of local masses. It can be estimated from below by $C_\epsilon N(\nu)^{1/2-\epsilon}$ (when ν is represented locally) where $\epsilon > 0$ and C_ϵ an ineffective positive constant depending on ϵ. It is ineffective since the lower bound appeals to Siegel's ineffective lower bound for $L(1, \chi)$ (see (30)). Now the bound (67) leads as above to the ν-th coefficient of C being $O(N(\nu)^{\frac{49}{200}})$. Thus we conclude: if $N(\nu)$ is sufficiently large and squarefree then ν is represented integrally iff it is represented locally. This yields a solution (albeit ineffective and for squarefree ν) of the representation problem for definite ternary forms. Using the results of Schulze-Pillot [Sc] one can extend these results to all ν except perhaps for an explicit finite set of square classes ($\nu = \nu_0 t^2, t > 0$) along which the local to global principle can fail.

Of special interest is the long studied problem of sums of squares in a number field. Over $\mathbb{Q}$ as is well-known all positive numbers ν are sums of four squares (Lagrange) and such a ν is a sum of three squares iff $\nu \neq 4^a(8b+7)$ (Legendre), that is iff there are no local obstructions. That the answers for these are so neat is a consequence of $x_1^2 + x_2^2 + x_3^2$ and $x_1^2 + x_2^2 + x_3^2 + x_4^2$ having one class in their genus. This happens for very few totally real fields. In fact Siegel [Si3] shows that $\mathbb{Q}(\sqrt{5})$ is the only field for which *every* totally positive number is a sum of three squares. In general Siegel [Si1] showed that every sufficiently large (in norm) totally positive ν is a sum of five squares and the result mentioned above settles four squares similarly. For three squares (67) implies that there is an ineffective C_K depending on K such that if $N(\nu) \geq C_K$ and ν is squarefree and totally positive, then ν is a sum of three squares iff it is so locally (the local condition only involves the primes dividing 2 and for many fields K there are no such local obstructions). Note these results are ones of equidistribution. Indeed for $N(\nu)$ large and ν satisfying the local solvability conditions, ν is represented by the genus of f in roughly $N(\nu)^{1/2}$ ways. The subconvexity estimate ensures that each class in the genus represents ν roughly equally often. We mention that Linnik [Li2] gave an interesting ergodic theoretic approach to equidistribution problems associated to ternary quadratic forms which yield some partial results.

We end this part of the discussion with some general comments about integer solutions to Diophantine equations. The problem of establishing the existence of any or many such solutions for equations for which solutions are expected, has proven formidable. Success has been limited to varieties which are homogeneous (such as the case of quadrics which were discussed above) for an action of an algebraic group, or to varieties defined by many variables compared to the degree and number of equations. For the latter the circle method of Hardy and Littlewood can be applied. An example of the last is due to Heath-Brown [He] who showed that any nonsingular cubic form in 10-variables over $\mathbb{Q}$ has infinitely many rational (projective) points (i.e. for $f(x) = 0$). New methods for exhibiting rational points on varieties would be very welcome.

The second equidistribution problem comes from "Quantum Chaos". One of the central problems in this subject concerns the behavior of individual eigenstates of the quantization of classically chaotic systems in the semi-classical limit. This problem cannot at present be addressed with the techniques of analysis or partial differential equations. To gain insight we therefore specialize to systems of classical mechanics defined by the geodesic motion on a hyperbolic manifold (compact or finite volume). These are well known examples of chaotic Hamiltonian dynamics. We even specialize to such manifolds of arithmetic type. For these it turns out that the questions that we have been discussing about general L-functions lie at the heart of the problem. Consider the case of a hyperbolic such surface $X = \Gamma\backslash\mathbb{H}$. A quantization of the geodesic motion on the cotangent space $T^*(X)$, is the Laplacian Δ. Let $dv(x)$ denote the Riemannian measure on X. Perhaps the most fundamental problem concerns the behavior as $\lambda \to \infty$ of the probability measures $\mu_\phi := |\phi(x)|^2 dv(x)$ on X, where $\Delta\phi + \lambda\phi = 0$; $\int_X |\phi(x)|^2\, dv(x) = 1$. These measures have the well-known interpretation of being the probability distribution in configuration space of a particle in eigenstate ϕ. These ϕ's are the familiar Maass forms, especially if we restrict to $X_0(N) = \Gamma_0(N)\backslash\mathbb{H}$, N fixed (one can similarly analyze the compact surfaces arising as quotients of $\mathbb{H}$ by units in quaternion groups). We consider the question of the behavior of these measures μ_ϕ for Maass forms ϕ which are also eigenforms for the Hecke operators. Since the multiplicity of cusp forms with eigenvalue λ is expected to be very small, the latter assumption is probably not necessary, but we will certainly exploit it. Once we are assuming that ϕ is a Hecke eigenform we can also allow holomorphic eigenforms as well. We formulate the problem precisely for these modular forms: Fix N and $X_0(N) = \Gamma_0(N)\backslash\mathbb{H}$. Each of the spaces of holomorphic newforms of weight k with given Hecke eigenvalues, and Maass newforms with given Hecke eigenvalues, are one dimensional. So there are unique normalizations so that the following are probability measures on $X_0(N)$:

$$\begin{aligned} \mu_f &:= y^k \left|f(z)\right|^2 dv(z) \\ \mu_\phi &:= \left|\phi(z)\right|^2 dv(z)\ , \end{aligned} \tag{70}$$

where $dv(z) = y^{-2}dxdy$. We have the following Equidistribution of Mass Conjecture [RudS2]:

$$\lim_{k\longrightarrow\infty} \mu_f = d\tilde{v} \tag{71}$$

$$\lim_{\lambda\longrightarrow\infty} \mu_\phi = d\tilde{v}\ , \tag{72}$$

where $d\tilde{v} = dv/\mathrm{Vol}\,(X_0(N))$.

Comments:

(1) This conjecture looks reasonable, but at the time it was made it went against certain beliefs (i.e. that eigenstates might concentrate on periodic geodesics for example). The conjecture is made more generally for compact manifolds of negative curvature, yet the only theoretical evidence is in the case of arithmetic manifolds (there is some numerical evidence as well [Hej], [AuS]).

(2) In the holomorphic case the condition that f be a Hecke eigenform is essential. For example the masses μ_{Δ^k} of the holomorphic forms $(\Delta(z))^k$ of weight $12k$ on $X_0(1)$ certainly don't become equidistributed.

(3) The Conjecture (71), if true, has the pleasant corollary that the zeros in $X_0(N)$ of such an $f(z)$ (there are about k of them) become equidistributed with respect to $d\tilde{v}$ as $k \to \infty$ [Rud].

The connection between Conjectures (71), (72) and L-functions is (61) and its generalizations. GRH for the triple product L-function $L(s, F \otimes F \otimes G)$ and even subconvexity in the k or λ aspects (with $s = \frac{1}{2}$) already imply these conjectures! To see this, consider the case of $GL_2(\mathbb{Z})\backslash\mathbb{H}$. The equidistribution (72) in this case is equivalent to

$$\mu_{\phi_\lambda}(\phi_0) = \int_{SL_2(\mathbb{Z})\backslash\mathbb{H}} \phi_0(z)|\phi_\lambda(z)|^2 y^{-2}\, dxdy \to 0 \tag{73}$$

as $\lambda \to \infty$, for any fixed Maass cusp form ϕ_0 (one needs also to consider the continuous spectrum, that is the unitary Eisenstein series in place of ϕ_0, but these are slightly easier to handle so we ignore them here). Now according to (61) the right hand side of (73) is up to gamma factors equal to $L(\frac{1}{2}, \phi_\lambda \otimes \phi_\lambda \otimes \phi_0)/L^2(1, \mathrm{sym}^2\phi_\lambda)L(1, \mathrm{sym}^2\phi_0)$. There is no problem dealing with $L(1, \mathrm{sym}^2\phi_\lambda)$ since it is bounded below by $\lambda^{-\epsilon}$ and above by λ^ϵ for any $\epsilon > 0$, and effectively so [Iw2], [HoL]. A simple analysis with Stirling formula then shows that a bound of $O(\lambda^{-\delta})$ for the right hand side in (73) is equivalent to the subconvexity estimate in λ;

$$L\left(\tfrac{1}{2}, \phi_\lambda \otimes \phi_\lambda \otimes \phi_0\right) \ll \lambda^{\frac{1}{2}-\delta_0}\,. \tag{74}$$

Unfortunately we have not been able to establish (74) in general. The L-function $L(s, \phi_\lambda \otimes \phi_\lambda \otimes \phi_0)$ factors into $L(s, \mathrm{sym}^2\phi_\lambda \otimes \phi_0)\; L(s, \phi_0)$. So the key case is subconvexity for $L\left(\frac{1}{2}, \mathrm{sym}^2\phi_\lambda \otimes \phi_0\right)$ which is an Euler product of degree six. For the special case of "CM forms" ϕ_λ or f on $\Gamma_0(N)\backslash\mathbb{H}$ (see (9) for example) this Euler product factors further into a Rankin Selberg L-function of degree four times $L(s, \phi_0)$ which is of degree two. So in this case the subconvexity estimate (68) (and its λ analogue) gives a proof of (71) and (72). That is (71) and (72) are true for CM forms.

These two applications show, of course, the power of GRH, however they also show that in certain problems a complete resolution can be achieved by finessing GRH and establishing the more approachable subconvexity estimate.

7 $GL(2)$ Tools

We give some flavor of some of the modern techniques that have been successful in studying L-functions by indicating how subconvexity estimates are proven. Suppose for example that we want to estimate $L\left(\frac{1}{2}, F\right)$, where

F is a self-dual cusp form on $GL_m(\mathbb{A}_K)$. Using the series representation (3) together with standard arguments involving contour shifts and the functional equations, we obtain

$$L\left(\tfrac{1}{2}, F\right) = 2\sum_{a\neq 0} \frac{c_F(a)}{\sqrt{Na}}\, W\left(\frac{Na}{X}\right), \tag{75}$$

where $W(t)$ is a smooth function which is essentially independent of F and is rapidly decreasing as $t \to \infty$ and $X = \sqrt{C(F)}$, $C(F)$ being the analytic conductor (31) (here $C(F)$ denotes $C(F,0)$ from (31)). The coefficients $c_F(a)$ are known in some cases to satisfy the GRC so that

$$c_F(a) \ll_\epsilon (Na)^\epsilon . \tag{76}$$

In any case for F on $GL_m(\mathbb{A}_k)$ we have (76) on average ([Iw2], [Mo])

$$\sum_{Na\leq Y} |c_F(a)|^2 \ll_\epsilon Y\big(C(F)\big)^\epsilon . \tag{77}$$

From (75) and (77) the "trivial" convexity bound

$$L\left(\tfrac{1}{2}, F\right) \ll_\epsilon \big(C(F)\big)^{\frac{1}{4}+\epsilon} \tag{78}$$

follows. To go beyond (78) one needs to exhibit some cancellation in the sum (75). We can no longer appeal to any functional equations since these have already been exploited in deriving (75) (that is if we "dualize" using the functional equation we arrive back at a similar sum). Also to directly estimate (75) is problematic since one knows very little about $c_F(a)$. We proceed according to the theme of this account and embed F in a family $\mathcal{F}$ (sometimes even fake families!). Finding suitable families is part of the problem. The idea is to consider averages

$$S(\mathcal{F}) = \sum_{F\in\mathcal{F}} \left|L(\tfrac{1}{2}, F)\right|^2 . \tag{79}$$

The conductors $C(F), F \in \mathcal{F}$ are all assumed to be the same (or nearly the same) size. In some cases one might take higher powers of L in (79) (the choice here of high moments rather than something like high tensor powers as in (43) and in section 3 is no doubt a poor one). Now GRH (via Lindelöff) asserts that $L(\frac{1}{2}, F) \ll_\epsilon \big(C(F)\big)^\epsilon$ so we can expect that

$$S(\mathcal{F}) \ll_\epsilon |\mathcal{F}|\big(C(F_0)\big)^\epsilon . \tag{80}$$

Here F_0 is our particular $F \in \mathcal{F}$ which we seek to bound. Using orthogonality and completeness of the family $\mathcal{F}$, (80) can often be established. By positivity (an apparently precious tool) we have from (80).

$$L(\tfrac{1}{2}, F_0) \ll_\epsilon |\mathcal{F}|^{1/2}\big(C(F_0)\big)^\epsilon . \tag{81}$$

So if the family $\mathcal{F}$ is sufficiently small (precisely $|\mathcal{F}| \ll (C(F))^{1/2-\delta}$) and at the same time rich enough to establish (80), then (81) will yield a subconvex bound. In practice when this method succeeds one finds that one can establish (80) with $|\mathcal{F}| \asymp (C(F))^{1/2}$ in a relatively straight-forward analysis involving summing over $\mathcal{F}$ and analyzing only the "diagonal" contribution. This however simply recovers the convexity bound (78) and the heart of the problem is to decrease somewhat the size of $|\mathcal{F}|$. This is done at the expense of off-diagonal terms now appearing in the analysis. Cancellations in these new sums has to be gotten from some new input. In some problems the decrease in size of $|\mathcal{F}|$ can be achieved analytically (e.g. in the case of (68), $F \otimes G$ with G fixed on $GL_2/\mathbb{Q}$, F varying over $\mathcal{F}_K$ the holomorphic forms of fixed level and weight $\frac{K}{2} \leq k \leq 2K$. Shortening here can be done by restricting $K - H \leq k \leq K + H$ with $H = K^{1-\delta}, \delta > 0$). In many interesting examples (e.g. (66) and (67)) shortening cannot be achieved by such a device. One appeals to a technique known as "amplification" which arithmetically shortens $\mathcal{F}$ by introducing weights. The method was introduced in [FI] in connection with estimating $L(s, \chi)$, however its true power has been shown in the more general setting of $L(s, F)$s. Roughly the idea is as follows: Let M be a small parameter ($M = X^{\delta}, \delta$ small) and let $\alpha(b)$ with $Nb \leq M$, be complex numbers of modulus at most 1. Consider the amplified sums

$$A = \sum_{F \in \mathcal{F}} \left| \sum_{Nb \leq M} \alpha(b)\, c_F(b) \right|^2 \left| L\left(\tfrac{1}{2}, F\right) \right|^2 . \tag{82}$$

This time we shoot for the expected bound

$$A \ll_{\epsilon} M |\mathcal{F}| X^{\epsilon} . \tag{83}$$

In establishing (83) one faces off-diagonal terms and if these can be successfully estimated then choosing $\alpha(b) = \overline{c_{F_0}(b)}$, i.e. amplifying F_0, we get

$$\left| L\left(\tfrac{1}{2}, F_0\right) \right|^2 \left| \sum_{Nb \leq M} |c_{F_0}(b)|^2 \right|^2 \ll_{\epsilon} |\mathcal{F}| M X^{\epsilon} . \tag{84}$$

This implies a subconvex bound for $L\left(\frac{1}{2}, F_0\right)$.

So the key features are the family and dealing with the off-diagonal sums. For example for (66) the family used is $L(s, F \otimes \chi)$ with χ a Dirichlet character of conductor q. For (67) one cannot use ray class characters of conductor $\mathcal{Q}$ since there may be very few of these (such characters have to be trivial on the units). One proceeds with nonnegative expressions which make sense for all numerical characters of $(\mathcal{O}_K/\mathcal{Q})^*$, but which have no meaning in terms of L-functions ("fake families"). In both (66) and (67) amplification is used. The key off-diagonal sums that need to be treated are of type

$$\sum_{\nu\alpha-\mu\beta=h} c_F(\alpha)\, c_F(\beta)\, W\left(\frac{N(\alpha)}{X}\right) W\left(\frac{N(\beta)}{X}\right) G(\alpha,\beta)\,, \tag{85}$$

where ν and μ are fixed small integers in K, $h \neq 0$ and G is a smooth function depending on the arguments of α and β in the embeddings of K into $\mathbb{R}$. Over $\mathbb{Q}$ a reasonably elementary treatment of these sums is given in [DuFI] (it appeals among other things to Weil's bounds on Kloosteman sums - see below). In general, one uses the full Maass form spectral theory for $GL_2(\mathbb{A}_K)$ and a suitable theory of Poincaré series. Crucial ingredients are the GRC bounds (41) (42) as well as the spectral analysis method in [S1].

For the case of $L(s, F\otimes G)$ and the estimate (69), where F varies over holomorphic cusp forms of a fixed weight for $\Gamma_0(N)$ and $N\to\infty$ (all over $\mathbb{Q}$), the averaging is carried out by means of the Petersson Formula [P]. Let $B_k(N)$ be an orthogonal basis for $S_k(N)$ (the space of holomorphic cusp forms of weight k for $\Gamma_0(N)$). Normalize the Fourier coefficients of $a_F(n)$ for $F\in S_k(N)$ by setting

$$\psi_F(n) = \left(\frac{\Gamma(k-1)}{(4\pi n)^{k-1}}\right)^{1/2} \frac{a_F(n)}{\langle f,f\rangle^{1/2}}\,. \tag{86}$$

Then the formula reads that for $m,n\geq 1$

$$\sum_{F\in B_k(N)} \overline{\psi_F(n)}\,\psi_F(m) = \delta(m,n) + 2\pi i^k \sum_{c\equiv 0(N)} \frac{S(m,n,c)}{c} J_{k-1}\left(\frac{4\pi\sqrt{mn}}{c}\right). \tag{87}$$

Here $\delta(m,n)$ is 0 if $m\neq n$ and is 1 if $m=n$, $J_k(x)$ is the Bessel function and $S(m,n,c)$ is the Kloosterman sum

$$S(m,n,c) = \sum_{\substack{x \bmod c\\ x\bar{x}\equiv 1(c)}} e\Big(\frac{mx+n\bar{x}}{c}\Big)\,. \tag{88}$$

Thus (87) converts averages of this family to sums of exponential sums over finite fields and allows one to analyze averages over the family of $L(s,F)$, for $F\in S_k(N)$. The Petersson formula (87) and its generalizations due to Kuznetsov [Ku] are important tools in the subject. They lie at the bottom of many of the applications of the $GL_2/\mathbb{Q}$ analytic theory. We mention one other such application, that is the Density Theorem for exceptional eigenvalues. Recall the Selberg Conjecture (36) which is a special case of GRC. In [Iw2] the following Density Theorem is proven; for any r with $0 \le r \le \frac{1}{2}$,

$$\#\Big\{\lambda < \tfrac{1}{4} - r^2; \lambda \text{ an eigenvalue of } \Delta \text{ on } L_0^2(\Gamma_0(N)\backslash\mathbb{H})\Big\} \\ \ll_\epsilon \Big[\mathrm{Vol}\big(\Gamma_0(N)\backslash\mathbb{H}\big)\Big]^{1-4r+\epsilon} . \quad (89)$$

The proof uses the generalizations of (87) as well as Weil's bound [We2] for Kloosterman sums

$$\big|S(m,n,p)\big| \le 2\sqrt{p} \quad (90)$$

if p does not divide m or n ((90) is a consequence of the function field RH for curves). A point to note about (89) is the exponent $1-4r$, which in particular implies Selberg's well-known bound

$$\lambda_1\big(\Gamma_0(N)\backslash\mathbb{H}\big) \ge \tfrac{3}{16} . \quad (91)$$

An exponent of $1-2r$ in (89) is easily deduced from the Selberg Trace Formula. Just as with the Density theorems for $GL(1)/\mathbb{Q}$ mentioned in section 4, (89) can sometimes be used as a substitute for (36) in applications.

Formula (87) and its generalizations are also useful when applied in the other direction, that is to capture cancellations in sums of Kloosterman sums (the Linnik-Selberg Conjecture [Li3], [Se]). Kuznetsov [Ku] used his formula and the fact that $\lambda_1(SL(2,\mathbb{Z})\backslash\mathbb{H}) \ge \frac{1}{4}$, to show that for m,n fixed

$$\sum_{c\le X} \frac{S(m,n,c)}{\sqrt{c}} \ll_\epsilon X^{\frac{2}{3}+\epsilon} . \quad (92)$$

Note that Weil's bound (which is sharp in c for m,n fixed [Mi]) gives $O_\epsilon(X^{1+\epsilon})$ for the sum in (92). So indeed (92) asserts cancellations in the sums of these sums. The development (42) in as much as it goes beyond (91) (i.e. $\lambda_1 \ge \frac{21}{100}$), shows that there is cancellation in the sums (92) for c in an arithmetic progression. Fix m,n,N and a, then

$$\sum_{\substack{c\le X\\ c\equiv a(q)}} \frac{S(m,n,c)}{\sqrt{c}} \ll_\epsilon X^{\frac{9}{10}+\epsilon} . \quad (93)$$

This concludes our brief discussion of GL_2 tools. Also fundamental is the Selberg Trace Formula which we have mentioned a few times in passing. So at least over $\mathbb{Q}$ and for GL_2, the analytic theory can be considered to be in quite good shape, much like $GL_1/\mathbb{Q}$ (see section 4) was at the end of the 70's.

The trace formula has been successfully extended to GL_n as well as other groups by Arthur [A]. His form is very suitable for comparisons of the geometric sides of the trace formula for different groups. This implies on the spectral side some striking conjectured "liftings" of automorphic forms between various groups. The analytic type of spectral and trace formula that have been developed for GL_n with applications such as those of this section in mind, have met only with mild success. For now this can be viewed as a challenging new direction which might provide important new information on the basic problems. For example we expect that such developments are needed to resolve the basic problem B in general.

8 Symmetry and Attacks on GRH

In the previous sections we described progress made not by climbing the summit (GRH), but by going around it. In this section we discuss some structural phenomenon and insights that might play a role in the accent.

As we have mentioned a number of times, families of L-functions play a central role even (or especially) when examining the deeper aspects of a *given* L-function. One might ask whether something like a monodromy group of a family of L-functions (section 3) exists in the number field setting. One way to detect such symmetry groups for families is to look at the local distribution of zeros of L-functions. For a fixed L-function $L(s,F)$, F cuspidal on $GL_m(\mathbb{A}_\mathbb{Q})$ (note here we demand over $\mathbb{Q}$ so that these L-functions do not factor further) one can examine the high zeros $\rho_F = \frac{1}{2} + i\gamma_F$ (for this part of the discussion we assume GRH). One can show that

$$N_F(T) := \#\{0 \le \gamma_F \le T\} \sim \frac{m \log T}{2\pi} T \,. \tag{94}$$

Hence, in studying the local distribution of spacings between the zeros one considers the unfolded numbers $\frac{m \log \gamma_F}{2\pi}\ \gamma_F$, whose mean spacing is 1. Remarkably these follow the local scaled spacing laws for eigenvalues of large unitary matrices, that is the CUE (Circular Unitary Ensemble) laws from random matrix theory (at least in leading order asymptotics). This was proven analytically in restricted ranges for the distribution of pairs of zeros ("pair correlation") in [Mon] and for higher correlations in [RudS1]. Moreover, extensive numerical experiments [O], [Rum], [Ru] confirm this phenomenon for various $GL_1/\mathbb{Q}$ and $GL_2/\mathbb{Q}$ L-functions. In [KS] an analogue of this phenomenon about local spacings of zeros is proven for the function field zeta functions of section 3. Moreover, the source of this universal behavior is identified. While in this case the spectral interpretation of the zeros is through the eigenvalues of Frobenius on cohomology groups, the local distribution of zeros is governed by the scaling limits of the eigenvalue distributions of the monodromy groups of families of L-functions. The calculations of these scaling limits appeals to methods from random matrix theory and these limits, at least for monodromies which come from the classical groups, are universally the CUE distributions. In the function field case one can also show [KS] that the low-lying zeros (i.e. zeros $\frac{1}{2}+i\gamma_F$ with γ_F near zero) as F runs over a family of L-functions, follow the laws governed by the corresponding scaling limit of monodromy groups of the family. This time the distributions are not universal and depend on the monodromy, or the symmetry of the family. Again, it is remarkable that this phenomenon of distribution of low-lying zeros persists for L-functions $L(s,F)$ for F in suitable families [KS]. This has been confirmed (again in restricted but wider ranges than for the high zeros) analytically for a number of families [IwLS], with different symmetry types. It has also been confirmed numerically in [Ru]. These distributions attached to each family and its symmetry also explain the specific fractions that appear in (62) and (63) (see also [So]). In (62) the symmetry is an orthogonal one $SO(\text{even})$ while for (63) it is $SO(\text{odd})$ (it is worth noting that subfamilies are independent entities and may have different symmetry types). Random matrix theory via these symmetries has recently been used to predict the asymptotics of all moments of L-functions on the line $\frac{1}{2}+it$ and for suitable families $L(\frac{1}{2},F)$, $F \in \mathcal{F}$ [ConrF], [KeS].

The results above about the distribution of zeros give ample evidence for there being a natural spectral interpretation of the zeros of $L(s,F)$ as well as the existence of a glue that marries different L-functions. However, these insights offer no real clue as to where such a spectral interpretation, or such symmetry groups may be found. There have been some interesting attempts to find nontautological spectral interpretations of the zeros of a given L-function such as $\zeta(s)$. In particular Connes [Con] suggests the singular space $X = K^*\backslash\mathbb{A}_K$ (P. Cohen has also pointed to this space and its intimate connection to the zeros of L-functions). The idele class group $J_K = K^*\backslash\mathbb{A}_K^*$ acts on X by multiplication $x \to yx$, $x \in X, y \in J_K$. He shows that with a suitable interpretation (via regularization) and assuming GRH, the decomposition of this action of multiplication over addition into irreducibles of J_K, yields exactly all the zeros of $\Lambda(s,F)$ where F is a $GL_1(\mathbb{A}_K)$ automorphic form. It turns out that this is closely connected to the explicit formula of Riemann which relates sums over zeros of an L-function to sums over primes and their powers, of the coefficients of the

L-function. Connes analysis gives a group action interpretation of the explicit formula. Weil [We3] had previously pointed to an arrangement of the terms in the explicit formula in attempting to interpret them in suggestive ways (so that they look like various key players in the function field setting). We note that anyway the explicit formula is a basic tool which is used analytically. For example, it is used directly in the analysis above concerning the distribution of the zeros. It is also used indirectly in the zero density theorems mentioned in section 4.

Whether these interpretations of the zeros or the explicit formula can be of any use in further understanding the zeros or attacking RH is unclear. Right now the use of families as a tool to study the zeros has been the most successful. We believe that families and understanding further what quantities to average as well as positivity will continue to play a central role perhaps even in the big ascent. After all, it is this analysis that "puts the zeros on the line $\frac{1}{2}$" in the general case of varieties over finite fields. One can imagine a scenario, a short cut, where GRH is established via families before a suitable spectral interpretation is given (for example fictitious zeros off the line might be ruled out spectrally before the true zeros are spectrally understood). More likely however is that suitable spaces and spectral interpretations of the zeros will be given and their analysis through families lead to the complete understanding (i.e. GRH, distribution of zeros ...). Anyway, all this is wild speculation and this is no doubt a good place to stop.

References

[A] J. ARTHUR, The trace formula for noncompact quotients, Proceedings of International Congr of Math., Warsaw (1983), 849-859.

[AC] J. ARTHUR, L. CLOZEL, Simple algebras, base change and the advanced theory of the trace formula, in "Annals of Math Studies", 120, Princeton University Press (1989).

[AuS] R. AURICK, F. STEINER, Exact theory for the quantum eigenstates of a strongly chaotic system, Physica D48 (1991), 445-470.

[B1] E. BOMBIERI, On the large sieve, Mathematika 12 (1965), 201-225.

[B2] E. BOMBIERI, Problems of the millennium: The Riemann hypothesis, in the home page of the Clay Mathematics Institute (webmaster@claymath.org).

[BFI] E. BOMBIERI, J. FRIEDLANDER, H. IWANIEC, Primes in arithmetic progressions to large moduli, Acta Math. 156 (1986), 203-251.

[Bu] D. BUMP, The Rankin-Selberg method, in "Number Theory, Trace Formulas and Discrete Groups" (Aubert, Bombieri and Goldfeld, eds.), Academic Press (1988), 49-109.

[Bur] D. BURGESS, On character sums and L-series. II, Proc. London Math. Soc. 3:13 (1963), 24-36.

[C] J. CASSELS, Rational Quadratic Forms, Academic Press, 1978.

[CoW] J. COATES, A. WILES, On the conjecture of Birch and Swinnerton-Dyer, Invent. Math. 39 (1977), 223-251.

[CogP] J. COGDELL, I. PIATETSKI-SHAPIRO, Converse theorems for $GL(n)$, Publ. IHES 79 (1994), 157-214.

[CogPS] J. COGDELL, I. PIATETSKI-SHAPIRO, P. SARNAK, in preparation.

[Con] A. CONNES, Formule de trace en géometrie non-commutative et hypothesése de Riemann, C.R. Acad. Sc. Paris 323, ser I (1996), 1231-1236.

[ConrF] B. CONREY, D. FARMER, Mean values of L-functions and symmetry, preprint, 2000.

[ConrI] B. CONREY, H. IWANIEC, The cubic moment of central values of automorphic L-functions, Annals of Math., to appear.

[D1] P. DELIGNE, Formes madulaires et représentations ℓ-adiques, Sém. Bourbaki, (1968-1969), exposé 355, Lecture Notes in Math. 179, Springer-Verlag (1971), 139-172.

[D2] P. DELIGNE, La Conjecture de Weil I, II, Publ. IHES 48 (1974), 273-308; 52 (1981), 313-428.

[DuFI] W. DUKE, J. FRIEDLANDER, H. IWANIEC, Bounds for automorphic L-functions, Invent. Math. 112 (1993), 1-8.

[DuI1] W. DUKE, H. IWANIEC, Estimates for coefficients of L-functions. I, in "Automorphic Forms and Analytic Number Theory" (1989), CRM, Montreal (1990), 43-47.

[DuI2] W. DUKE, H. IWANIEC, Estimates for coefficients of L-functions. II, Proceedings of the Amalfi Conference on Analytic Number Theory (1989), Universita di Salerno (1992), 71-82.

[E] M. EICHLER, Quaternäre quadratische formen und die Riemannsche vernutung für die kongruenz zetafunktion, Arch. Math. 5 (1954), 355-366, 113-120.

[FI] J. FRIEDLANDER, H. IWANIEC, A mean value theorem for character sums, Michigan Math. J. 39 (1992), 153-159.

[G] P. GARRET, Decomposition of Eisenstein series; Rankin triple products, Annals of Math. 125 (1987), 209-235.

[GeJ] S. GELBART, H. JACQUET, A relation between automorphic representations of $GL(2)$ and $GL(3)$, Ann. Sci. Ecole Norm. Sup 11 (1978), 411–452.

[GelP] I. GELFAND, I. PIATETSKI-SHAPIRO, Automorphic functions and the theory of representations, Trupy Moskov. Mat 12 (1963), 389-412.

[GoJ] R. GODEMENT, H. JACQUET, Zeta functions of simple algebras, L.N.M. 260, Springer (1972).

[Gol] D. GOLDFELD, Gauss class number problem for imaginary quadratic fields, BAMS 13:1 (1985), 23-27.

[Goo] A. GOOD, The square mean of Dirichlet series associated with cusp forms, Mathematika 29 (1982), 2778-295.

[GrZ] B. GROSS, D. ZAGIER, Heegner points and derivatives of L-series, Invent. Math. 84 (1986), 225-320.

[Gu] J. GUO, On the positivity of the central value of automorphic L-functions for $GL(2)$, Duke Math. J. 83 (1996), 1-18.

[HK] M. HARRIS, S. KUDLA, The central value of a triple product L-function, Ann. of Math. 133 (1991), 605-672.

[He] R. HEATH-BROWN, Cubic forms in ten variables, Proc. London Math. Soc. 47 (1983), 225-257.

[Hec] E. HECKE, Mathematische Werke, Göttingin, 1959.

[Hee] K. HEEGNER, Diophantische Analysis und Modulfunktionen, Math. Z. 56 (1952), 227-253.

[Hej] D. HEJHAL, On the topography of Maass wave forms for $PSL(2,\mathbb{Z})$, J. Exp. Math. 1:4 (1992), 275-305.

[HoL] J. HOFFSTEIN, P. LOCKHART, Coefficients of Maass forms and the Siegel zero, Ann. Math. 140 (1994), 177-181.

[HsKK] J. HSIA, Y. KITAOKA, M. KNESER, Representations of positive definite quadratic forms, J. Reine Angew Math. 301 (1978), 132-141.

[I] J. IGUSA, Fibre systems of Jacobian varieties III. Fibre systems of elliptic curves, Amer. J. Math. 81 (1989), 453-476.

[Ih] Y. IHARA, Hecke polynomials as congruence ζ functions in elliptic modular case, Ann. Math. 85 (1967), 267-295.

[IwS] H. IWANIEC, P. SARNAK, The nonvanishing of central values of automorphic L-functions and Landau-Siegel zeros, Israel J. Math., to appear.

[Iw1] H. IWANIEC, Fourier coefficients of modular forms of half-integral weight, Invent. Math. 87 (1987), 385-401.

[Iw2] H. IWANIEC, Small eigenvalues of Laplacian for $\Gamma_0(N)$, Acta Arith. 56 (1990), 65-82.

[IwLS] H. IWANIEC, W. LUO, P. SARNAK, Low lying zeros of families of L-functions, I.H.E.S., to appear.

[JS] H. JACQUET, J. SHALIKA, On Euler products and the classification of automorphic representations I, Amer. J. Math. 103 (1981), 499-558.

[JPS] H. JACQUET, I. PIATETSKI-SHAPIRO, J. SHALIKA, Rankin-Selberg convolutions, Amer. J. Math. 105 (1983), 367-464.

[KS] N. KATZ, P. SARNAK, Zeros of zeta functions and symmetry, BAMS 36 (1999), 1-26.

[KeS] J. KEATING, N. SNAITH, Random matrix theory and L-functions at $s = 1/2$, preprint, 2000.

[KiS] H. KIM, P. SARNAK, Refined estimates towards the Ramanujan and Selberg conjectures, preprint, 2000.

[KiSh] H. KIM, F. SHAHIDI, Symmetric cube L-functions for GL_2 are entire, Annals of Math. 150 (1999), 645-662.

[Kn] M. KNESER, Darstelungsmasse indefiniter quadratische Formen, Math. Zeit. 77 (1961), 188-194.

[KoL] V. KOLYVAGIN, D. LUGACHEV, Finiteness of the Shafarevich-Tate group and the group of rational points for some modular abelian varieties, Leningrad Math. J. 1:5 (1990), 1229-1253.

[KowMV1] E. KOWALSKI, P. MICHEL, J. VANDERKAM, Rankin-Selberg L-functions in the level aspect, preprint, 2000.

[KowMV2] E. KOWALSKI, P. MICHEL, J. VANDERKAM, Nonvanishing of high derivatives of automorphic L-functions at the center of the critical strip, J. Reine Angew Math., to appear.

[Ku] N. KUZNETZOV, Petersson's conjecture for cusp forms of weight zero and Linnik's conjecture, sums of Kloosterman sums, Math. SB 111 (1980), 334-383.

[L] L. LAFFORGUE, preprint, 2000.

[La] R. LANGLANDS, Problems in the theory of automorphic forms, S.L.N. 170, Springer (1970), 18-86.

[Li1] Y.V. LINNIK, Dokl. Akad. Nauk SSSR 30 (1941), 292-294.

[Li2] Y.V. LINNIK, Additive problems and eigenvalues of the modular operators, Proc. ICM (Stockholm) (1962), 270-284.

[Li3] Y.V. LINNIK, Ergodic properties of algebraic fields, Ergebnisse, 45 Springer Verlag (1968).

[Lu] W. LUO, Nonvanishing of L-values and the strong Weyl law, preprint, 2000.

[LuRS] W. LUO, Z. RUDNICK, P. SARNAK, On the generalized Ramanujan conjecture for $GL(n)$, Proc. Symp. Pure Math. 66, part 2, AMS, (1999).

[M] L. MEREL, Bornes pour la torsion de courbes elliptiques sur les corps de nombres, Invent. Math. 124 (1996), 437-449.

[Me] T. MEURMAN, On the order of the Maass L-function on the critical line, Number Theory, Vol. I, Budapest Colloq. Math. Soc., Janos Bolyai, 51 (1990).

[Mi] P. MICHEL, Autour de la conjecture de Sato-Tate pour les sommes de Kloosterman. I, Invent. Math. 121 (1995), 61-68; Minorations de sommes d'exponentielles, Duke Math. J. 95:2 (1998), 227-240.

[Mo] G. MOLTENI, Upper and lower bounds at $s = 1$ for certain Dirichlet series with Euler product, preprint, 2000.

[Mon] H. MONTGOMERY, The pair-correlation of zeros of the zeta function, Proc. Symp. Pure Math., AMS 24 (1973), 181-193.

[O] A. ODLYZKO, The 10^{20}-th zero of the Riemann zeta function and 70 million of its neighbors, preprint, A.T.T., 1989.

[P] H. PETERSSON, Zur analytischeu theorie de grenzkreisgruppen, Math. Z. 44 (1938), 127-155.

[PhS] R. PHILLIPS, P. SARNAK, On cusp forms for cofinite subgroups of $PSL(2,\mathbb{R})$, Invent. Math. 80 (1985), 339-364.

[R] R. RANKIN, Contributions to the theory of Ramanujan's function $\tau(n)$, Proc. Camb. Phil. Soc. 35 (1939), 351-356.

[Ru] M. RUBINSTEIN, Evidence for a spectral interpretation of zeros of L-functions, 1998, thesis, Princeton.

[Rud] Z. RUDNICK, Letter to Peter Sarnak (1999).

[RudS1] Z. RUDNICK, P. SARNAK, Zeros of principle L-functions and random matrix theory, Duke Math. J. 82 (1996), 269-322.

[RudS2] Z. RUDNICK, P. SARNAK, The behavior of eigenstates of arithmetic hyperbolic manifolds, CMP 161 (1991), 195-213.

[Rum] R. RUMELY, Numerical computations concerning ERH, Math. Comp. 61 (1993), 415-440.

[S1] P. SARNAK, Integrals of products of eigenfunctions, IMRN 6 (1994), 251-260.

[S2] P. SARNAK, Subconvexity estimates for Rankin-Selberg L-functions, in preparation.

[Sa1] I. SATAKE, Spherical functions and the Ramanujan conjecture, Proc. Symp., AMS 9 (1971), 258-264.

[Sa2] I. SATAKE, Theory of spherical functions on reductive algebraic groups over p-adic fields, Publ. IHES 18 (1983).

[Sc] R. SCHULZE-PILLOT, Exceptional integers for genera of integral ternary positive definite quadratic forms, Duke Math. J. 102 (2000), 351-357.

[Se] A. SELBERG, On the estimation of Fourier coefficients of modular forms, Proc. Symp. Pure Math. VIII, AMS, Providence (1965), 1-15.

[Sh1] F. SHAHIDI, Automorphic L-functions-a survey, in "Automorphic Forms, Shimura Varieties and L-functions" (L. Clozel and A. Milne, eds.), Academic Press (1988), 49-109.

[Sh2] F. SHAHIDI, On Ramanujan's conjecture and the finiteness of poles of certain L-functions, Ann. of Math. 127 (1988), 547-584.

[Shi1] G. SHIMURA, On modular forms of half integral weight, Annals of Math. 97 (1973), 440-481.

[Shi2] G. SHIMURA, On the holomorphy of certain Dirichlet series, Proc. London Math. Soc. 31 (1975), 79-98.

[Shi3] G. SHIMURA, Hilbert modular forms of half-integral weight, Duke Math. J. 71 (1993), 501-557.

[Si1] C.L. SIEGEL, Additive Theorie der Zahlkörper II, Math. Ann. 88 (1923), 184-210.

[Si2] C.L. SIEGEL, Über die classenzahl quadratischer Zahlkörper, Acta Arith. 1 (1935), 83-86.

[Si3] C.L. SIEGEL, Sums of m-th powers of algebraic integers, Ann. of Math. 46 (1945), 313-339.

[Si4] C.L. SIEGEL, Lectures on the Analytic Theory of Quadratic Forms, Robert Peppermüller, Göttingen, 1963.

[So] K. SOUNDARARAJAN, Nonvanishing of quadratic L-functions at $s = \frac{1}{2}$, Annals of Math., to appear.

[T] T. TAMAGAWA, On the zeta functions of a division algebra, Annals of Math. 77 (1963), 387-405.

[TaW] R. TAYLOR, A. WILES, Ring-theoretic properties of certain Hecke algebras, Annals of Math. 141 (1995), 553-572.

[W] J. WALDSPURGER, Sur les coefficients de Fourier de formes modulaires de poids demi entier, J. Math. Pures et Appl. 60 (1981), 365-384.

[Wa] T. WATSON, Central value of the Rankin triple L-function for unramified Maass cuspforms, in preparation.

[We1] A. WEIL, On the Riemann hypothesis in function fields, Proc. Nat. Acad. Sci. 27 (1941), 345-349.

[We2] A. WEIL, On some exponential sums, Proc. Nat. Acad. Sci., USA, 34 (1948), 204-207.

[We3] A. WEIL, Sur les formules explicites de la théorie des nombres, Izv. Mat. Nauk. 36, 3-18.

[Wey] H. WEYL, Zur abschätzung von $\zeta(1+it)$, Math. Zeit. 10 (1921), 88-101.

[Wi] A. WILES, Modular elliptic curves and Fermat's last theorem, Annals of Math. 141 (1995), 443-551.

[Wo] S. WOLPERT, Disappearance of cusp forms in special families, Annals of Math. 139 (1994), 239-291.

HENRYK IWANIEC, Department of Mathematics, Rutgers University, New Brunswick, NJ 08903-2101, USA iwaniec@math.rutgers.edu

PETER SARNAK, Department of Mathematics, Princeton University, Princeton, NJ 08544, USA sarnak@math.princeton.edu