Equidistribution, L-functions and Ergodic theory: on some problems of Yu. V. Linnik

Philippe Michel, Akshay Venkatesh *

Contents

Abstract. An old question of Linnik asks about the equidistribution of integral points on a large sphere. This question proved to be very rich: it is intimately linked to modular forms, to subconvex estimates for L-functions, and to dynamics of torus actions on homogeneous spaces. Indeed, Linnik gave a partial answer using ergodic methods, and his question was completely answered by Duke using harmonic analysis and modular forms. We survey the context of these ideas and their developments over the last decades.

Mathematics Subject Classification (2000). Primary: 11F66; Secondary: 11F67, 11M41

Keywords. Automorphic L-functions, Central Values, Subconvexity, Equidistribution

1. Linnik's problems

Given Q a homogeneous polynomial of degree m in n variables with integral coefficients, a classical problem in number theory is to understand the integral representations of an integer d by the polynomial Q, as $|d| \to +\infty$. Let $V_{Q,d}(\mathbf{Z}) = \{\mathbf{x} \in \mathbf{Z}^n,\ Q(\mathbf{x}) = d\}$ denote the set of such representations (possibly modulo some obvious symmetries). If $|V_{Q,d}(\mathbf{Z})| \to +\infty$ with d, it is natural to investigate the distribution of the discrete set $V_{Q,d}(\mathbf{Z})$ inside the affine variety "of level d"

$$V_{Q,d}(\mathbf{R}) = \{\mathbf{x} \in \mathbf{R}^n,\ Q(\mathbf{x}) = d\}.$$

In fact, one may rather consider the distribution, inside the variety of fixed level $V_{Q,\pm 1}(\mathbf{R})$, of the radial projection $|d|^{-1/m}.V_{Q,d}(\mathbf{Z})$ (here $\pm$ is the sign of d) and one would like to show that, as $|d| \to +\infty$, the

*The research of the first author is partially supported by the Marie Curie RT Network "Arithmetic Algebraic Geometry" and by the "RAP" network of the Région Languedoc-Roussillon. The research of the second author is supported by a Clay Mathematics Research Fellowship and NSF grant DMS-0245606.

set $|d|^{-1/m}.V_{Q,d}(\mathbf{Z})$ becomes equidistributed on $V_{Q,\pm1}(\mathbf{R})$ with respect to some natural measure $\mu_{Q,\pm1}$ on $V_{Q,\pm1}(\mathbf{R})$. Here, to take care of the case where $V_{Q,d}(\mathbf{Z})$ and $\mu_{Q,\pm1}(V_{Q,\pm1}(\mathbf{R}))$ are infinite, equidistribution w.r.t. $\mu_{Q,\pm1}$ is defined by the following property: for any two sufficiently nice compact subsets $\Omega_1, \Omega_2 \subset V_{Q,\pm1}(\mathbf{R})$ one has

$$\frac{\big||d|^{-1/m}.V_{Q,d}(\mathbf{Z}) \cap \Omega_1\big|}{\big||d|^{-1/m}.V_{Q,d}(\mathbf{Z}) \cap \Omega_2\big|} \rightarrow \frac{\mu_{Q,\pm1}(\Omega_1)}{\mu_{Q,\pm1}(\Omega_2)} \text{ as } |d| \rightarrow +\infty. \tag{1.1}$$

The most general approach to this kind of problems is the circle method of Hardy/Littlewood. (Un)fortunately, that method is fundamentally limited to cases where the number of variables n is large compared with the degree m. To go further, one is led to make additional hypotheses on the varieties $V_{Q,d}$. It was anticipated by Linnik in the early 60's and systematically suggested by Sarnak in the 90's [Lin60, Lin63, Sar91], that for varieties which are homogeneous with respect to the action of some algebraic group $\mathbf{G}_{\mathbf{Q}}$, one should be able to take advantage of this action either via harmonic analysis or via ergodic theory or via a combination of both. Equidistribution problems on such homogeneous varieties are called (after Sarnak), equidistribution problems of Linnik's type.

By now, this expectation is largely confirmed by the resolution of wide classes of problems of Linnik's type ([Lin68, LS64, DRS93, EM93, EMS96, COU01, GO03, EO03]); and the methods developed to deal with them rely heavily on powerful techniques from harmonic analysis (Langlands functoriality, equidistribution of Hecke points and approximations to the Ramanujan/Petersson conjecture) or from ergodic theory (especially Ratner's classification of measures invariant under unipotent subgroups) possibly complemented by methods from number theory.

In this lecture we will not discuss that much the resolution of these important and general cases (for this we refer to [Esk98, Ull02]); instead, we wish to focus on three, much older, examples of low dimension and degree ($m = 2$, $n = 3$) which were originally studied in the sixties by Linnik and his school. Our point in highlighting these examples, is that the various methods developed to handle them (some of which even go back to Linnik in the 50's) are fairly different from the aforementioned ones which, in fact, may not apply[1] or at least not directly. In fact, we believe that the approches we are about to survey will be a starting point of new interesting developments.

The three problems are relative to polynomials of ternary quadratic forms of signatures $(3,0)$ and $(1,2)$ and these are problems of Linnik's type with respect to the action of the orthogonal group $\mathbf{G} = \mathrm{SO}(Q)$ on $V_{Q,d}$.

The first problem is for the definite quadratic form $Q(A,B,C) = A^2 + B^2 + C^2$. For d an integer, $V_{Q,|d|}(\mathbf{Z})$, is the set of representations of $|d|$ as a sum of three squares

$$V_{Q,|d|}(\mathbf{Z}) = \{(a,b,c) \in \mathbf{Z}^3,\ a^2 + b^2 + c^2 = |d|\}$$

and $V_{Q,1}(\mathbf{R}) = S^2$ is the unit sphere. We denote by

$$\mathscr{G}_d = |d|^{-1/2}.V_{Q,|d|}(\mathbf{Z})$$

the radial projection of $V_{Q,|d|}(\mathbf{Z})$ on S^2:

Theorem 1 (Duke [Duk88]). *For $d \rightarrow -\infty$, and $d \not\equiv 0, 1, 4 (\mathrm{mod}\, 8)$ the set $\mathscr{G}_d$ is equidistributed on S^2 w.r.t. the Lebesgue measure μ_{S^2}.*

It will be useful to recall the "accidental" isomorphism of $\mathrm{SO}(Q)$ with $\mathbf{G} = \mathrm{PG}(\mathrm{B}^{(2,\infty)}) = \mathrm{B}^\times_{2,\infty}/Z(\mathrm{B}^\times_{2,\infty})$ where $\mathrm{B}^{(2,\infty)}$ is the algebra of the Hamilton quaternions. This arises from the identification of the quadratic space $(\mathbf{Q}^3, Q)$ with the trace-0 Hamilton quaternions endowed with the norm form $N(z) = z.\overline{z}$ via the map $(a,b,c) \rightarrow z = a.i + b.j + c.k$.

The second and third problems are relative to the indefinite quadratic form $Q(A,B,C) = B^2 - 4AC$ which is the discriminant of the binary quadratic forms $q_{A,B,C}(X,Y) = AX^2 + BXY + CY^2$. In that case, there is another "accidental" isomorphism of $\mathrm{SO}(Q)$ with PGL_2 via the map

$$(a,b,c) \rightarrow q_{a,b,c}(X,Y) = aX^2 + bXY + cY^2$$

[1] for instance, in the problems presented below, the natural groups which acts on the sets $V_{Q,d}(\mathbf{Z})$, are (quotients of adelic) tori which, obviously, have no unipotent elements.

which identifies $V_{Q,d}$ with the set $\mathcal{Q}_d$ of binary quadratic forms of discriminant d; PGL_2 acts on the latter by linear change of variables, twisted by inverse determinant. As $\mathrm{PGL}_2(\mathbf{Z})$ acts on $\mathcal{Q}_d(\mathbf{Z})$, one sees that, if $V_{Q,d}(\mathbf{Z}) = \mathcal{Q}_d(\mathbf{Z})$ is non empty (i.e. if $d \equiv 0, 1 \bmod 4$), it is infinite; so the proper way to define the equidistribution of $|d|^{-1/2}.V_{Q,d}(\mathbf{Z})$ inside $V_{Q,\pm 1}(\mathbf{R}) = \mathcal{Q}_{\pm 1}(\mathbf{R})$ is via (1.1). However, instead of formulating the problems in these terms, it is useful to put them in a slightly different (although equivalent) form which has a modular interpretation and will prove suitable for number theoretic applications. Let $\mathbb{H}^{\pm} = \mathbb{H}^+ \cup \mathbb{H}^- = \mathbf{C} - \mathbf{R} = \mathrm{PGL}_2(\mathbf{R})/\mathrm{SO}_2(\mathbf{R}))$ denote the union of the upper and lower half-planes and $Y_0(1)$ denote the (non-compact) modular surface of full level i.e. $\mathrm{PGL}_2(\mathbf{Z})\backslash\mathbb{H}^{\pm} \simeq \mathrm{PSL}_2(\mathbf{Z})\backslash\mathbb{H}^+$.

As is well known, the quotient $\mathrm{PSL}_2(\mathbf{Z})\backslash\mathcal{Q}_d(\mathbf{Z})$ is finite, of cardinality some *class number* $h(d)$. For negative discriminants d, one associates to each $\mathrm{PSL}_2(\mathbf{Z})$-orbit $[q] \subset \mathcal{Q}_d(\mathbf{Z})$, the point $z_{[q]}$ in $Y_0(1)$ defined as the $\mathrm{PGL}_2(\mathbf{Z})$-orbit of the unique root of $q(X, 1)$ contained in $\mathbb{H}^+$. These points are called *Heegner points of discriminant*[2] d and we set

$$\mathscr{H}_d := \{z_{[q]},\ [q] \in \mathrm{PSL}_2(\mathbf{Z})\backslash\mathcal{Q}_d(\mathbf{Z})\} \subset Y_0(1).$$

An equivalent form to (1.1) for $Q(A, B, C) = B^2 - 4AC$ and $d \to -\infty$ is the following:

Theorem 2 (Duke [Duk88]). *As $d \to -\infty$, $d \equiv 0, 1 \bmod 4$, the set $\mathscr{H}_d$ becomes equidistributed on $Y_0(1)$ w.r.t. the Poincaré measure $d\mu_P = \frac{3}{\pi}\frac{dxdy}{y^2}$.*

For positive discriminants d, one associates to each class of integral quadratic form $[q] \in \mathcal{Q}_d(\mathbf{Z})$ the positively oriented geodesic, $\gamma_{[q]}$, in $Y_0(1)$ which is the projection to $Y_0(1)$ of the geodesic line in $\mathbb{H}^+$ joining the two (real) roots of $q(X, 1)$. This is a closed geodesic – in fact, all closed geodesics on $Y_0(1)$ are of that form – whose length is essentially equal to the logarithm of the fundamental solution to Pell's equation $x^2 - dy^2 = 4$. We denote by

$$\Gamma_d := \{\gamma_{[q]},\ [q] \in \mathrm{PSL}_2(\mathbf{Z})\backslash\mathcal{Q}_d(\mathbf{Z})\}$$

the set of all geodesics of discriminant d.

Theorem 3 (Duke [Duk88]). *As $d \to +\infty$, $d \equiv 0, 1 \bmod 4$, d not a perfect square, the set Γ_d becomes equidistributed on the unit tangent bundle of $Y_0(1)$, $S_1^*(Y_0(1))$, w.r.t. the Liouville measure $d\mu_L = \frac{3}{\pi}\frac{dxdy}{y^2}\frac{d\theta}{2\pi}$.*

These three problems (in their form (1.1)) where first proved by Linnik and by his student Skubenko by means of Linnik's *ergodic method*; we will return this method in section 6. The proof however is subject to an additional assumption which we call *Linnik's condition*, namely:

Theorem 4 (Linnik [Lin55, Lin60], Skubenko [Sku62]). *Let p be an arbitrary fixed prime, then the equidistribution statements of Theorems 1, 2, 3 hold for the subsequence of d such that* p is split *in the quadratic extension $K_d = \mathbf{Q}(\sqrt{d})$.*

As we'll see in section 6 Linnik's condition has a natural ergodic interpretation; the method of Linnik amounts to studying the dynamics of a $\mathbf{Q}_p^\times$-action on a homogeneous space. It can be somewhat relaxed to the condition that for each d there is a prime $p = p(d) \leqslant |d|^{\frac{1}{10^{10}\log\log|d|}}$ which splits in $\mathbf{Q}(\sqrt{d})$. The latter condition is satisfied by assuming that the L-functions of quadratic characters satisfying the generalized Riemann hypothesis or even by assuming the much weaker (still unproven) statement that these L-functions have no zeros in a $\frac{\log\log|d|}{\log|d|}$-neighborhood of 1. In particular, Linnik's condition (resp. the weaker one) is automatically fullfilled for subsequences of d such that K_d is a *fixed* quadratic field (resp. $\mathrm{disc}(K_d) = \exp(O(\frac{\log|d|}{\log\log|d|})))$; however, in these cases, the proof of Theorems 1, 2, 3 is much simpler (see [CU04] for instance), so, as it is the hardest and (from our perspective at least) the most important case, we will limit ourselves to d's which are fundamental discriminants (i.e. $d = \mathrm{disc}(K_d)$).

Acknowledgements. The first author is scheduled to give a presentation based on this work in the 2006 ICM. Since much of it is based on our joint work, we have decided to write this paper jointly. The results of Section 6 are all joint work with M. Einsiedler and E. Lindenstrauss and will also be discussed in their contribution to these proceedings [EL06].

It is our pleasure to thank Bill Duke, Henryk Iwaniec and Peter Sarnak for both their consistent encouragement and for many beautiful ideas which underlie the whole field. Peter Sarnak carefully read an early draft and provided many helpful comments and corrections. We also would like to thank our collaborators Manfred Einsiedler and Elon Lindenstrauss, for patiently explaining ergodic ideas and methods to us.

[2] for simplicity, we will ignore non-primitivite forms

2. Linnik's problems via harmonic analysis

Duke's unconditional solution of Linnik's problems is via harmonic analysis and is very different from Linnik's original ergodic approach. In a sense, it is more direct as it proceeds by verifying Weyl's equidistribution criterion. Let (X,μ) denote any of the probability spaces (S^2,μ_{S^2}), $(Y_0(1),\mu_P)$, $(S^1_*(Y_0(1)),\mu_L)$ and for each case and for appropriate d, let μ_d denote the probability measure formed out of the respective sets $\mathscr{G}_d$, $\mathscr{H}_d$ or Γ_d: for instance for $X=S^2$,

$$\int_{S^2}\varphi\mu_d=\frac{1}{|\mathscr{G}_d|}\sum_{\substack{(a,b,c)\in\mathbf{Z}^3\\ a^2+b^2+c^2=|d|}}\varphi(\frac{a}{\sqrt{|d|}},\frac{b}{\sqrt{|d|}},\frac{c}{\sqrt{|d|}}).$$

Showing that μ_d weak-$*$ converges to μ amounts to show that, for any φ ranging over a fixed orthogonal basis (made of continuous functions) of the L^2-space $L^2_0(X,\mu)$, the Weyl sum

$$W(\varphi,d):=\int_X\varphi\mu_d,\ \text{converges to 0 as } |d|\rightarrow+\infty. \tag{2.1}$$

In the context of Theorem 1 (resp. Theorem 2, resp. Theorem 3) such bases are taken to consist of non-constant harmonic polynomials (resp. Maass forms and Eisenstein series of weight 0, resp. Maass forms and Eisenstein series of non-negative, even, weight).

2.1. Duke's proof.

The decay of the period integral $W(\varphi,d)$ is achieved by realizing it in terms of the d-th Fourier coefficient of a modular form of half-integral weight and level 4; this modular form – call it $\tilde{\varphi}$ – is obtained from φ through a theta correspondance.

In the case of Theorem 1, and when φ is a non-constant harmonic polynomial of degree r, this comes from the well known fact that the theta-series

$$\tilde{\varphi}(z)=\theta_\varphi(z)=\sum_{|d|\geqslant1}\Big(\sum_{\substack{(a,b,c)\in\mathbf{Z}^3\\ a^2+b^2+c^2=|d|}}\varphi(a,b,c)\Big)e(|d|z)$$

is a modular form of weight $k=3/2+r$ for the modular group $\Gamma_0(4)$. This is a special case of a (theta) correspondance of Maass, which itself is now a special case of the theta correspondance for dual pairs; it associates to an automorphic form φ for an orthogonal group $\mathrm{SO}_{p,q}$ of signature (p,q), a Maass form of weight $(q-p)/2$, $\tilde{\varphi}$. Moreover, Maass provided a formula expressing the Fourier coefficients of $\tilde{\varphi}$ in terms of a certain integral of φ.

By the accidental isomorphisms recalled above, this provides a correspondance between automorphic forms either for $\mathrm{B}^\times_{2,\infty}$ or for PGL_2, and modular forms of half-integral weight. Under this correspondance, one has, for d a fundamental discriminant ($d=\mathrm{disc}(\mathcal{O}_K)$ for $\mathcal{O}_K$ the ring of integers of the quadratic field $K=\mathbf{Q}(\sqrt{d})$)

$$W(\varphi,d)=c_{\varphi,d}\frac{\rho_{\tilde{\varphi}}(d)|d|^{-1/4}}{L(\chi_d,1)} \tag{2.2}$$

where $c_{\varphi,d}$ is a constant depending on φ and mildly on d (i.e. one has $|d|^{-\varepsilon}\ll_{\varphi,\varepsilon}c_{\varphi,d}\ll_\varepsilon|d|^\varepsilon$ for any $\varepsilon>0$), $\rho_{\tilde{\varphi}}(d)$ denotes the d-th fourier coefficient of $\tilde{\varphi}$ (here, the Fourier expansion of $\tilde{\varphi}$ is normalized so that the analog of the Ramanujan/Petersson conjecture for half-integral weight forms is $\rho_{\tilde{\varphi}}(d)\ll_{\varepsilon,\tilde{\varphi}}|d|^\varepsilon$ for d squarefree) and χ_d is the quadratic character corresponding to K_d.

In particular, by Siegel's lower bound $L(\chi_d,1)\gg_\varepsilon|d|^{-\varepsilon}$, (2.1) is consequence of a bound of the form

$$\rho_{\tilde{\varphi}}(d)\ll|d|^{1/4-\delta} \tag{2.3}$$

for some absolute $\delta>0$. The bound (2.3) is to be expected; indeed the half-integral weight analog of the Ramanujan/Petersson conjecture – which follows from GRH – predicts that any $\delta<1/4$ is admissible. Unlike the situation in integral weight, this half-integral weight analogue does not follow from the Weil conjectures.

The problem of bounding Fourier coefficient of modular forms can be approached through a Petersson/Kuznetzov type formula (due to Proskurin in the half-integral weight case): (un)fortunately the standard bound for the Salié sums occuring in the formula yield the above estimate only for $\delta < 0$. This "barricade" was eventually surmounted by Iwaniec (with the value $\delta = 1/28$, [Iwa87]) for $\tilde{\varphi}$ a holomorphic form of weight $\geqslant 5/2$ and by B. Duke for general weight as well as Maass forms by adapting Iwaniec's argument, and thus concluding the first fully unconditional proof of Theorems 1, 2, 3.

We will not discuss the proof of Iwaniec's bound, excepted to say that it uses the half-integral weight Petersson/Kuznetzov formula, the very special structure of the Salié sums, and finally a fundamental trick of performing an averaging over the level of the forms in some optimal range. We mention this last point because, although we are unable to offer any idea of "why" Iwaniec's trick works, it would be extremely useful to have a conceptual interpretation of it; there are indeed other interesting equidistribution problems in which the Weyl sums are connected to Fourier coefficients of automorphic forms but on more exotic groups. A very interesting case, pointed out to us by Dick Gross, is equidistribution problem for the flats associated to totally cubic rings of large discriminant in $\mathrm{PGL}_3(\mathbf{Z})\backslash\mathrm{PGL}_3(\mathbf{R})$ (see Section 6 for an ergodic approach to this problem), the associated Weyl sums should be related to Fourier coefficients of automorphic forms for the exceptional group G_2 (see [GGS02, GS03, GS04] for the first steps in the direction of such a connection).

As Duke pointed out in [Duk88], some other values of the signature (p, q) yield exceptional isomorphisms between $\mathrm{SO}(p, q)$ and groups which carry a nice modular interpretation (i.e. connected to Shimura varieties or their reductions): namely the cases of signature $(3, 2)$ and $(2, 2)$. In [Coh05], Cohen extended Duke's approach to these cases and obtained, amongst other, new equidistribution results for special, positive dimensional, Siegel subvarieties of 3-folds.

2.2. Equidistribution and subconvex bounds for L-functions.

Shortly after Duke's proof, another approach emerged which turned out to be very fruitful, namely the connection between the decay of Weyl's sums 2.1 and the *subconvexity problem* for automorphic L-function. In Section 3 we give a full definition of the subconvexity problem; in short, it consists of giving "nontrivial" bounds on the size of an automorphic L-function on the critical line.

2.2.1. Weyl's sums as period integrals: Waldspurger type formulae. It goes back to Gauss that the set of classes of quadratic forms $\mathrm{SL}_2(\mathbf{Z})\backslash\mathcal{Q}_d(\mathbf{Z})$ has the structure of a finite commutative group (the class group) $\mathrm{Cl}(d)$. In particular for the second problem ($d < 0$), $\mathscr{H}_d$ is a homogeneous space under the action of $\mathrm{Cl}(d)$ and the Weyl sums can be seen as period integrals for this action:

$$W(\varphi, d) = \int_{\mathrm{Cl}(d)} \varphi(\sigma.z_d) d\mu_{\mathrm{Haar}}(\sigma).$$

In a similar way, the Weyl's sums over $\mathscr{G}_d$ and Γ_d can be realized as orbital integrals for the action of some class group.[3] The connection between such orbital integrals and L-functions follows from a formula basically due to Waldspurger. To describe it in greater detail it is useful and convenient to switch an adelic description of the Weyl's sums. Such description makes clear the unity of Theorems 1–3.

Let us recall recall that in the context of Theorem 1 with $Q(A, B, C) = A^2 + B^2 + C^2$ (resp. Theorems 2 and 3, with $Q(A, B, C) = B^2 - 4AC$) a solution $Q(a, b, c) = d$ gives rise to an embedding of the quadratic $\mathbf{Q}$-algebra K_d into the $\mathbf{Q}$-algebra $\mathrm{B}^{(2,\infty)}$ (resp. $M_{2,\mathbf{Q}}$) by sending $\sqrt{d}$ to $a.i + b.j + c.k$ (resp. $\begin{pmatrix} b & -2a \\ 2c & -b \end{pmatrix}$). This yields an embedding of $\mathbf{Q}$-algebraic groups, $\mathbf{T}_d := \mathrm{res}_{K/Q}\mathbb{G}_m/\mathbb{G}_m \hookrightarrow \mathbf{G}$, where $\mathbf{G} = \mathrm{PG}(\mathrm{B}^{(2,\infty)})$ (resp. $= \mathrm{PGL}_2$).

Let $\mathrm{K}_{f,\max}$ be a maximal compact subgroup of $\mathbf{G}(\mathbf{A}_f)$ in all three cases. In the context of Theorem 1 (resp. Theorem 2, resp. Theorem 3) take $\mathrm{K}_\infty = \mathbf{T}_d(\mathbf{R}) \cong \mathrm{SO}_2 \subset \mathbf{G}$ (resp. $\mathrm{K}_\infty = \mathbf{T}_d(\mathbf{R})$, resp. $\mathrm{K}_\infty = \{1\}$) and set $\mathrm{K} = \mathrm{K}_{f,\max}\mathrm{K}_\infty$; the quotient $\mathbf{G}(\mathbf{Q})\backslash\mathbf{G}(\mathbf{A_Q})/\mathrm{K}$ then equals a quotient of S^2 by a finite group of rotations (resp. $Y_0(1)$, resp. the unit tangent bundle of $Y_0(1)$).

It transpires, with these identifications, the subsets

$$\mathscr{G}_d \subset S^2,\ \mathscr{H}_d \subset Y_0(1),\ \Gamma_d \subset S^1_*(Y_0(1))$$

[3] For the first case, a connection between $V_{Q,|d|}(\mathbf{Z})$ and $\mathrm{Cl}(d)$ ($d < 0$) goes back to Gauss: to any $(a, b, c) \in V_{Q,|d|}(\mathbf{Z})$, one associates the $\mathrm{SL}_2(\mathbf{Z})$-class of the quadratic form of discriminant d obtained by restricting the square of the euclidean norm in $\mathbf{R}^3$ to the $\mathbf{Z}$-lattice $(a, b, c)^\perp \cap \mathbf{Z}^3$. Similarly, one can relate Γ_d to the idele class group of a real quadratic field.

may be uniformly described, after choosing a solution z_d, as a compact orbit of the adelic torus $\mathbf{T}_d$:

$$\mathbf{T}_d(\mathbf{Q})\backslash z_d.\mathbf{T}_d(\mathbf{A_Q})/\mathrm{K}_{\mathbf{T}_d} \subset \mathbf{G}(\mathbf{Q})\backslash\mathbf{G}(\mathbf{A_Q})/\mathrm{K}$$

where $\mathrm{K}_{\mathbf{T}_d} := \mathbf{T}_d(\mathbf{A_Q}) \cap \mathrm{K}$. In this notation the Weyl sum is given as a *toric integral*

$$W(\varphi, d) = \int_{\mathbf{T}_d(\mathbf{Q})\backslash\mathbf{T}_d(\mathbf{A_Q})/K_{\mathbf{T}_d}} \varphi(z_d.t)dt \tag{2.4}$$

when dt is the Haar measure on the toric quotient. A superficial advantage of this notation is that it allows for a uniform presentation of many equidistribution problems for "cycles" associated with quadratic orders in locally symmetric spaces associated to quaternion algebras. Indeed, as we shall see below that one can, under suitable compatibiility hypotheses, consider the above equidistribution problems while changing

- the group $\mathbf{G}$ to $\mathbf{G} = \mathrm{B}^\times/Z(\mathrm{B}^\times)$ for B any quaternion algebra over $\mathbf{Q}$ (definite or indefinite);
- the compact $\mathrm{K}_{f,\max}$ to a compact subgroup $\mathrm{K}'_f \subset \mathrm{K}_{f,\max}$ (i.e. changing the level structure)
- the subgroup $\mathrm{K}_{\mathbf{T}_d}$ to a subgroup $\mathrm{K}'_{\mathbf{T}_d}$ (i.e. considering cycle associated to suborders $\mathcal{O}$ of the maximal order $\mathcal{O}_d$).
- the base field $\mathbf{Q}$ to a fixed totally real number field F.

When φ is a *new cuspform* (the L^2 normalized *new vector* in some automorphic representation π), Waldspurger's formula [Wal85] relates $|W(\varphi, d)|^2$ (and correspondingly the square of the d-th Fourier coefficient $|\rho_{\tilde{\varphi}}(d)|^2$) to the central value of an automorphic L-function. In its original form, the formula was given up to some non-zero proportionality constant; as we are interested in the size $W(\varphi, d)$ a more precise expression is needed. Thanks to the work of many people ([KZ81, GZ86, Gro87, Hat90, KS93, CU05, Zha01, Zha01b, Pop06, Xue05, Xue05b]) notably Gross, Zagier and Zhang such an expression is by now available in considerable generality. Under suitable hypotheses (which in the present cases are satisfied), it has the following form

$$|W(\varphi, d)|^2 = c_{\varphi,d}\frac{L(\pi, 1/2)L(\pi \times \chi_d, 1/2)}{L(\chi_d, 1)^2\sqrt{|d|}} \tag{2.5}$$

where π' is a GL_2-automorphic representation corresponding to π by the Jacquet/Langlands correspondance and $c_{\varphi,d} > 0$ is a constant which depends mildy on d.

The Waldspurger formula (2.5) is more powerful than (2.2) as it may be extended to a formula for more general toric integrals. Indeed, let χ be a character of the torus $\mathbf{T}_d(\mathbf{Q})\backslash\mathbf{T}_d(\mathbf{A_Q}) = K_d^\times\mathbf{A}_\mathbf{Q}^\times\backslash\mathbf{A}_{K_d}^\times$ trivial on $\mathrm{K}_{\mathbf{T}_d}$. Under suitable compatibility assumptions between χ and φ and possibly under additional coprimality assumptions between the conductors of of π, χ, the relation (2.5) generalizes to

$$|W_\chi(\varphi, d)|^2 = c_{\varphi,d_\chi,\chi_\infty}\frac{L(\pi \times \pi_\chi, 1/2)}{L(\chi_d, 1)^2\sqrt{|d_\chi|}} \tag{2.6}$$

where $W_\chi(\varphi, d)$ is a *twisted toric integral* of the form

$$W_\chi(\varphi, d_\chi) = \int_{\mathbf{T}_d(\mathbf{Q})\backslash\mathbf{T}_d(\mathbf{A_Q})} \chi(t)\varphi(z_{d_\chi}.t)dt,$$

π_χ is the GL_2-automorphic representation (of conductor d_χ) corresponding to χ by quadratic automorphic induction and $L(\pi \times \pi_\chi, s)$ is the Rankin/Selberg L-function of the pair (π, π_χ).

2.3. Subconvexity, equidistribution and sparse equidistribution.

We see, from formula (2.5) and Siegel's lower bound that (2.1) follows from the bound

$$L(\pi \times \chi_d, 1/2) \ll_\pi |d|^{1/2-\delta}; \tag{2.7}$$

for some absolute $\delta > 0$; subject to this bound, one obtains another proof of Linnik's equidistribution problems. More generally, we see from (2.6) that the twisted Weyl sums are decaying, i.e.

$$W_\chi(\varphi, d_\chi) \to 0 \text{ for } d_\chi \to +\infty, \tag{2.8}$$

as soon as

$$L(\pi \times \pi_\chi, 1/2) \ll |d_\chi|^{1/2-\delta}. \quad (2.9)$$

Both (2.7) and (2.9) are special cases of subconvex bounds for central values of automorphic L-functions and have been proven (see below).

One should note that the decay of the twisted toric integral is useful if one needs to perform *harmonic analysis along* the toric orbit $\mathbf{T}_d(\mathbf{Q})\backslash z_d.\mathbf{T}_d(\mathbf{A_Q})/\mathrm{K}_{\mathbf{T}_d}$: this is particular the case when one needs equidistribution only for a stricly smaller suborbit of the full orbit, a problem we call a *sparse equidistribution problem.*

For instance one has:

Theorem 5 ([HM06]). *There is an absolute constant* $0 < \eta < 1$ *such that: for each fundamental discriminant* $d < 0$*, choose* $z_{0,d} \in \mathscr{H}_d$ *a Heegner point and choose* G_d *a subgroup of* $\mathrm{Cl}(d)$ *of size* $|G_d| \geqslant |\mathrm{Cl}(d)|^\eta$ *then the sequence of suborbits*

$$\mathscr{H}'_d := G_d.z_{0,d} = \{\sigma.z_{0,d},\ \sigma \in G_d\}$$

is equidistributed on $Y_0(1)$ *w.r.t.* μ_P*.*

One has also similar sparse equidistribution results for sufficiently large suborbits of $\mathscr{G}_d$ on the sphere and for suffciently large geodesic segments of Γ_d [Mic04, Pop06]. Note however that the present method has fundamental limitations as one cannot take η too close to 0: even under the GRH, one would prove equidistribution only for $\eta > 1/2$. Nevertheless we would like to formulate the following

Conjecture 1. (Equidistribution of subgroups) Fix any $\eta > 0$ and for each fundamental discriminant $d < 0$, choose $z_{0,d} \in \mathscr{H}_d$ a Heegner point and choose G_d a subgroup of $\mathrm{Cl}(d)$ of size $|G_d| \geqslant |d|^\eta$. Then as $|d| \to +\infty$, the sequence of suborbits

$$\mathscr{H}'_d := G_d.z_{0,d} = \{\sigma.z_{0,d},\ \sigma \in G_d\}$$

is equidistributed on $Y_0(1)$ w.r.t. μ_P.

This conjecture is certainly difficult in general; however, we expect that, by ergodic methods, significant progress might be made, at least for subgroups G_d that satisfy suitable versions of Linnik's condition for some fixed prime p (see Section 6 for evidence in that direction). One might consider Conjecture 1 to be a homogeneous space analogue of the results of Bourgain-Konyagin [BK03] on small subgroups of $(\mathbf{Z}/p\mathbf{Z})^\times$. In a related vein we formulate:

Conjecture 2. (Mixing conjecture) For each fundamental discriminant $d < 0$, let $\sigma_d \in \mathrm{Cl}(d)$ be such that the minimal norm of any integral ideal representing the class σ_d approaches ∞ as $|d| \to \infty$. Then, as $|d| \to \infty$, the sequence

$$\mathscr{H}''_d = \{(z, \sigma_d z) : z \in \mathscr{H}_d\} \subset Y_0(1) \times Y_0(1)$$

is equidistributed w.r.t. $\mu_P \times \mu_P$.

One can formulate natural generalizations of Conjecture 2: e.g. one can replace $Y_0(1) \times Y_0(1)$ by $Y_0(1) \times \big(S^2/\mathrm{SO}_3(\mathbf{Z})\big)$ and, fixing base points $z_1 \in \mathscr{H}_d$, $z_2 \in \mathscr{G}_d$, replace the role of $\mathscr{H}''_d$ by the set $\{(\sigma.z_1, \sigma.z_2) : \sigma \in \mathrm{Cl}(d)\}$. Similarly one can formulate a version with even more factors, or replacing $\mathrm{Cl}(d)$ by a subgroup of size d^η. See also Section 6.4.1.

2.4. Equidistribution and non-vanishing of L-functions.

Before continuing with the subconvexity problem, we would like to point out another interesting application. It combines subconvexity, equidistribution and the period relation (2.5) and applies them to the non-vanishing of L-functions. It should be noted that the field of proving non-vanishing for L-functions is a vast one, with many techniques that have proved successful in different contexts; the technique we present here is only one of many, and in fact achieves much weaker results than have been obtained in other contexts.

Consider, for simplicity, the context of Theorem 2 (see also [MV06]): let φ be a Maass-Hecke eigenform of weight 0 and π be its associated automorphic representation. If one averages (2.5) over the characters of $\mathrm{Cl}(d)$, one obtains by orthogonality (in that case the constants $c_{\varphi,d_\chi,\chi_\infty} = c > 0$ are all equal to an absolute constant)

$$c\frac{\sqrt{d}}{|\mathrm{Cl}(d)|^2}\sum_{\chi \in \widehat{\mathrm{Cl}}(d)} L(\pi \times \pi_\chi, 1/2) = \int_{Y_0(1)} |\varphi|^2.\mu_d$$

and since by Theorem 2

$$\int_{Y_0(1)} |\varphi|^2.\mu_d \to \int_{Y_0(1)} |\varphi(z)|^2 d\mu_P(z) > 0, \text{ as } |d| \to +\infty$$

this shows that for some χ the central value $L(\pi \times \pi_\chi, 1/2)$ does not vanish.

Moreover, by the subconvex bound (2.9), one obtains a quantitative form of non-vanishing

$$|\{\chi \in \widehat{\mathrm{Cl}}(d),\ L(\pi \times \pi_\chi, 1/2) \neq 0\}| \gg |d|^\eta \tag{2.10}$$

for some absolute $\eta > 0$.

By considering equidistribution relative to definite quaternion algebras, one can obtain similar non-vanishing results for central values $L(\pi \times \pi_\chi, 1/2)$ where π_∞ is in the discrete series and the sign of the functional equation of $L(\pi \times \pi_\chi, s)$ is $+1$. In particular when $\pi = \pi_E$ is the automorphic representation associated to an elliptic curve $E/\mathbf{Q}$, such estimates provide a lower bound for the size of the "rank-0" part of the group $E(H_K)$ of points of E which are rational over the Hilbert class field of K as $|d| \to +\infty$.

Remark 2.1. When π corresponds to an Eisenstein series, stronger results where obtained before by Duke/Friedlander/Iwaniec and Blomer [DFI95, Blo04]; although this it appears in a somewhat disguised (and more elaborate) form, the basic principle underlying the proof is the same.

An interesting problem is to adress the case where the sign of the functional equation is -1. In this case, $L(\pi \times \pi_\chi, 1/2) = 0$ and one considers instead the question of non-vanishing of the first derivative $L'(\pi \times \pi_\chi, 1/2)$. At least when π_∞ is in the holomorphic discrete series and π has trivial central character , the Gross/Zagier formula (and its extensions by Zhang) interprets $L'(\pi \times \pi_\chi, 1/2)$ as the "height" of some Heegner cycle above some modular (or Shimura) curve. This is not quite a period integral; however the height decomposes as a sum of local heights indexed by the places v of $\mathbf{Q}$. These local heights are either simple or can be interpreted as periods integrals over quadratic cycles associated with K which live over appropriate adelic quotients $\mathbf{G}^{(\mathbf{v})}(\mathbf{Q})\backslash\mathbf{G}^{(\mathbf{v})}(\mathbf{A})/\mathrm{K}_\mathbf{v}$ where $\mathbf{G}^{(v)}$ is associated to a quaternion algebra $\mathrm{B}^{(v)}$ ramified at v.

It seem then plausible that one can compute the asymptotic of the average $\sum_\chi L'(\pi \times \pi_\chi, 1/2)$ by using the equidistribution property of quadratic cycles on these infinitely many quotients. One consequence of this would then be, for compatible E and K, a lower bound for the rank of $E(H_K)$:

$$\mathrm{rank}_\mathbf{Z}\ E(H_K) \gg |d|^\eta$$

for some $\eta > 0$ as $|d| \to +\infty$.

Remark 2.2. A few years ago, Vatsal and Cornut [Vat02, Vat03, CV04] used period relations and equidistribution in a similar way to obtain somewhat stronger non-vanishing results for Rankin/Selberg L-functions but associated to anti-cyclotomic[4] characters of a *fixed* imaginary quadratic field. This is in constrast with the present case where one can allow the quadratic field to vary. Note that one of their main ingredient to obtain equidistribution came for ergodic theory and precisely from Ratner's classification of measures invariant under unipotent subgroups. We expect that the ergodic methods to be described in section 6 (which are distinct from Ratner's) will enable one to obtain results stronger than say (2.10) for a possibly varying quadratic field.

3. The subconvexity problem

Although the subconvexity problem is a venerable topic in number theory – its study begins with Weyl's estimate $|\zeta(1/2+it)| \ll_\epsilon t^{1/6+\epsilon}$– there has been a renaissance of interest in it recently. This owes largely to the observation that a resolution of the subconvexity problem for automorphic L-functions on GL has many striking applications, as we have just seen to Linnik's equidistribution problems or to "Arithmetic Quantum Chaos." We refer to [IS00] for a discussion of all these questions in the broader context of the analytic theory of automorphic L-functions.

[4]The case of cyclotomic characters was carried out even earlier by Rohrlich, by more direct methods.

Let $\Pi = \Pi_\infty \otimes \bigotimes_p' \Pi_p$ some reasonable "automorphic object": by automorphic object we mean, for instance an automorphic representation or more generality an admissible representation constructed out of automorphic representations via the formalism of L-groups (for instance the Rankin/Selberg convolution $\pi_1 \times \pi_2$ of two automorphic representations on some linear groups). To Π, one can usually associate a collection of local L-factors

$$L(\Pi_p, s) = \prod_{i=1}^{d}(1 - \frac{\alpha_{\Pi,i}(p)}{p^s})^{-1},\ p \text{ prime},\ L(\Pi_\infty, s) = \prod_{i=1}^{d}\Gamma_{\mathbf{R}}(s - \mu_{\Pi,i})$$

where $\Gamma_{\mathbf{R}}(s) = \pi^{-s/2}\Gamma(s/2)$ and $\{\alpha_{\Pi,i}(p)\},\ \{\mu_{\Pi,i}\}$ are called the local numerical parameters of Π at p and at infinity; from these local datas one forms a global L-function

$$L(\Pi, s) = \sum_{n\geqslant 1}\frac{\lambda_\Pi(n)}{n^s} = \prod_p L(\Pi_p, s).$$

In favourable cases, one can show that $L(\Pi, s)$ has analytic continuation to $\mathbf{C}$ and to satisfies a functional equation which we normalize into the form

$$q_\Pi^{s/2}L(\Pi_\infty, s)L(\Pi, s) = w_\Pi q_\Pi^{(1-s)/2}\overline{L(\Pi_\infty, 1-\overline{s})L(\Pi, 1-\overline{s})},$$

where $|w_\Pi| = 1$ and $q_\Pi > 0$ is an integer called the conductor of Π. We recall (after Iwaniec/Sarnak [IS00]) that *the analytic conductor* of Π is the function of the complex variable s given by

$$C(\Pi, s) = q_\Pi \prod_{i=1}^{d}|s - \mu_{\Pi,i}|.$$

It is expected, and known in many cases, that the following *convexity bound* for the values of $L(\Pi, s)$ holds on the critical line $\Re es = 1/2$: for any $\varepsilon > 0$, one has

$$L(\Pi, s) \ll_{\varepsilon,d} C(\Pi, s)^{1/4+\varepsilon}.$$

This is known, in particular, when Π is an automorphic cuspidal representation of $\mathrm{GL}(n)$ over any number field, [Mol02]. The Lindelöf conjecture, which is a consequence of the GRH, asserts that in fact $L(\Pi, s) \ll_{\varepsilon,d} C(\Pi, s)^\varepsilon$. In many applications, however, it is sufficient to improve the convexity bound.

The subconvexity problem consists in improving the exponent $1/4$ to $1/4 - \delta$ for some positive absolute δ. In fact, for most applications it is sufficient to improve that exponant only with respect to one of the three type of parameters s, q_Π or $\prod_{i=1}^{d}(1 + |\mu_{\Pi,i}|)$; these variants of the subconvexity problem are called the s-aspect, the q-aspect (or *level*-aspect) and the ∞-aspect (or *eigenvalue*-aspect) respectively. See [Fri95, IS00] for an introduction to the subconvexity problem in this generality. In this lecture, we mainly discuss the recent progress made on the subconvexity problem in the q-aspect, although the other aspects are very interesting, both for applications and for conceptual reasons (see [Iwa92, Ivi01, Sar01, Blo04b, JM05]).

During the last decade, there has been considerable progress on the subconvexity problem for L-functions associated to GL_1 and GL_2 automorphic forms. For the level aspect the current situation is the following

Theorem 6. *Let F be a fixed number field and π_2 be a* fixed *cuspidal automorphic representation of $GL_2(\mathbf{A}_F)$. Let χ_1, π_1 denote respectively a $GL_1(\mathbf{A}_F)$-automorphic representation (i.e. a grossencharacter) , a $GL_2(\mathbf{A}_F)$-automorphic representation and let $\mathfrak{q}_1$ denote either the conductor of χ_1 or π_1 and $q_1 = N_{F/\mathbf{Q}}(\mathfrak{q}_1)$. There exists an absolute $\delta > 0$ (independent of χ_1, π_1, π_2 and F) such that for $\Re es = 1/2$ one has*

$$L(\chi_1, s) \ll_s q_1^{1/4-\delta}, \tag{3.1}$$

$$L(\chi_1 \times \pi_2, s) \ll_{s,\pi_2,\chi_{1,\infty}} q_1^{1/2-\delta}, \tag{3.2}$$

$$L(\pi_1, s) \ll_{s,\pi_{1,\infty}} q_1^{1/4-\delta}, \tag{3.3}$$

$$L(\pi_1 \times \pi_2, s) \ll_{s,\pi_2,\pi_{1,\infty}} q_1^{1/2-\delta}. \tag{3.4}$$

Thus the subconvexity problem is solved in the q_1-aspect for all these L-functions.

- For $F = \mathbf{Q}$, the bound for Dirichlet L-functions (3.1) is due to Burgess (see also [CI00] for a very strong subconvex exponent when χ is quadratic). The bound for twisted L-function (3.2) is basically due to Duke/Friedlander/Iwaniec [DFI93] (see also [Byk96, BHM06] for the general bound over $\mathbf{Q}$ with a good subconvex exponent). The bound 3.3 is mainly due to a series of works by Duke/Friedlander/Iwaniec: [DFI94] for π_1 with trivial central character and [DFI97, DFI01, DFI02] for the much harder case of a central character of conductor q_1; it has been recently completed for π_1 with arbitrary central character in [BHM06b] building on the method of [Mic04]. The bound for Rankin/Selberg L-functions (3.4) for π having trivial central character is due to the first author, Kowalski and Vanderkam ([KMV02]) by generalizing the methods of [DFI94] and to the first author and Harcos for for π_1 with an arbitrary central character by a different approach which is discussed in Section 3.4 [Mic04, HM06].
- In the case of a number field of higher degree, the first general subconvex result is due to Cogdell-/Piatetski-Shapiro/Sarnak [Cog03]: it consists of (3.2) when F is a totally real field and $\pi_{2,\infty}$ is in the holomorphic discrete series (i.e. corresponds to a holomorphic Hilbert modular form). Recently, the second author developed a new method which we discuss in section 4 below and established, amongst other things, the bounds (3.1), (3.2), (3.3) and (3.4) for F an arbitrary number field, π_2 fixed but arbitrary and π_1 with a trivial central character [Ven05]. Eventually the authors combined their respective methods from [Mic04] and [Ven05] to obtain (3.3) and (3.4) for π_1 with an arbitrary central character.

3.1. The amplification method and the shifted convolution problem. Arguably, the most successful approach to subconvexity in the q-aspect is via the method of moments or more precisely via its variant, the *amplification method*. For the sake of completness we briefly recall the mechanism and refer to [Fri95] and [Iwa99] for the philosophy underlying this method.

Given Π_1 and a (well choosen) family of automorphic objects $\mathcal{F} = \{\Pi\}$ containing Π_1, the amplification method builds on the possibility to obtain a bound for the amplified k-th moment of the $\{L(\Pi, s),\ \Pi \in \mathcal{F}\}$, $\Re es = 1/2$, of the form

$$\sum_{\Pi\in\mathcal{F}} |L(\Pi, s)|^k |\sum_{\ell\leqslant L} \lambda_\Pi(\ell)a_\ell|^2 \ll_\varepsilon |\mathcal{F}|^{1+\varepsilon} \sum_{\ell\leqslant L} |a_\ell|^2. \tag{3.5}$$

for any $\varepsilon > 0$, where the $(a_\ell)_{\ell\leqslant L}$ are *a priori* arbitrary complex coefficients and where L is some positive power of $|\mathcal{F}|$. Such a bound is expected if L is sufficiently small compared with $|\mathcal{F}|$, since the individual bound $|L(\Pi, s)|^k \ll_\varepsilon |\mathcal{F}|^\varepsilon$ would follow from the Generalized Riemann Hypothesis and the estimate

$$\sum_{\Pi\in\mathcal{F}} |\sum_{\ell\leqslant L} \lambda_\Pi(\ell)a_\ell|^2 = |\mathcal{F}|(1 + o(1)) \sum_{\ell\leqslant L} |a_\ell|^2$$

should be a manifestation of the *quasi-orthogonlity* of the $\{(\lambda_\Pi(\ell))_{\ell\leqslant L}\}_{\Pi\in\mathcal{F}}$ which is a frequent theme in harmonic analysis.

Assuming (3.5), one deduces that

$$L(\Pi_1, s) \ll_\varepsilon \Big(\frac{\sum_{\ell\leqslant L} |a_\ell|^2}{|\sum_{\ell\leqslant L} \lambda_{\Pi_1}(\ell)a_\ell|^2}\Big)^{1/k} |\mathcal{F}|^{1/k+\varepsilon} \ll_\varepsilon \Big(\frac{\sum_{\ell\leqslant L} |a_\ell|^2}{|\sum_{\ell\leqslant L} \lambda_{\Pi_1}(\ell)a_\ell|^2}\Big)^{1/k} C(\Pi_1, s)^{1/4+\varepsilon}$$

if $\mathcal{F}$ is such that $|\mathcal{F}| \ll C(\Pi_1, s)^{k/4}$. Often it is possible to choose the coefficient $(a_\ell)_{\ell\leqslant L}$ (depending on Π_1) so that

$$\frac{\sum_{\ell\leqslant L} |a_\ell|^2}{\left|\sum_{\ell\leqslant L} \lambda_{\Pi_1}(\ell)a_\ell\right|^2} \ll L^{-\alpha}$$

for some $\alpha > 0$; eventually this yields a subconvex bound.

In fact, all the subconvex bounds presented in Theorem 6 can be obtained by considering for $L(\Pi, s)$ an L-function of *Rankin/Selberg* type, i.e. either of the form $L(\chi_1 \times \pi_2, s)$ or of the form $L(\pi_1 \times \pi_2, s)$ with π_2 a fixed (not necessarily cuspidal) GL_2-automorphic representation . The families $\mathcal{F}$ considered are then essentially of the form $\{\chi \times \pi_2,\ q_\chi = q_1\}$ or $\{\pi \times \pi_2,\ q_\pi = q_1,\ \omega_\pi = \omega_{\pi_1}\}$ and the bound (3.5) is achieved for the second moment ($k = 2$).

The next step is to analyze effectively the lefthand side of (3.5) and in particlular to have a manageable expression for $L(\Pi, s)$ for s on the critical line. The traditional method to do so is to apply an *approximate* *functionnal equation* technique which expresses $L(\Pi, s)$ essentially as a partial sum of the form

$$\Sigma(\Pi) := \sum_{n\geqslant 1} \frac{\lambda_\Pi(n)}{n^s} W(\frac{n}{\sqrt{q_\Pi}})$$

with W a rapidly decreasing function (which depends on s and on Π_∞). In the context of Theorem 6 the second amplified moment of the families of partial sums $\{\Sigma(\chi \times \pi_2),\ q_\chi = q_1 =: q\}$ or $\{\Sigma(\pi \times \pi_2),\ q_\pi = q_1 =: q,\ \omega_\pi = \omega_{\pi_1}\}$ are computed and transformed by spectral methods. These involve, in particular, the orthogonality relations for characters and the Kuznetsov-Petersson formula. These computations reduce the subconvex estimates to the problem of estimating non-trivially sums of the form

$$\Sigma_\pm(\varphi_2, \ell_1, \ell_2, h) := \sum_{\ell_1 m \pm \ell_2 n = h} \overline{\rho_{\varphi_2}(m)} \rho_{\varphi_2}(n) \mathcal{W}(\frac{m}{q}, \frac{n}{q}), \tag{3.6}$$

the trivial bound being $\ll_{\varphi_2} q^{1+o(1)}$; here $h = O(q)$ is a non-zero integer, $\rho_{\varphi_2}(n)$ denote the n-th Fourier coefficients of some automorphic form φ_2 in the representation space of π_2, $\mathcal{W}(x, y)$ is a rapidly decreasing function and $\ell_1, \ell_2 \leqslant L$ are the parameters occuring as indices of the amplifier $(a_\ell)_{\ell\leqslant L}$. These sums are classical in analytic number theory and are called *shifted convolution sums*; the problem of estimating them non-trivially for various ranges of h, m, n is called a *shifted convolution problem.*

Observe that when $h = 0$ the sum in (3.6) is is a partial sum of Rankin/Selberg type and can be analyzed by means of the analytic properties of the Rankin/Selberg convolution L-function. The non-homogeneous case $h \neq 0$ is different. Historically, the shifted convolution problem already occured in the work of Kloosterman on the number of representations of an integer n by the quadratic form $a_1.x^2 + a_2.y^2 + a_3.z^2 + a_4.t^2$, and also in Ingham's work on the additive divisor problem. In Kloosterman's case φ_2 is a theta-series of weight 1, whereas in Ingham's case φ_2 is the standard non-holomorphic Eisenstein series.

3.2. Shifted convolutions via the circle method.

In order to solve a shifted convolution type problem, one needs an analytically manageable expression of the linear constraint $\ell_1 m \pm \ell_2 n = h$; one is to suitably decompose the integral

$$\delta_{\ell_1 m \pm \ell_2 n - h = 0} = \int_{\mathbf{R}/\mathbf{Z}} \exp(2\pi\imath(\ell_1 m \pm \ell_2 n - h)\alpha) d\alpha,$$

and there are several methods to achieve this; the first possibility in this context was Kloosterman's refinement of the circle method; other possibilities are the Δ-symbol method, used in [DFI93] and [DFI94] to prove some cases of (3.2) and (3.3) or Jutila's method of overleaping intervals which is particularly flexible [Jut99, Har03]. These methods provide an expression of the above integral into weighted sums of Ramanujan type sums of the form

$$\sum_{\substack{a \mod c \\ (a,c)=1}} e(\frac{(\ell_1 m \pm \ell_2 n - h).a}{c})$$

for c ranging over relatively small moduli. Such decomposition makes it possible to essentially "separate" the variable m from n and to reduce $\Sigma(\varphi_2, \ell_1, \ell_2, h)$ to sums over moduli c on additively twisted sums of Fourier coefficients

$$\sum_c \cdots \sum_{\substack{a \mod c \\ (a,c)=1}} e(\frac{-ha}{c}) \times \Big(\sum_m \overline{\rho_{\varphi_2}(m)} e(\frac{\ell_1 ma}{c}) \mathcal{W}(\frac{m}{q})\Big) \Big(\sum_n \rho_{\varphi_2}(n) e(\pm\frac{\ell_2 na}{c}) \mathcal{W}(\frac{n}{q})\Big).$$

The independent m and n-sums are then transformed via the Voronoïsummation formula with the effect of replacing the test functions $\mathcal{W}(\frac{\cdot}{q})$ by some Bessel transform and the additive shift $e\big(\pm\frac{\ell_2 a.}{c}\big)$ by $e\big(- \pm \frac{\overline{\ell_2 a}.}{c}\big)$ where $\overline{a}$ denote the multiplicative inverse of $a \mod c$. After these transformations and after averaging over $a \mod c$ the sum $\Sigma_\pm(\varphi_2, \ell_1, \ell_2, h)$ takes essentially the following form (possibly up to a main term which occurs if φ_2 is an Eisenstein series [DFI94])

$$MT_\pm(\varphi_2, \ell_1, \ell_2, h) + \sum_{c\equiv 0(\ell_1\ell_2 q_{\pi_0})} \sum_{h'} \Big(\sum_{\mp \ell_1 n - \ell_2 m = h'} \alpha_m \overline{\rho_{\varphi_2}(m)} \beta_n \rho_{\varphi_2}(n)\Big) \mathrm{Kl}(-h, h'; c) \mathcal{V}(h, h'; c) \tag{3.7}$$

where $MT_\pm(\varphi_2,\ell_1,\ell_2,h)$ is non-zero only if φ_2 is an *Eisenstein series* (in which case it is a *main term* of size $\approx_{\ell_1,\ell_2,\varphi_2} q^{1+o(1)}$), α_m,β_n are smooth coefficients, $\mathrm{Kl}(-h,h';c)$ is a Kloosterman sum and $\mathcal{V}$ is a smooth function. Eventually, Weil's bound for Kloosterman sums

$$\mathrm{Kl}(-h,h';c)\ll (h,h',c)^{1/2}c^{1/2+o(1))}$$

gives the formula

$$\Sigma_\pm(\varphi_2,\ell_1,\ell_2,h)=MT_\pm(\varphi_2,\ell_1,\ell_2,h)+O_{\varphi_2,\varepsilon}\Big((\ell_1\ell_2)^Aq^{3/4+\varepsilon}\Big) \tag{3.8}$$

for some absolute constant A (in fact any non-trivial bounds for Kloosterman sums, like Kloosterman's original bound, is sufficient to obtain an asymptotic formula with 3/4 replaced by some exponent <1). Finally, from (3.8) one can deduce (3.2), (3.3), (3.4) when π has trivial central character although the derivation may be quite delicate if φ_2 is an Eisenstein series (cf. [DFI94] and see also [KMV00]).

3.3. Shifted convolutions and spectral theory of automorphic forms. In [Sar01], Sarnak, inspired by ideas of Selberg, developed a purely spectral approach to the shifted convolution sums (3.6) (previously some special cases have been treated by others, for instance by A. Good). This method, which at present has been entirely worked out when φ_2 is a classical holomorphic cuspform (say of weight $k\geqslant 2$ and level q_2), is based on the analytic properties of the Dirichlet series

$$D(\varphi_2,\ell_1,\ell_2,h,s)=\sum_{\substack{m,n\geqslant 1\\ \ell_1m-\ell_2n=h}}\frac{\overline{\rho_{\varphi_2}(m)}\rho_{\varphi_2}(n)}{(\ell_1m+\ell_2n)^s}\Big(\frac{\sqrt{\ell_1\ell_2mn}}{\ell_1m+\ell_2n}\Big)^{k-1}.$$

Note that for $h=0$ this series is essentially a Rankin/Selberg L-function. As in the Rankin/Selberg case, the analytic properties of D follows from an appropriate integral representation in the form of a triple product integral; however, for $h\neq 0$ one needs to replace the Eisenstein series by a Poincaré series. Precisely, one has $D(s)=(2\pi)^{s+k-1}(\ell_1\ell_2)^{1/2}\Gamma^{-1}(s+k-1)I(s)$ with

$$\begin{aligned}I(\varphi_2,\ell_1,\ell_2,h,s) &:= \langle(\ell_1y)^{k/2}\varphi_2(\ell_1z).(\ell_2y)^{k/2}\overline{\varphi_2}(\ell_2z),P_h(z,s)\rangle\\ &=\int_{\Gamma_0(q_2\ell_1\ell_2)\backslash\mathbb{H}}(\ell_1y)^{k/2}\overline{\varphi_2}(\ell_1z).(\ell_2y)^{k/2}\varphi_2(\ell_2z)P_h(z,s)\frac{dxdy}{y^2}\end{aligned}$$

where $P_h(z,s)$ the non-holomorphic Poincaré series of weight 0:

$$P_h(z,s)=\sum_{\gamma\in\Gamma_0(q_2\ell_1\ell_2)_\infty\backslash\Gamma_0(q_2\ell_1\ell_2)}(\Im m\gamma z)^se(h\Re e\gamma z).$$

The analytic continuation for D follows from that of $P_h(.,s)$; in particular, from its spectral expansion one deduce that the latter is absolutely convergent for $\Re es>1$ and has holomorphic continuation in the half-plane $\Re es>1/2+\theta$ where θ measures the quality of available results towards the Ramanujan/Petersson conjecture:

Hypothesis H_θ**.** *For any cuspidal automorphic form π on $GL_2(\mathbf{Q})\backslash GL_2(\mathbf{A_Q})$ with local Hecke parameters $\{\alpha_{\pi,i}(p),\ i=1,2\}$ for $p<\infty$ and $\{\mu_{\pi,i},\ i=1,2\}$ one has the bounds*

$$|\alpha_{\pi,i}(p)|\leqslant p^\theta,\ i=1,2,$$

$$|\Re e\mu_{\pi,i}|\leqslant\theta,\ i=1,2,$$

provided π_p, π_∞ are unramified, respectively.

Remark 3.1. Hypothesis H_θ is known for $\theta>3/26$ thanks to the works of Kim and Shahidi [KS02, Kim03].

A bound for $D(s)$ in a non-trivial domain, is deduced from the spectral expansion of the inner product $I(s)$ over an suitable orthonormal basis of Maass forms, $\{\psi\}$ say, and of Eisenstein series of weight 0 and level $\ell_1\ell_2q_0$: one has

$$\sum_\psi\langle(\ell_1y)^{k/2}\varphi_2(\ell_1z).(\ell_2y)^{k/2}\overline{\varphi_2(\ell_2z)},\psi\rangle\langle\psi,P_h(z,s)\rangle+\text{Eisenstein spectrum contr.} \tag{3.9}$$

For ψ in the cuspidal basis, let it_ψ denote the archimedean parameter $\mu_{\pi,1}$ of the representation π containing ψ; the second inner product $\langle\psi, P_h(z,s)\rangle$ equals the Fourier coefficient of ψ, $\overline{\rho_\psi(-h)}$ times a factor bounded by $(1+|t_\psi|)^B e^{\frac{\pi}{2}|t_\psi|}$. The Fourier coefficient $\overline{\rho_\psi(-h)}$ is bounded by $O(|h|^{\theta+o(1)})$ by Hypothesis H_θ at the non-archimedean places. The problem now, as was pointed out by Selberg, is to have a bound for the triple product integral $\langle(\ell_1 y)^{k/2}\varphi_2(\ell_1 z).(\ell_2 y)^{k/2}\overline{\varphi_2}(\ell_2 z), \psi\rangle$ which exhibits an exponential decay for the form $O((1+|t_\psi|)^C e^{-\frac{\pi}{2}|t_\psi|})$, so as to compensate the exponential growth of $\langle\psi, P_h(z,s)\rangle$. In this generality, this exponential decay property for triple product was achieved by Sarnak in [Sar94]; later, a representation theoretic version of Sarnak's arguments as well as some improvments were given by Bernstein/Reznikov [BR99]. The final consequence of these bounds is the following estimate

$$\Sigma_-(\varphi_2,\ell_1,\ell_2,h) = O_{\varphi_2,\varepsilon}\Big((\ell_1\ell_2)^A q^{1/2+\theta+\varepsilon}\Big) \tag{3.10}$$

This approach is important for several reasons:

- It ties more closely the subconvexity problem for GL_2 L-functions – a problem whose origin lies in analytic number theory – to the Ramanujan/Petersson conjecture for $GL_2(\mathbf{A_Q})$; or, in other words, to the spectral gap property which is a classical problem in the harmonic analysis of groups;
- it gives an hint that automophic period integrals might be useful tool in the study of the subconvexity problem: this will be largely confirmed in section 4.
- This approach is sufficiently smooth that it can be extended to number fields of higher degree: a few years ago, Cogdell/Piatetski-Shapiro/Sarnak used the amplification method in conjunction with this approach to obtain (3.2) when F is totally real and π_∞ is a holomorphic discrete series (see [Cog03] for an account). (Note that no suitable version of the Kloosterman refinement has been carried out over a number field.)

A challenging point is that, so far this approach has not been carried out succesfully for the case where φ_2 is non-holomorphic (or at least not in the ranges required to obtain the bounds (3.2), (3.3) or (3.4) when the fixed representation π_2 is not a holomorphic discrete series): apparently, the main problem is to find an appropriate test vector $\varphi_2\in V_{\pi_2}$ so that the shifted convolution sum can be represented in terms of a triple product integral $I(\varphi_2,\ell_1,\ell_2,h)$. See [Mot04] for a discussion on this issue and some hints.

Remark 3.2. These two approaches of the shifted convolution problem are closely related. This can be seen already at a superficial level by remarking that Weil's bound for Kloosterman sums yield the saving $q^{3/4+\varepsilon}$ in (3.8) which is precisely the saving one gets from Hypothesis $H_{1/4}$ in (3.10); moreover $H_{1/4}$ (a.k.a the Selberg/Gelbart/Jacquet bound) can be obtained by applying Weil's bound to the Kloosterman sums occuring in the Kuznetzov/Petersson formula.

One can push this coincidence further, by applying, in (3.7) the Kuznetzov/Petersson formula *backwards* in order to transform the sums of Kloosterman sums into sums of Fourier coefficients of Maass forms:

$$(3.7) = \sum_\psi\sum_{h'}\Big(\sum_{\mp\ell_1 n-\ell_2 m=h'}\alpha_m\overline{\rho_{\varphi_2}(m)}\beta_n\rho_{\varphi_2}(n)\Big)\overline{\rho}_\psi(-h)\rho_\psi(h')\tilde{\mathcal{V}}(h,h',it_\psi)$$
$$+\text{ Discrete Spectrum} + \text{Eisenstein spectrum}. \tag{3.11}$$

Thus, we have realized the spectral expansion of the shifted convolution sum $\Sigma_\pm(\varphi_2,\ell_1,\ell_2,h)$ in a way similar to that obtained in (3.9); from there, we may use again the full force of spectral theory. This may look like a rather circuituous path to obtain the spectral expansion; this method however has some technical advantage over the method discussed in section 3.3: it works even if φ_2 is a Maass form, without the need to find appropriate test vector or to obtain exponential decay for triple product integrals ! The spectral decomposition (3.11) will be very useful in the next section.

3.4. The case of a varying central character.

The methods discussed so far are sufficient to establish (3.2), (3.3), (3.4) as long as the conductor of the central character, ω_1 say, of π_1 is significantly smaller than q_1. The case of a varying central character reveals new interesting features which we discuss here. To simplify, we consider the extremal (in a sense hardest) case where both conductors are equal $q_{\omega_1} = q_1 =: q$.

3.4.1. Subconvexity of Hecke L-function via bilinear Kloosterman fractions. As usual for the subconvexity problem, the first result is due to Duke/Friedlander/Iwaniec for the case (3.3) [DFI01, DFI02]. As pointed out above, the problem of bounding $L(\pi_1, s)$ for $\Re es = 1/2$ may be formulated as the problem of bounding a Rankin/Selberg L-function

$$L(\pi_1 \times \pi_2, s) = L(\pi_1, s)^2$$

where $\pi_2 = 1 \boxplus 1$ is the representation corresponding to the the fully unramified Eisenstein series. Eventually, another approach was considered in [DFI01, DFI02], which comes from the identity

$$|L(\pi_1, s)|^2 = L(\pi_1 \times \overline{\chi}, 1/2)L(\tilde{\pi}_1 \times \chi, 1/2)$$

where $\tilde{\pi}_1$ is the contragredient and $\chi = \omega_1 |.|^{-it},\ t = \Im ms$. The amplification method applied to the family

$$\Big\{L(\pi \times \overline{\chi}, 1/2)L(\tilde{\pi} \times \chi, 1/2),\ q_\pi = q_1 := q, \omega_\pi = \omega_1\Big\}$$

yields in practice to shifted convolution sums of the form ([DFI01, DFI02]) of the form

$$\sum_{\ell_1 ad - \ell_2 bc = h} \overline{\chi}(a)\chi(c)\mathcal{W}(\frac{a}{q^{1/2}}, \frac{b}{q^{1/2}}, \frac{c}{q^{1/2}}, \frac{d}{q^{1/2}}),$$

with $h \approx q,\ h \equiv 0(q)$. The later is essentially a truncated version of the shifted convolution sums associated to the Eisenstein series $E(1, \chi)$ of the representation $1 \boxplus \chi$ (whose fourier coefficients are $\rho_{E(1,\chi)}(n) = \sum_{bc=n} \chi(c)$); the new feature by comparison with the previous shifted convolution problems is that the coefficients $\rho_{E(1,\chi)}(n)$ vary with q, which is essentially the range of the variables $m = ad$ and $n = bc$. Since χ has conductor q and a, c vary in ranges of size $\approx q^{1/2}$ one cannot really use the arithmetical structure of the weights $\overline{\chi}(a),\ \chi(c)$ so this shifted convolution problem is basically reduced to the non-trivial evaluation of a quite general sum:

$$\sum_{\ell_1 ad - \ell_2 bc = h} \alpha_a \gamma_c \mathcal{W}(\frac{a}{q^{1/2}}, \frac{b}{q^{1/2}}, \frac{c}{q^{1/2}}, \frac{d}{q^{1/2}}) = MT((\alpha_a)_a, (\gamma_c)_c, \ell_1, \ell_2, h) + O((\ell_1\ell_2)^A q^{1-\delta}) \tag{3.12}$$

for some $\delta > 0$ absolute and where $MT((\alpha_a)_a, (\delta_d)_d, \ell_1, \ell_2, h)$ denote a natural main term and with $(\alpha_a)_{a \sim q^{1/2}}$, $(\gamma_c)_{c \sim q^{1/2}}$ arbitrary complex numbers of modulus bounded by 1. Since the b variable is smooth, the condition $\ell_1 ad - \ell_2 bc = h$ is essentially equivalent to the congruence condition $\ell_1 ad \equiv h \mod c\ell_2$. One can then analyze this congruence by Poisson summation applied on the remaining smooth variable d which yields a sums of Kloosterman fractions of the shape

$$\sum_{\substack{a \sim A, c \sim C \\ (a,c)=1}} \alpha_a \gamma_c e(h\frac{\overline{a}}{c}),\ \text{for } h \neq 0 \text{ and}$$

where the values of a, c, h and $\alpha_a,\ \gamma_c$ may be different from the previous ones. In [DFI97] such sums are bounded non-trivially for any ranges A, C (the most crucial one being $A = C$).

A remarkable feature of this proof is that the bound is obtained from an non-trivial bound for the related sum

$$\sum_{a \sim A} \Big| \sum_{\substack{c \sim C \\ (a,c)=1}} \gamma_c e(h\frac{\overline{a}}{c}) \Big|^2$$

by an application of the amplification method in a very unexpected direction, namely by amplifying the trivial (!) multiplicative characters $\chi_{0,a}$ of modulus a in the family of sums

$$\{ \sum_{\substack{c \sim C \\ (a,c)=1}} \gamma_c \chi(c) e(h\frac{\overline{a}}{c}),\ \chi \bmod a,\ a \sim A \}.$$

Remark 3.3. One should note that (3.12) is quite a bit more general than what is needed to establish (3.3). In fact, estimates of that kind may be used in other context than subconvexity: for example, to establish Bombieri/Vinogradov type results. On the other hand, in the context of the subconvexity problem, this method uses the special shape of the Fourier coefficients of Eisenstein series and does not seem to generalize to Rankin/Selberg L-functions.

3.4.2. Subconvexity of Rankin/Selberg L-functions via subconvexity for twisted L-functions. The case of Rankin/Selberg L-functions over $\mathbf{Q}$, $L(\pi_1 \times \pi_2, s)$ when π_2 is essentially fixed and π_1 has a central character ω_1 of large conductor was treated in [Mic04] (when the fixed representation π_2 corresponds to a holomorphic form) and in [HM06] (in the general case). In the case of a varying central characters, the shifted convolution problem that needs to be resolved is one on average over h. More precisely subconvexity comes from an estimate for a sum of shifted convolution sums:

$$\Sigma_{\pm}(\varphi_2, \ell_1, \ell_2, \omega) := \sum_{0<|h|\ll q} \overline{\omega}(h)\Sigma_{\pm}(\varphi_2, \ell_1, \ell_2, h); \tag{3.13}$$

The goal, however, is not to achieve the shifted convolution problem on average over h but to prove rather more, namely

$$\sum_{0<|h|\ll q} \overline{\omega}(h)\Sigma_{\pm}(\varphi_2, \ell_1, \ell_2, h) \ll_{\varphi_2} (\ell_1\ell_2)^A q^{3/2-\delta} \tag{3.14}$$

for some $A, \delta > 0$ absolute. In particular even under the Ramanujan/Petersson conjecture (H_0), the individual bound (3.10) is "just" not sufficient; this means that one has to account for the averaging over the h variable.

This is achieved through the spectral decomposition of the shifted convolution sums (3.11): plugging this formula into (3.13) one obtains a sum over the orthonormal basis $\{\psi\}$ of sums of the form

$$\sum_{0<|h|\ll q} \overline{\omega}(h)\rho_{\psi}(-h)$$

if ψ belong to the space V_τ of some automorphic representation, the later sums are partial sums associated to the twisted L-function $L(\tau \times \overline{\omega}, s)$. In that case, the subconvexity bound for twisted L-functions (3.2) is exactly sufficient to give (3.14).

Remark 3.4. Hence the subconvexity bound for an L-functions of degree 4 has been reduced to a collection of subconvex bounds for L-functions of automorphic forms of small level twisted by the original central character ω ! This surprising phenomenon is better explained via the approach described in the next section. It would be interesting if this reduction of degree persists in higher rank.

4. Subconvexity of L-functions via periods of automorphic forms

4.1. The various perspectives on an L-function.

From the perspective of analytic number theory, the definition of L-function might be "an analytic function sharing the key features of $\zeta(s)$: analytic continuation, functional equation, Euler product."

However, there are various "incarnations" of L-functions attached to automorphic forms; although equivalent, different features become apparent in different incarnations. For instance, one can define L-functions via constant terms of Eisenstein series (the Langlands-Shahidi method), via periods of automorphic forms (the theory of integral representations, which begins with the work of Hecke, or indeed already with Riemann), or via a Dirichlet series (which is often taken as their defining property).

To give a highly simplified illustration of the difference in perspective these provide, consider the following standard analytic results:

1. Nonvanishing on the line $\Re(s) = 1$;
2. Analytic continuation to the whole complex plane;
3. Convexity bound.

The first is (perhaps) most "transparent" from the Langlands-Shahidi perspective; the second is often carried out via integral representations, i.e. by periods; the only generally known proof of the third is via Dirichlet series.

Thus far in this article, we have discussed the subconvexity from the perspective of Dirichlet series. In other words, writing

$$L(s) = \sum_{n=1}^{\infty} a_n n^{-s}$$

we ask: What properties of the sequence (a_n) give rise to subconvex estimates for the L-function?

Relatively recently, the subconvexity question has also been succesfully approached via the "period" perspective on L-functions. The first such result (in the eigenvalue aspect) was given by Bernstein-Reznikov [BR04], [BR05], and a little later a result in the level aspect was given by Venkatesh [Ven05]. These two methods seem to be quite distinct.

To phrase in the context of the present text: thus far we have studied periods (e.g. (2.4)) via reducing them to questions about L-functions (via (2.2)) and then proving subconvexity for the L-function. The viewpoint of [BR05], [Ven05] reverses this general process, although for different families of L-functions. These papers deduce subconvexity from a geometric study of the period.

These approaches are closely related to existing work, but in many cases the period perspective allows certain conceptual simplifications and it brings together harmonic analysis and ideas from dynamics. Such conceptual simplifications are particularly of value in passing from $\mathbf{Q}$ to a general number field; so far, with the exception of the result of Cogdell/Piatetski-Shapiro/Sarnak, all the results in Theorem 6 in the case $F \neq \mathbf{Q}$ are proven via the period approach.

On the other hand, it might be noted that a slight drawback to the period approach to subconvexity is that, especially for automorphic representations with complicated ramification, one must face the difficulty of choosing appropriate test vectors. More generally, the proof of the precise formulas relating periods to L-functions often involve formidable technical difficulties.

4.2. The work of Bernstein-Reznikov and Venkatesh: Triple product period and triple product L-function.

At present, all known results towards the subconvexity of triple product L-functions $L(\pi_1 \times \pi_2 \times \pi_3, \frac{1}{2})$ arise from the "period" perspective.

The period of interest is

$$\int_{\mathrm{PGL}_2(\mathbf{Q})\backslash \mathrm{PGL}_2(\mathbf{A})} \varphi_1(g)\varphi_2(g)\varphi_3(g)dg$$

where $\varphi_i \in \pi_i$, and each π_i is an automorphic cuspidal representation of GL_2. It is expected that this period, and the variants when GL_2 is replaced by the multiplicative group of a quaternion algebra, is related to the central value of the triple product L-function $L(\pi_1 \times \pi_2 \times \pi_3, \frac{1}{2})$, see [HK91]. A precise relationship has been computed for the case of Maass forms at full level in [Wat06]; indeed, the following formula is established:

$$\left| \int_{\mathrm{SL}_2(\mathbf{Z})\backslash \mathbb{H}} \varphi_1\varphi_2\varphi_3 d\mu \right|^2 = \frac{\Lambda(\varphi_1 \times \varphi_2 \times \varphi_3, 1/2)}{\Lambda(\wedge^2\varphi_1, 1)\Lambda(\wedge^2\varphi_2, 1)\Lambda(\wedge^2\varphi_3, 1)} \tag{4.1}$$

where Λ denotes completed L-function and $d\mu$ is a suitable multiple of $\frac{dx\,dy}{y^2}$.

4.2.1. The eigenvalue aspect.

Let φ_1, φ_2 be fixed Hecke-Maass forms on $\mathrm{SL}_2(\mathbf{Z})\backslash\mathbb{H}$, and φ_λ a Hecke-Maass form of eigenvalue $\lambda = 1/4 + r^2$. Let $\pi_1, \pi_2, \pi_\lambda$ be the associated cuspidal automorphic representations. In [BR05] the following bound is established:

$$L(\pi_1 \times \pi_2 \times \pi_\lambda, \frac{1}{2}) \ll_{\epsilon, \pi_1, \pi_2} r^{5/3+\epsilon} \tag{4.2}$$

The convexity bound for the left-hand side is $r^{2+\epsilon}$.

Let Γ be a cocompact subgroup of $\mathrm{SL}_2(\mathbf{R})$, let $\mathbb{H}$ be the upper half-plane, let φ_1, φ_2 be fixed eigenfunctions of the Laplacian on $\Gamma\backslash\mathbb{H}$ and φ_λ an eigenfunction with eigenvalue $\lambda := 1/4 + r^2$. In the paper [BR05], Bernstein and Reznikov establish the following bound.

$$r^2 e^{\pi r/2} \left| \int_{\Gamma\backslash\mathbb{H}} \varphi_1(z)\varphi_2(z)\varphi_\lambda(z) d\mu_z \right|^2 \ll_\epsilon r^{5/3+\epsilon} \tag{4.3}$$

They note that their method is valid also when $\varphi_1, \varphi_2, \varphi_\lambda$ are cusp forms and Γ is noncompact. In the case when $\Gamma = \mathrm{SL}_2(\mathbf{Z})$ and $\varphi_1, \varphi_2, \varphi_\lambda$ are Hecke-Maass forms associated to automorphic representations $\pi_1, \pi_2, \pi_\lambda$ respectively, the bound (4.3) translates, via (4.1), to the subconvex bound (4.2).

We remark only on one key ingredient: for $G = \mathrm{PGL}_2(\mathbf{R})$ and three irreducible, admissible representations π_1, π_2, π_3 the space of G-invariant functionals on $\pi_1 \otimes \pi_2 \otimes \pi_3$ is *at most* one dimensional. An consequence of this is that one may study the period (4.3) by replacing $\varphi_1, \varphi_2, \varphi_\lambda$ by any other vectors $v_1 \in \pi_1, v_2 \in \pi_2, v_\lambda \in \pi_\lambda$. An important observation of Bernstein and Reznikov is that the spherical vectors $\varphi_1, \varphi_2, \varphi_\lambda$ are, in a sense, "poor" test vectors for the trilinear functional defined by (4.3); by switching to more suitable v_1, v_2, v_λ one can eliminate, for instance, the exponential decay implicit in (4.3). Further ideas are needed, however, to obtain subconvexity.

It is worth emphasizing that, by contrast to what will be discussed in the following section, the following two aspects of this method:

1. This method does *not* use any information about the spectral gap.
2. It does not use Hecke operators: it is purely *local*, i.e. using only properties of the real group $\mathrm{PGL}_2(\mathbf{R})$ rather than the group $\mathrm{PGL}_2(\mathbf{A}_\mathbf{Q})$.

4.2.2. The level aspect. Let F be a number field. Let π_2, π_3 be fixed automorphic cuspidal representations on $\mathrm{PGL}_2(\mathbf{A}_F)$ – say with coprime conductor – and let π_1 be a third automorphic cuspidal representation with conductor $\mathfrak{q}$, a prime ideal of F. In [Ven05] it is established that

$$L(\pi_1 \times \pi_2 \times \pi_3, \frac{1}{2}) \ll_{\pi_{1,\infty}, \pi_2, \pi_3} \mathrm{N}(\mathfrak{q})^{1-\frac{1}{13}} \tag{4.4}$$

contingent on a suitable version of (4.1) when the level of one factor varies.[5] The convexity bound for the left-hand side is $\mathrm{N}(\mathfrak{q})^{1+\varepsilon}$.

Remark 4.1. In [Ven05], a form of (4.4) is proved when π_2 and/or π_3 are Eisenstein series: in that case, (4.1) corresponds to simple computations in the Rankin-Selberg method and so is unconditional. In particular, this yields the bounds (3.3) and (3.4) for π_1 with trivial central character.

We explain the period bound underlying (4.4) in the classical case $F = \mathbf{Q}$. In this context, one can see the role of equidistribution clearly; in the next Section 4.3, we explain a more general period bound using more explicit spectral methods, which have the disadvantage of concealing the underlying equidistribution results.

Let $\Gamma_0(p) = \{\begin{pmatrix} a & b \\ c & d \end{pmatrix} : p|c\}$ and let $Y_0(p) = \Gamma_0(p)\backslash\mathbb{H}$. Then the map $z \mapsto (z, pz)$ descends to a map

$$\iota : Y_0(p) \to Y_0(1) \times Y_0(1).$$

Let φ_2, φ_3 be Hecke-Maass cusp forms for $Y_0(1)$ and φ_1 a Hecke-Maass cusp form for $Y_0(p)$. We equip $Y_0(1)$ with the Poincaré measure $d\mu_P$ defined in Theorem 2, which is a probability measure; similarly, we equip $Y_0(p)$ with the probability measure induced from the measure $\frac{1}{[\Gamma_0(1):\Gamma_0(p)]}\frac{3}{\pi}\frac{dx\,dy}{y^2}$ on $\mathbb{H}$.

In the paper [Ven05], the second author establishes the bound

$$\left|\int_{Y_0(p)} \iota^*(\varphi_2 \times \varphi_3) \cdot \varphi_1\right| \ll p^{-\frac{1}{26}} \|\varphi_1\|_{L^2(Y_0(p))} \|\varphi_2\|_{L^4(Y_0(1))} \|\varphi_3\|_{L^4(Y_0(1))}$$

where ι^* denotes the pullback on functions, i.e. $\iota^*(\varphi_2 \times \varphi_3)(z) = \varphi_2(z)\varphi_3(pz)$. This estimate yields (4.4) – contingent, as remarked above, on a suitable generalization of (4.1).

The proof uses a certain "geometrization" of the amplification method of Friedlander/Iwaniec, together with ideas of equidistribution and mixing from ergodic theory. Here we just indicate the rough idea; it is applicable in various other settings, as is explained in [Ven05], and a more abstract explanation of it as well as applications to other families of L-functions is given in that paper.

Let T_ℓ be the ℓth Hecke operator on $Y_0(p)$, where $(\ell, p) = 1$. Let $\lambda_{\pi_1}(\ell)$ be the eigenvalue of T_ℓ acting on φ_1. We take a suitable linear combination

$$\mathbb{T} := \sum_{\ell \in \mathcal{L}} a_\ell T_\ell \tag{4.5}$$

[5]This has not appeared in the literature to our knowledge, except in the case where one of the π_j are Eisenstein; however, it should amount to a routine though very involved computation of p-adic integrals.

of the Hecke operators so that the eigenvalue $\Lambda_{\pi_1} = \sum_{\ell \in \mathcal{L}} a_\ell \lambda_{\pi_1}(\ell)$ is "large." We regard $\mathbb{T}$ as an operator on functions on $Y_0(p)$; it is self-adjoint, and so

$$\Lambda_{\pi_1} \int_{Y_0(p)} \iota^*(\varphi_2 \times \varphi_3) \cdot \varphi_1 = \int_{Y_0(p)} \iota^*(\varphi_2 \times \varphi_3) \cdot \mathbb{T}\varphi_1$$
$$= \int_{Y_0(p)} \mathbb{T}(\iota^*(\varphi_2 \times \varphi_3)) \cdot \varphi_1 \leqslant \|\varphi_1\|_{L^2(Y_0(p))} \|\mathbb{T}(\iota^*(\varphi_2 \times \varphi_3))\|_{L^2(Y_0(p))} \quad (4.6)$$

Thus far this rather closely resembles the amplification method.

However, at this point, one proceeds to "directly" bound $\int_{Y_0(p)} |\mathbb{T}(\iota^*(\varphi_2 \times \varphi_3))|^2$. To see how this might work, consider first the corresponding integral without $\mathbb{T}$, i.e. $\int_{Y_0(p)} |\iota^*(\varphi_2 \times \varphi_3)|^2$.

It is known that $Y_0(p) \subset Y_0(1) \times Y_0(1)$ becomes *equidistributed* as $p \to \infty$; this is the equidistribution of Hecke points, which was known "classically" in this case, and which was quantified in great generality (higher rank groups) in [COU01]. This equidistribution shows at once that

$$\int_{Y_0(p)} |\iota^*(\varphi_2 \times \varphi_3)|^2 \to \int_{Y_0(1)} |\varphi_2|^2 \int_{Y_0(1)} |\varphi_3|^2, \quad (4.7)$$

as $p \to \infty$, with a quantitative rate of decay. Translating to L-functions via (4.1), we see that (4.7) amounts to a mean value theorem, and indeed it seems often that one can obtain mean value theorems for L-functions from equidistribution results; another example was seen in Sec. 2.4.

In order to incorporate the $\mathbb{T}$ into this argument, one needs to use a slight refinement of the equidistribution of $Y_0(p) \subset Y_0(1) \times Y_0(1)$. This refinement is not difficult but notationally is best expressed adelically; we discuss it briefly in the next section. In classical terms, the key point is the following. If F, G are continuous function on $Y_0(1) \times Y_0(1)$, then

$$\int_{Y_0(p)} T_\ell F.G \sim \int_{Y_0(1)\times Y_0(1)} ((T_\ell \times T_\ell)F).G$$

whenever ℓ is small compared to p. Note on the left-hand side the Hecke operator T_ℓ is acting on functions on $Y_0(p)$, whereas on the right-hand side it is acting on functions on $Y_0(1)$; in both cases it is normalized to act as the identity on the constants.

Both equidistribution and mixing properties of Hecke operators reduce to a bound towards Ramanujan on $\mathrm{GL}_2(\mathbf{A}_{\mathbf{Q}})$; one needs in fact a bound H_θ with $\theta < 1/4$, in the notation of Sec. 3.3.

4.3. Central character. In this section, we return to Sec. 3.4 and explain, via periods, the bound (3.4). In particular, this sheds light on the "reason" for the reduction to a lower rank subconvexity problem that was encountered in that Section. The content of this section is carried out in detail in [MV].

Let π_1, π_2 be automorphic cuspidal representations of $\mathrm{GL}_2(\mathbf{A}_F)$. Let ω be the central character of π_1. For simplicity, we restrict ourselves to the case where π_2, the "fixed" form, has level 1 and trivial central character; and where "all the ramification of π_1 comes from the central character," i.e. π_1 and ω have the same conductor $\mathfrak{q}$. However, as we remark at the end, the process becomes strictly simpler in the general case.

Let us first give a very approximate "philosophical" overview of the proof. There is an identity between mean values of L-functions of the following type:

$$\sum_{\pi_1} L(\pi_1 \times \pi_2, \frac{1}{2}) \longleftrightarrow \sum_{\tau \text{ level } 1} L(\tau, \frac{1}{2}) L(\tau \times \omega, \frac{1}{2}) \quad (4.8)$$

where the left-hand summation is over π_1 of central character ω and conductor $\mathfrak{q}$, whereas the right-hand summation is over automorphic representations τ of trivial central character and level 1. It includes the trivial (one-dimensional) automorphic representation, which is in fact the dominant term and actually needs to be handled by regularization. [6]

By means of a suitable amplifier, one can restrict the left-hand summation to pick out a given π_1. When one does this, the necessary bounds on the right-hand side follow from two different inputs:

[6]Note that "morally", when τ is trivial, the L-function $L(\tau, s)L(\tau \times \omega, s) = \zeta(s+1/2)\zeta(s-1/2)L(\omega, s-1/2)L(\omega, s+1/2)$. Thus we obtain a pole at $s = 1/2$.

1. Subconvexity for $L(\tau \times \omega, \frac{1}{2})$ (in the aspect where ω varies), to handle the nontrivial τ.
2. A bound for decay of matrix coefficients of p-adic groups, to handle the case of τ the trivial representation (this will become clearer in the discussion below).

Now we explain this in a little more detail. A philosophical discussion of the significance of (4.8) may be found in Section 4.3.3.

4.3.1. The source of (4.8) via periods. Writing $Y_{\mathbf{A}} = \mathrm{PGL}_2(F)\backslash \mathrm{PGL}_2(\mathbf{A}_F)$, we note that the Rankin-Selberg L-function may be expressed as a period integral:

$$L(s, \pi_1 \times \pi_2) \sim \int_{Y_{\mathbf{A}}} \varphi_1(g)\varphi_2(g)E_s(g)dg$$

where $\varphi_i \in \pi_i$ are the respective newforms, and E_s is the Eisenstein series corresponding to the new vector of the automorphic representation $|\cdot|^s \boxplus \omega^{-1}|\cdot|^{-s}$. Here $\sim$ means that there is a suitable constant of proportionality, depending on the archimedean types of the representations.

Let $\mathcal{B}_{\omega,\mathfrak{q}}$ be an orthogonal basis for the space of forms on $\mathrm{GL}_2(F)\backslash\mathrm{GL}_2(\mathbf{A}_F)$ of level $\mathfrak{q}$ and central character ω; let $\mathcal{B}_{1,1}$ be an orthogonal basis for the space of forms on $Y_{\mathbf{A}}$ of full level and trivial central character. By spectral expansion, we have the following identity:

$$\sum_{\varphi_1 \in \mathcal{B}_{\omega,\mathfrak{q}}} \left| \int_{Y_{\mathbf{A}}} \varphi_1\varphi_2 E_s \right|^2 = \int_{Y_{\mathbf{A}}} |\varphi_2.E_s|^2 = \int_{Y_{\mathbf{A}}} |\varphi_2|^2 \, |E_s|^2 = \sum_{\psi \in \mathcal{B}_{1,1}} \langle |E_s|^2, \psi\rangle \overline{\langle |\varphi_2|^2, \psi\rangle} \tag{4.9}$$

Here the ψ- summation is, *a priori*, over an orthonormal basis for $L^2(Y_{\mathbf{A}})$; however, the summand $\langle |\varphi_2|^2, \psi\rangle$ vanishes unless ψ is of level 1 and trivial central character. Note that the ψ-summation should, strictly, include a continuous contribution for the Eisenstein series, which also needs to be suitably regularized. This is not a trivial matter and occupies a good deal of [MV]; we shall supress it for now.

In any case, if ψ belongs to the space of an automorphic representation τ, the Rankin-Selberg method shows that $\langle |E_s|^2, \psi\rangle$ is a multiple of $L(\tau, 2s - \frac{1}{2})L(\tau \times \omega, \frac{1}{2})$. Thus (4.9) basically yields (4.8)!

4.3.2. Introducing an amplifier, and using decay of matrix of coefficients. We restrict to $s = 1/2$ for concreteness, although the method works for any s. The identity (4.9) does not suffice to obtain a nontrivial bound on $\int_{Y_{\mathbf{A}}} \varphi_1\varphi_2 E_{1/2}$, for the left-hand summation is too large. To localize it, one introduces an amplifier. We phrase it adelically, but it should be made clear this is still the amplifier of Friedlander/Iwaniec.

For any function f on $\mathrm{GL}_2(F)\backslash\mathrm{GL}_2(\mathbf{A}_F)$ and any $g_0 \in \mathrm{GL}_2(\mathbf{A}_F)$, we write $f^{g_0}(g) = f(gg_0^{-1})$. Then one has the following tiny variant of (4.9), for $g_1, g_2 \in \mathrm{GL}_2(\mathbf{A}_F)$:

$$\sum_{\varphi_1 \in \mathcal{B}_{\omega,\mathfrak{q}}} \left(\int_{Y_{\mathbf{A}}} \varphi_1^{g_1^{-1}} \varphi_2 E_{1/2} \right) \overline{\left(\int_{Y_{\mathbf{A}}} \varphi_1^{g_2^{-1}} \varphi_2 E_{1/2} \right)} = \sum_{\psi \in L^2(Y_{\mathbf{A}})} \langle E_{1/2}^{g_1} \overline{E_{1/2}^{g_2}}, \psi\rangle \overline{\langle \overline{\varphi_2^{g_1}} \varphi_2^{g_2}, \psi\rangle} \tag{4.10}$$

This is again an identity of this shape (4.8), but with slightly more freedom due to the insertion of g_1, g_2. The left-hand (resp. right-hand) side is still proportional, by the Rankin-Selberg method, to $L(\pi_1 \times \pi_2, 1/2)$ (resp. $L(\tau, 1/2)L(\tau \times \omega, 1/2)$, if $\psi \in \tau$) but the constants of proportionality depend – in a precisely controllable way – on g_1, g_2. In effect, this allows one to introduce a "test function" $h(\pi_1)$ into the identity (4.8), thereby shortening the effective range of summation. It should be noted that in (4.10), by contrast with (4.8), the right hand ψ-summation is no longer over ψ of level 1; however, it involves only those ψ which are invariant by $\mathrm{PGL}_2(\hat{\mathbf{Z}}) \cap g_1^{-1}\mathrm{PGL}_2(\hat{\mathbf{Z}})g_1 \cap g_2^{-1}\mathrm{PGL}_2(\hat{\mathbf{Z}})g_2$, and in particular their level is bounded in a way that depends predictably on g_1, g_2.

A subconvex bound for $L(\pi_1 \times \pi_2, \frac{1}{2},)$ follows from any method to get nontrivial bounds on the right-hand side of (4.10) for general g_1, g_2.

1. To deal with the case when ψ is perpendicular to the constants, we note that the terms $\langle E_{1/2}^{g_1} E_{1/2}^{g_2}, \psi\rangle$ are, by Rankin-Selberg, certain multiples of $L(\tau, \frac{1}{2})L(\tau \times \omega, \frac{1}{2})$ whenever $\psi \in \tau$, the space of an automorphic representation. We then apply subconvex bounds for $L(\tau \times \omega, \frac{1}{2})$, in the aspect when ω varies.

2. To deal with the case $\psi = \text{Const}$, we note that the term $\langle \varphi_2^{g_1}\overline{\varphi_2^{g_2}}, \psi\rangle$ is, in that case, simply a multiple of the matrix coefficient $\langle \varphi_2^{g_1 g_2^{-1}}, \varphi_2\rangle$. Thus it is bounded by bounds on the decay of matrix coefficients.

Obviously this description is dishonest, for the term $\langle |E_{1/2}|^2, \psi\rangle$ is not even convergent for $\psi = \text{Const}$! However, this is a technical and not a conceptual difficulty: the only two analytic ingredients required are the two above.

Finally we comment that, in the general case where not all the ramification of π_1 arises from central character, one proceeds precisely as above, but the vector $E_{1/2}$ is not the new vector in the underlying automorphic representation. Rather, $E_{1/2} = E_{1/2,new}^{g_{\mathfrak{q}'}}$ for a suitable $g_{\mathfrak{q}'} \in \mathrm{GL}_2(\mathbf{A}_F)$, i.e. one takes a certain translate of the new vector. In that case, the proof becomes strictly simpler: as long as g_1, g_2 commute with $g_{\mathfrak{q}'}$, we write

$$\langle E_{1/2}^{g_1}\overline{E_{1/2}^{g_2}}, \psi\rangle = \langle (E_{1/2,new}^{g_1}\overline{E_{1/2,new}^{g_2}})^{g_{\mathfrak{q}'}}, \psi\rangle$$

and again simply use decay of matrix coefficients! We hope these examples indicate the power of the decay of matrix coefficients when studying periods.

4.3.3. Mysterious identities between families of L-functions and their interpretation. In a sense, (4.8) is the key point of the above discussion; it explains immediately why one has the "reduction of degree" discussed in Sec. 3.4. As discussed, (4.8) is essentially the period identity (4.9). This is another example of the phenomenon discussed in Sec. 4.1: we have an identity that is obvious from the spectral (period) viewpoint, but not at all clear from the viewpoint of L-functions considered as Dirichlet series.

This type of identity is not an isolated phenomenon. Perhaps a better-known example of such an identity between *a priori* different families of L-functions is Motohashi's beautiful formula [Mot97] for the 4th moment of ζ. Roughly speaking, it relates integrals of $|\zeta(1/2+it)|^4$ to sums of $L(\varphi, \frac{1}{2})^3$, where φ varies over Maass forms. Again, one can gain some insight into this from the "period" perspective. If φ is a suitably normalized Maass form on $\mathrm{SL}_2(\mathbf{Z})\backslash\mathbb{H}$, its completed L-function is given by the Hecke period $\Lambda(\varphi, \frac{1}{2}+it) = \int_0^\infty \varphi(iy)y^{it}d^\times y$. Applying Plancherel's formula shows that $\frac{1}{2\pi}\int_{-\infty}^{\infty}|\Lambda(\varphi, \frac{1}{2}+it)|^2 dt = \int_{y=0}^{\infty}|\varphi(iy)|^2 d^\times y$. Again, one can spectrally expand $|\varphi|^2$, yielding:

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}|\Lambda(\varphi, \frac{1}{2}+it)|^2 dt = \sum_\psi \frac{\langle|\varphi|^2, \psi\rangle}{\langle\psi, \psi}\int_0^\infty \psi(iy)d^\times y = \sum_\psi \frac{\langle|\varphi|^2, \psi\rangle}{\langle\psi, \psi\rangle}\Lambda(\psi, \frac{1}{2}) \tag{4.11}$$

where, again, the ψ-sum is over an orthogonal basis, suitably normalized, for $L^2(Y_0(1))$, and, again, we suppress the continuous spectrum. (4.11) expresses a relation between mean values of $L(\varphi, \frac{1}{2}+it)$, where t varies, and the family of L-functions $L(\psi, \frac{1}{2})$, where ψ varies. Specializing (4.11) to the case of φ the Eisenstein series at the center of symmetry yields a formula "of Motohashi type." We emphasize that this argument is not rigorous (for the integrals diverge in the Eisenstein case) and has not been carried out rigorously to our knowledge; it would likely involve considerable technical difficulty. Nevertheless, this approach may have value insofar as it offers some insight into the origin of such formulae.

A. Reznikov has given a very general and elegant formalism [Rez05] that encapsulates such identities as (4.8) and (4.11); one hopes that further analytic applications will stem from his formalism.

5. Applications

5.1. Subconvexity and functoriality. Via the functoriality principle of Langlands, it is now understood that the same L-function may be attached to automorphic forms on different groups. This gives rise to the possibility of studying the same L-function in different ways; this is a very powerful idea (implicitly already encountered when we discussed connections between Weyl sums and central values of L-functions).

A recent instance where this kind of idea played a decisive role was the attempt to solve the subconvexity problem for the L-functions of the class group characters of a quadratic field K of large discriminant. In [DFI95], Duke/Friedlander/Iwaniec were able to solve the problem but only under the assumption that K has sufficiently many small *split* primes (it would follow from the Generalized Riemann Hypothesis, but so far, has been established unconditionally only for special discriminants). This assumption, which was also encountered by the second author in [Ven05] in the context of periods and which is closely related to

Linnik's condition, is a fundamental and major unsolved issue that arises in many contexts, e.g. in work on the André/Oort conjecture [Edi05, Yaf03].

A key observation of [DFI02], is that, by functoriality, (in that case due to Hecke and Maass) a class group character L-function is the L-function of a Maass form of weight 0 or 1, with Laplace eigenvalue 1/4. For these, as we have just seen, the subconvexity problem can be solved independently of any assumption. For that reason, we find it useful to spell out explicitly some direct consequences of the subconvex bounds of Theorem 6 and of functoriality.

Corollary 5.1. *Let F be a fixed number field and $\rho : \mathrm{Gal}(\bar{F}/F) \to \mathrm{GL}_2(\mathbf{C})$ be a modular Galois representation (for instance, if the image of ρ in $\mathrm{PGL}_2(\mathbf{C})$ is soluble). Let $\mathfrak{q}_\rho$ be the Artin conductor of ρ and let $L(\rho, s)$ be its Artin L-function, then for $\Re es = 1/2$*

$$L(\rho, s) \ll_{F,s} N_{F/\mathbf{Q}}(\mathfrak{q}_\rho)^{1/4-\delta}$$

for $\delta > 0$ some absolute constant.

Corollary 5.2. *Let F be a fixed number field and K be an extension of F of absolute discriminant $\mathrm{disc}(K/\mathbf{Q}) =: \Delta_K$ and let $\zeta_K(s)$ be the Dedekind zeta function of K; then, if K/F is abelian or cubic, one has for $\Re es = 1/2$*

$$\zeta_K(s) \ll_{F,s} |\Delta_K|^{1/4-\delta}$$

for $\delta > 0$ some absolute constant.

Corollary 5.3. *Let F be a fixed number field, π be a fixed $\mathrm{GL}_2(\mathbf{A}_F)$-automorphic cuspidal representation and let K be an extension of F of absolute discriminant $\mathrm{disc}(K/\mathbf{Q}) =: \Delta_K$. If K/F is abelian or cubic, we denote by π_K the base change lift of π from F to K (which exist by the works of Saito/Shintani/Langlands and Jacquet/Piatetski-Shapiro/Shalika). For $\Re es = 1/2$, one has*

$$L(\pi_K, s) \ll_{F,\pi,s} |\Delta_K|^{1/2-\delta}$$

for $\delta > 0$ some absolute constant.

5.2. Equidistribution on quaternionic varieties.

We define a quaternionic variety as the locally homogeneous space given as an adelic quotient of the following form: for F a totally real number field, B a quaternion algebra over F, let $\mathbf{G}$ be the $\mathbf{Q}$-algebraic group $\mathrm{res}_{F/\mathbf{Q}}\mathrm{B}^\times/Z(\mathrm{B}^\times)$; one has

$$G(\mathbf{R}) \simeq \mathrm{PGL}_2(\mathbf{R})^{f'} \times \mathrm{SO}(3,\mathbf{R})^{f-f'}$$

where $f = \deg F$ and f' is the number of real place of F for which B splits. Let K_∞ be a compact subgroup of $\mathbf{G}(\mathbf{R})$ of the form

$$\mathrm{SO}(2,\mathbf{R})^{f'} \times \prod_{v=1}^{f-f'} \mathrm{K}_v$$

with $\mathrm{K}_v =$ either $\mathrm{SO}_2(\mathbf{R})$ or $\mathrm{SO}_3(\mathbf{R})$ and let X denote the quotient $\mathbf{G}(\mathbf{R})/\mathrm{K}_\infty$; finally let K_f be an open compact subgroup of $\mathbf{G}(\mathbf{A}_f)$ and $\mathrm{K} := \mathrm{K}_\infty.\mathrm{K}_f$.

The quaternionic variety $V_\mathrm{K}(\mathbf{G}, X)$ is defined as the quotient

$$V_\mathrm{K}(\mathbf{G}, X) := \mathbf{G}(\mathbf{Q})\backslash X \times \mathbf{G}(\mathbf{A}_f)/\mathrm{K}_f = \mathbf{G}(\mathbf{Q})\backslash\mathbf{G}(\mathbf{A}_\mathbf{Q})/\mathrm{K}.$$

It has the structure of a Riemannian manifold whose connected components are quotients, by a discrete subgroup of $\mathbf{G}(\mathbf{R})$, of the product of $(\mathbb{H}^\pm)^{f'} \times (S^2)^{f''}$ for $f'' \leqslant f - f'$. The case of the sphere and of the modular surface correspond to the case $F = \mathbf{Q}$, B the algebra of 2×2 matrices $M_2(\mathbf{R})$ or the Hamilton quaternions $\mathrm{B}^{(2,\infty)}$, with $K_\infty = \mathrm{SO}(2,\mathbf{R})$ and K_f the maximal compact subgroup of $\mathbf{G}(\mathbf{A}_f)$.

Let K/F be a quadratic extension with an embedding into B, and let $\mathbf{T}$ denote the $\mathbf{Q}$-torus "$\mathrm{res}_{F/\mathbf{Q}}K^\times/F^\times$". As was pointed out in section 2.2.1, there exists, in great generality, a precise relationship between:

1. Central values of some Rankin/Selberg L-function $L(\pi_\chi \times \pi_2, s)$ (for which the sign of the functional equation $w(\pi_\chi \times \pi_2)$ is $+1$); and

2. (the square of) twisted Weyl sums

$$\int_{\mathbf{T}(\mathbf{Q})\backslash\mathbf{T}(\mathbf{A}_{\mathbf{Q}})}\chi(t)\varphi_2(z.t)dt$$

These Weyl sums describe the distribution properties of toric orbits, $\mathbf{T}(\mathbf{Q})\backslash z.\mathbf{T}(\mathbf{A}_{\mathbf{Q}})$ of cycles associated to (orders of K) inside $V_{\mathrm{K}}(\mathbf{G}, X)$.

The general scheme is that, in cases where these formula have been written out explicitly, the subconvex bound (3.2) (along possibly with hypothesis H_θ for some $\theta < 1/2$) yields at once the equidistribution of the *full* orbit and the subconvex bounds (3.4) yield the equidistribution of *big enough* suborbits of the toric orbit. We present below further sample of results that have been proven along these lines. However, we note that there is considerable scope for generalization of these equidistribution results, in particular regarding the formulas of Waldspurger/Gross/Zagier/Zhang type which can be worked out precisely for more general quadratic fields (say not necessarily totally real or totally imaginary).

5.3. Hilbert's eleventh problem. Similarly to what was explained before in the case $F = \mathbf{Q}$, when B is totally definite, $\mathrm{K}_\infty = \mathrm{SO}_2(\mathbf{R})^f$, $X = (S^2)^f$ is a product of spheres. In this case, the equidistribution of toric orbits (relative to a totally imaginary quadratic field) above can be interpreted in terms of the integral representations of a totally positive integer $d \in \mathcal{O}_F$ by a totally positive definite quadratic form q (more precisely $-q$ "is" the norm form $N_{B/F}(\mathbf{x})$ on the space of quaternions of trace 0) .

More generally, Hilbert's eleventh problem asked, amongst other things, which integers are represented by integral quadratic form over (the ring of integers in) a fixed number field. When F is not totally real or q is indefinite at some place or q has more than 4 variables this question was basically solved through various methods during the 20th century. Recently, Cogdell/Piatetski-Shapiro/Sarnak settled the last remaining case of positive definite ternary forms by following an approach similar to the original method of Duke. In this approach, however, the non-trivial bound for Fourier coefficients of half-integral weight forms (2.3) is obtained by combining a formula of Baruch/Mao [BM03] (generalizing the Kohnen/Zagier formula) with their subconvex bound (3.2) for π_2 holomorphic. Their first result is (see also [DSP90] for the case $F = \mathbf{Q}$)

Theorem 7. *Let F be a totally real number field and q be an integral positive definite quadratic form over F; there is an absolute (ineffective) constant $N_{F,q} > 0$ such that if d is a squarefree totally positive integer with $N_{F/\mathbf{Q}}(d) > N_F$ then d is integraly represented by q iff d is everywhere locally integrally represented. Moreover, in the later case, the number, $r_q(d)$, of all such integral representation satisfies*

$$r_q(d) \gg_{q,F} N_{F/\mathbf{Q}}(d)^{1/2+o(1)}, \text{ as } N_{F/\mathbf{Q}}(d) \to +\infty.$$

Remark 5.1. The question of the integral representability of d by some form in the genus of q was completely settled a long time ago by Siegel, in a quantitative way, through the Siegel mass formula. The present theorem (or a slightly more precise form of it) can then be interpreted by saying that the various representations d are *equidistributed* amongst the various genus classes of q. Their next theorem is an even stronger version of this equidistribution (see [DSP90] for the case $F = \mathbf{Q}$).

Theorem 8. *With the same notations as above, let $\sigma : x \to (\sigma_1(x), \dots, \sigma_f(x))$ be the standard embedding of F into $\mathbf{R}^f$ and let $V_{q,1}(\mathbf{R})$ denote the variety of level 1 of q (a product of ellipsoids in $(\mathbf{R}^3)^f$):*

$$V_{q,1}(\mathbf{R}) = \{(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_f),\ \mathbf{x}_i \in \mathbf{R}^3,\ \sigma_i(q)(\mathbf{x}_i) = 1,\ i = 1\dots f\} \subset (\mathbf{R}^3)^f.$$

If $R_q(d) > 0$, let $\mathscr{G}_d$ denote

$$\mathscr{G}_d = \{(\frac{1}{\sigma_1(d)^{1/2}}\sigma_1(\mathbf{x}), \dots, \frac{1}{\sigma_f(d)^{1/2}}\sigma_f(\mathbf{x})),\ \mathbf{x} \in \mathcal{O}_F,\ q(\mathbf{x}) = d\} \subset V_{q,1}(\mathbf{R})$$

the projection of the integral solutions of the equation $q(x) = d$ to $V_{q,1}(\mathbf{R})$. Then, as $N_{F/\mathbf{Q}}(d) \to +\infty$, with $r_q(d) > 0$, the set $\mathscr{G}_d$ becomes equidistributed on $V_{q,1}(\mathbf{R})$.

5.4. Equidistribution of CM points on quaternionic Shimura varieties.

When B is indefinite at some real place and $\mathrm{K}_\infty = \mathrm{SO}_2(\mathbf{R})^{f'} \times \mathrm{SO}_3(\mathbf{R})^{f-f'}$ the quaternionic variety $V_{\mathrm{K}}(\mathbf{G}, X)$ is a Shimura variety, $\mathrm{Sh}_{\mathrm{K}_f}(\mathbf{G}, X)$ (a Hilbert modular variety of complex dimension f'). It has the structure of the complex points of an algebraic variety defined over some reflex field E/F.

In this setting, the generalization of the set of Heegner point is the so called set of "CM" points, $\mathscr{H}_{\mathfrak{d}}$, which is associated to a quadratic order $\mathcal{O}_{\mathfrak{d}}$ (say of discriminant $\mathfrak{d}$) of a (not necessarily fixed) totally imaginary K/F. In that case and under some natural local condition, the equidistribution of

$$\mathscr{H}_{\mathfrak{d}} = \mathbf{T}(\mathbf{Q})\backslash z_{\mathfrak{d}}.\mathbf{T}(\mathbf{A}_{\mathbf{Q}})/\mathbf{T}(\widehat{\mathcal{O}}_{\mathfrak{d}})$$

on $\mathrm{Sh}_{\mathrm{K}_f}(\mathbf{G}, X)$ as $|N_{F/\mathbf{Q}}(\mathfrak{d})| \to +\infty$ was established independently by Clozel/Ullmo, Cohen and Zhang [CU05, Coh05, Zha04] by using the subconvex bound (3.2) of the second author. For instance, one has

Theorem 9. *Suppose $\mathrm{K}_f = \mathrm{K}_{f,\max}$ is a maximal compact subgroup of $\mathbf{G}(\mathbf{A}_F)$, then for $|N_{F/\mathbf{Q}}(\mathfrak{d})| \to +\infty$ and $\mathfrak{d}$ coprime with $\mathrm{disc}(F)$, the set $\mathscr{H}_{\mathfrak{d}}$ becomes equidistributed on $\mathrm{Sh}_{\mathrm{K}_f}(\mathbf{G}, X)$ w.r.t. the hyperbolic measure.*

Similarly, as in Theorem 5, the bound (3.4) allows one to show the equidistribution of strict suborbits of $z_{\mathfrak{d}}$:

Theorem 10. *With the notations as above, there is an absolute constant $0 < \eta < 1$ such that, for any subtoric orbit $\mathscr{H}'_{\mathfrak{d}} \subset \mathscr{H}_{\mathfrak{d}}$ of size satisfying $|\mathscr{H}'_{\mathfrak{d}}| \geqslant |\mathscr{H}_{\mathfrak{d}}|^{\eta}$, then $\mathscr{H}'_{\mathfrak{d}}$ is equidistributed on $\mathrm{Sh}_{\mathrm{K}_f}(\mathbf{G}, X)$ as $|N(\mathfrak{d})| \to +\infty$,.*

As was pointed out by Zhang [Zha04], the possibility of considering strict suborbits of the full toric orbit has a very nice arithmetic interpretation. Indeed any CM point $z_{\mathfrak{d}} \in \mathscr{H}_{\mathfrak{d}}$ is defined over some ring class extension above K and a natural question (motivated in part from the André/Oort conjecture [Ull02]) is the question of the distribution of the Galois orbits of such points. By Shimura's theory of complex multiplication, the galois orbit of $z_{\mathfrak{d}}$ is a (strict in general) subtoric orbit of the toric orbit $\mathbf{T}(\mathbf{Q})\backslash z_{\mathfrak{d}}.\mathbf{T}(\mathbf{A}_{\mathbf{Q}})/\mathbf{T}(\widehat{\mathcal{O}}_{\mathfrak{d}})$, thus it follows from Theorem 10, that if the Galois orbit is big enough, it is equidistributed. This however is not to be expected, in full generality, for there exists families of CM points whose the Galois orbit sits entierely on fixed strict sub-Shimura variety of $\mathrm{Sh}_{\mathrm{K}_f}(\mathbf{G}, X)$. On the other hand, in the (generic) case of CM points whose Mumford/Tate group is as big as possible (equal to the full torus $\mathbf{T} = \mathrm{res}_{F/\mathbf{Q}} K^\times/F^\times$) it is expected (and proven in a very limited number of cases) that the size of the galois orbit is bigger than $|\mathscr{H}_{\mathfrak{d}}|^{1-\varepsilon}$ for any $\varepsilon > 0$ which would be more than sufficient to obtain equidistribution [Zha04].

Remark 5.2. In the same vein, the generalization of the sets of geodesics, Γ_d are the sets, $\Gamma_{\mathfrak{d}}$, of compact flats of maximal dimension $\mathrm{Sh}_{\mathrm{K}_f}(\mathbf{G}, X)(\mathbf{C})$ associated to totally real quadratic orders $\mathcal{O}_{\mathfrak{d}}$. To our knowledge the generalization of (2.2) and (2.5) has not been written up yet for fields F larger that $\mathbf{Q}$. In any case, once such generalizations become available, the above subconvex bound will imply similar equidistribution statement of these flats.

6. A modern perspective on Linnik's original ergodic method

6.1. The source of dynamics.

Although the relevance of dynamics to integral points on the sphere is not immediately apparent, it is not difficult to see from an adelic perspective. We have already mentioned in Section 2.2.1 that all three theorems (Theorems 1 – 3) may be considered as questions about the distribution of an orbit of an adelic torus $z_d.\mathbf{T}_d(\mathbf{A})$ inside $\mathbf{G}(\mathbf{Q})\backslash\mathbf{G}(\mathbf{A})$.

One can, therefore, hope to use results about the dynamics of $\mathbf{T}_d(\mathbf{Q}_v)$-action on $\mathbf{G}(\mathbf{Q})\backslash\mathbf{G}(\mathbf{A})$. This is of use, however, only if $\mathbf{Q}(\sqrt{d})$ splits at v; for if this is not the case, $\mathbf{T}_d(\mathbf{Q}_v)$ is a compact group and has no dynamics of interest.

This leads to Linnik's condition: if p is a prime such that $\mathbf{Q}(\sqrt{d})$ is split at p, then one may analyze Theorems 1 – 3 by studying the dynamics of $\mathbf{T}_d(\mathbf{Q}_p) \cong \mathbf{Q}_p^\times$ acting on $\mathbf{G}(\mathbf{Q})\backslash\mathbf{G}(\mathbf{A})$. This is precisely Linnik's method in modern language.

In the case of Theorem 3, we may take instead $v = \infty$, which is split in the (real) quadratic field $\mathbf{Q}(\sqrt{d})$. This amounts to studying the collection of closed geodesics described in Theorem 3 through the dynamics of the geodesic flow. Curiously, this was apparently never done by Linnik, who only used the action of p-adic tori.

6.2. Linnik's method in the light of modern ergodic theory. As discussed, Linnik acheived partial results towards Theorems 1 – 3 by using some ingenious ideas which he collectively referred to as "the ergodic method." As Linnik pointed out (see, e.g., [Lin68, Chapter XI, comments on Chapters IV–VI]) despite this name, this method remained rather *ad hoc* and did not fit into ergodic theory as it is normally understood: that is to say, dynamics of a measure-preserving transformation. The joint work of the authors with M. Einsiedler and E. Lindenstrauss, in [ELMV] and [ELMVb], remedies this, both putting Linnik's original work into a more standard ergodic context, and giving the first higher rank results.

Much of this joint work is based on the recent work of Einsiedler and Lindenstrauss on classification of invariant measures for higher rank tori, which is discussed in their contribution to these proceedings [EL06]. An important difference between the higher rank case and the rank 1 case considered by Linnik is the phenomenon of *measure rigidity*; in general, the actions of higher rank tori are far more rigid (have fewer invariant measures and closed invariant sets) than the actions of rank 1 tori. A hint of this distinction between rank 1 and rank 2 actions may already be seen in a more elementary context: Although a real number x can have an irregular expansion to base p, i.e. $\{p^n x\}$ can behave irregularly modulo 1, it is much more difficult for $\{p^n q^m x\}$ to behave irregularly mod 1, if p and q are coprime. This phenomenon was first studied by Furstenberg; for a recent survey we refer the reader to [Lin].

A central concept here is that of *entropy*; we briefly reprise the definition. We recall that if $\mathscr{P}$ is a partition of the probability space (X, ν), the entropy of $\mathscr{P}$ is defined as:

$$h_\nu(\mathscr{P}) := \sum_{S \in \mathscr{P}} -\nu(S) \log \nu(S). \tag{6.1}$$

This has the following basic subadditivity property: if $\mathscr{P}_1, \mathscr{P}_2$ are two partitions, then $h_\nu(\mathscr{P}_1 \vee \mathscr{P}_2) \leqslant h_\nu(\mathscr{P}_1) + h_\nu(\mathscr{P}_2)$, where $\vee$ denotes common refinement. If T is a measure-preserving transformation of (X, ν), then the measure entropy of T is defined as:

$$h(T) = \sup_{\mathscr{P}} \lim_{n \to \infty} \frac{h_\nu(\mathscr{P} \vee T^{-1}\mathscr{P} \vee \cdots \vee T^{-(n-1)}\mathscr{P})}{n} \tag{6.2}$$

where the supremum is taken over all finite partitions of X.

Here are two results that illustrate the importance of this concept. We denote by Haar the G-invariant probability measure on a quotient space $\Gamma \backslash G$.

Fact 1. Let μ on $\mathrm{SL}_2(\mathbf{Z}) \backslash \mathrm{SL}_2(\mathbf{R})$ be invariant by the diagonal subgroup, and let a be a nontrivial diagonal matrix. Then $h_\mu(a) \leqslant h_{\mathrm{Haar}}(a)$, with equality if and only if $\mu = \mathrm{Haar}$.[7]

Fact 2. Let μ be a probability measure on $\mathrm{SL}_3(\mathbf{Z}) \backslash \mathrm{SL}_3(\mathbf{R})$ invariant by the diagonal subgroup A and let $a \in A$ be nontrivial. If $h_\mu(a) > 0$ and μ is ergodic (w.r.t. A), then $\mu = \mathrm{Haar}$. This lies much deeper than *Fact 1* and is a result of Einsiedler, Katok and Lindenstrauss [EKL06].

The scheme of [ELMV] and [ELMVb] is to treat Linnik problems by combining results of the above type – towards the classification of measures with positive entropy – with Diophantine ideas that establish positive entropy.

The power of results such as *Fact 2* for number-theoretic purposes becomes manifest in this context. The reason that one obtains much stronger results in the $\mathrm{SL}_3(\mathbf{R})$ context than the $\mathrm{SL}_2(\mathbf{R})$ context is another manifestation of "measure rigidity."

In the subsequent sections we discuss some applications of this general scheme that are carried out in these papers; we have aimed for concreteness, but we believe that these methods will be much more generally applicable.

6.3. Linnik's original proof: the "Linnik principle" and entropy. In [ELMVb] we give a purely dynamical proof of Theorem 3, concerning equidistribution of geodesics of fixed discriminant. Although, unlike the prior work of Skubenko, it requires no auxiliary prime splitting, this proof is still based heavily on Linnik's ideas. However, it introduces considerable conceptual simplification using the notion of entropy discussed in the previous Section, and uses in particular *Fact 1.*

---

[7] More generally, this result is true for the geodesic flow on the unit tangent bundle of a surface of constant negative curvature: the Liouville measure is the unique measure of maximal entropy.

Let $d > 0$ be a fundamental discriminant. The unit tangent bundle of $Y_0(1)$ is identified with $\mathrm{PGL}_2(\mathbf{Z})\backslash\mathrm{PGL}_2(\mathbf{R})$, and so the subset Γ_d described in Theorem 3 may be regarded as a subset $\Gamma_d \subset \mathrm{PGL}_2(\mathbf{Z})\backslash\mathrm{PGL}_2(\mathbf{R})$. Considered in this way, Γ_d is invariant by the subgroup

$$A = \{a(t) = \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix}, t \in \mathbb{R}\}$$

of diagonal matrices with positive entries. It supports a natural A-invariant probability μ_d (the one which assigns the same mass to each connected component) and Theorem 3 asserts precisely that μ_d converge weakly to the $\mathrm{PGL}_2(\mathbf{R})$-invariant probability measure on $\mathrm{PGL}_2(\mathbf{Z})\backslash\mathrm{PGL}_2(\mathbf{R})$.

The dynamical proof – which is based very much on Linnik's ideas – uses *Fact 1* together with a Diophantine computation (based on "Linnik's basic Lemma") to show that any weak limit of the μ_d has maximal entropy w.r.t. the action of $a(1)$.

For now we shall merely briefly indicate how Diophantine considerations enter. In view of the definition (6.1), a lower bound on the entropy $h_\nu(\mathscr{P})$ of a partition follows if one knows a lower bound for the mass $\nu(S)$ of any $S \in \mathscr{P}$. Now, let μ_∞ be a weak limit of the μ_d. We shall assume here (for simplicity) that μ_∞ is a probability measure, i.e. that there is no "escape of mass" in the noncompact space $\mathrm{PGL}_2(\mathbf{Z})\backslash\mathrm{PGL}_2(\mathbf{R})$.

To give a positive lower bound on the entropy of $a(1)$ w.r.t. μ_∞ it suffices to show that there is a sequence of integers n_d so that:

$$\mu_d(\mathscr{P} \vee a(1)^{-1}\mathscr{P} \vee \cdots \vee a(1)^{-(n_d-1)}\mathscr{P}) \geqslant c^{n_d}$$

for some absolute constant c.

As it turns out, each piece of such a refined partition looks like a very small tube around a piece of an A-orbit. A hint on how to bound the μ_d-mass of such a tube is already given by the following Lemma, which shows that points in Γ_d are well-separated tranverse to the A-direction.

Lemma 6.1. *Fix any compact subset* $\Omega \subset \mathrm{PGL}_2(\mathbf{Z})\backslash\mathrm{PGL}_2(\mathbf{R})$ *and a Riemannian metric on* $\mathrm{PGL}_2(\mathbf{Z})\backslash\mathrm{PGL}_2(\mathbf{R})$*. Then there is* $c_\Omega > 0$ *such that, for any* $y, y' \in \Omega \cap \Gamma_d$ *such that* $d(y, y') < c_\Omega d^{-1/2}$*, there exists* $t \leqslant 1$ *such that* $y = y'a(t)$*.*

The proof of this is very simple: it is a translation of the fact that any two distinct integral binary quadratic forms $ax^2 + bxy + cy^2$, considered as points in the affine space of (a, b, c), are separated by at least 1! (Recall that Γ_d was constructed from a set $\mathcal{Q}_d(\mathbf{Z})$ of binary quadratic forms of discriminant d).

Thus even a relatively trivial Diophantine consideration (any two integer points are separated by $\geqslant 1$) already gives a nontrivial bound on entropy. To apply Fact 1, however, one needs an *optimal* bound, and this requires a slightly more sophisticated Diophantine argument; it requires a version of "Linnik's basic Lemma," cf. [Lin68, Theorem III.2.1], which in turn may be deduced from Siegel's mass formula.

6.4. A rank 3 version of Duke's theorem. Supersparse equidistribution.

A natural "rank 2" version of Theorem 3 considers adelic torus orbits inside $\mathrm{PGL}_3(\mathbf{Q})\backslash\mathrm{PGL}_3(\mathbf{A})$. One of the main theorems (part 1 of Theorem 11 below) of [ELMV] obtains a weak form of equidistribution ("well-distribution") in this context. A key ingredient in this theorem is *Fact 2* mentioned above.

Let D be a $\mathbf{R}$-split central simple algebra of rank 3 over $\mathbf{Q}$, i.e. $\dim_{\mathbf{Q}} \mathrm{D} = 9$, so that $\mathrm{D} \otimes_{\mathbf{Q}} \mathbf{R} = M_3(\mathbf{R})$. Let $\mathcal{O}_{\mathrm{D}}$ be a fixed order in D. Let $\mathbf{G}$ be the algebraic group $\mathrm{PG}(\mathrm{D}) = \mathrm{D}^\times/Z(\mathrm{D})^\times$; we fix a maximal split torus $A = (\mathbf{R}^\times)^2$ inside $\mathbf{G}(\mathbf{R})$. Let U be the standard maximal compact subgroup $\prod_p \mathrm{PG}(\mathcal{O}_{\mathrm{D},p})$ of $\mathbf{G}(\mathbf{A}_f)$. We will assume, for simplicity, that the class number of $\mathcal{O}_{\mathrm{D}}$ is 1, i.e. that $\mathbf{G}(\mathbf{A}_f) = \mathbf{G}(\mathbf{Q}).U$.

Let $K \subset \mathrm{D}$ be a totally real cubic field, together with an isomorphism $\theta : K \otimes \mathbf{R} \to \mathbf{R}^3$. We assume for simplicity that $K \cap \mathcal{O}_{\mathrm{D}}$ is the maximal order $\mathcal{O}_K$ of K. This yields, in particular, an embedding of the torus $\mathbf{T}_K = \mathrm{res}_{K/\mathbf{Q}} K^\times/\mathbf{Q}^\times$ into the algebraic group $\mathrm{PG}(\mathrm{D})$. The choice of θ determines a unique $g_\theta \in \mathbf{G}(\mathbf{R})$ so that $g_\theta A g_\theta^{-1} = \mathbf{T}_K(\mathbf{R})$.

Setting $U_T = \mathbf{T}_K(\mathbf{A}_f) \cap U$, we consider

$$\Gamma_K := (\mathbf{T}_K(\mathbf{Q})\backslash\mathbf{T}_K(\mathbf{A})/U_T)g_\theta \subset \mathbf{G}(\mathbf{Q})\backslash\mathbf{G}(\mathbf{A})/U \cong \mathcal{O}_{\mathrm{D}}^\times\backslash\mathrm{PGL}_3(\mathbf{R}).$$

This is a collection of compact A-orbits on $\mathcal{O}_{\mathrm{D}}^\times\backslash\mathrm{PGL}_3(\mathbf{R})$, indexed by the set

$$\mathbf{T}_K(\mathbf{Q})\backslash\mathbf{T}_K(\mathbf{A}_f)/U_T$$

which is precisely the class group $\mathrm{Cl}(\mathcal{O}_K)$.

Thus we may associate to any subset $S \subset \mathrm{Cl}(\mathcal{O}_K)$ of the class group a collection of $|S|$ closed A-orbits on $\mathcal{O}_{\mathrm{D}}^{\times}\backslash \mathrm{PGL}_3(\mathbf{R})$; call this $\Gamma_{K,S}$. This supports a natural A-invariant probability measure (which assigns the same mass to each of the constitutent orbits); call this measure $\mu_{K,S}$. For $S = \mathrm{Cl}(\mathcal{O}_K)$ we write simply μ_K.

Theorem 11. *1. There is an absolute constant $c > 0$ so that, as* $\mathrm{disc}(K) \to \infty$*, any weak limit of the measures μ_K contains a Haar component of size $\geqslant c$.*[8]

2. Suppose D *is not* **Q***-split. Fix $\delta < 1/2$. There is a constant $c = c(\delta) > 0$ such that, if each set $S \subset \mathrm{Cl}(\mathcal{O}_K)$ satisfies $\frac{|S|}{|\mathrm{Cl}(\mathcal{O}_K)|} \geqslant \mathrm{disc}(K)^{-\delta}$, then any weak limit of $\mu_{K,S}$ contains a Haar component of size $\geqslant c(\delta)$.*

The condition "D is not **Q**-split" is used to avoid the difficulties associated with noncompactness. To handle the case of D **Q**-split (i.e. the case of $\mathrm{PGL}_3(\mathbf{Z})\backslash\mathrm{PGL}_3(\mathbf{R})$) for the first item requires, in addition to the general strategy outlined at the end of Sec. 6.2, an *analytic* result that shows that any weak limit of the μ_K is in fact a probability measure, i.e. that there is no "escape of mass" to ∞. The analytic result needed is, in fact, precisely Corollary 5.2 in the case $F = \mathbf{Q}$.

The second part of this theorem illustrates a major advantage of the ergodic approaches of [ELMV] in comparison to harmonic-analysis methods: they allow nontrivial results about *very small* torus orbits ("supersparse equidistribution"; the result allows any exponent $\delta < 1/2$; since the size of the class group of $\mathcal{O}_K$ is at most $\mathrm{disc}(K)^{1/2+\varepsilon}$, this is as strong as could be hoped for). This seems to be a general advantage, at present, of ergodic methods over harmonic analysis. It is not specific to torus orbits; it is also reflected in many applications of Ratner's theorems to questions that have not been successfully analyzed by harmonic analysis; see also Section 7.

In any case, such *small* torus orbits arise naturally in arithmetic questions – for instance, in the André/Oort conjecture, and we expect that measure rigidity results for actions of p-adic torii should allow for partial progress towards Zhang's measure-theoretic refinement of the André/Oort conjecture [Zha04, "Equidistribution Conjecture"].

Remark 6.1. One should compare the result of the second part of Theorem 1 with Conjecture 1; indeed, the Theorem is giving a result of the quality of Conjecture 1, but without the requirement that S be a *subgroup* of the Picard group $\mathrm{Cl}(\mathcal{O}_K)$. We note that one can certainly formulate an analogous question to the second part of this Theorem in the rank 1 context. However, it is very unlikely that anything as strong as the above result holds; this is again the phenomenon of measure rigidity alluded to at the beginning of the present Section.

6.4.1. The joinings theorem of Einsiedler and Lindenstrauss. We have indicated above some examples of the power of ergodic methods to yield results about quite sparse orbits. We briefly return to an earlier problem to give one more instance of this. As remarked after Conjecture 2, it is possible to consider the analogous Conjecture on $Y_0(1) \times S^2$ or more generally on products of quaternionic varieties associated to *distinct* quaternion algebras. It seems that this analogue is an immediate corollary of a recent result of Einsiedler and Lindenstrauss [EL06, Theorem 2.7], under a "stronger" Linnik type condition: if one varies d through a sequence of admissible discriminants for which $\mathbf{Q}(\sqrt{d})$ is split at *two* fixed primes p, q. The point of this assumption is that it allows application of the Theorem of Einsiedler/Lindenstrauss to the rank 2 action of $\mathbf{T}_d(\mathbf{Q}_p) \times \mathbf{T}_d(\mathbf{Q}_q)$.

On the other hand, these techniques do not appear to yield Conjecture 2, because of the difficulty of ruling out certain intermediate measures as possible limits. (How does one know that the limit measure associated to $\mathscr{H}_d''$ is not concentrated on the diagonal of $Y_0(1) \times Y_0(1)$?)

6.5. An application to Minkowski's theorem in higher rank.

We first recall Minkowski's theorem: if K is a number field of degree d and with maximal order $\mathcal{O}_K$, any ideal class for $\mathcal{O}_K$ possesses an integral representative $J \subset \mathcal{O}_K$ of norm $N(J) = O(\sqrt{\mathrm{disc}(K)})$ where the implicit constant depends only on d. We conjecture that this is not sharp for totally real number fields of degree $d \geqslant 3$:

Conjecture 3. Suppose $d \geqslant 3$ is fixed. Then any ideal class in a totally real number fields of degree d has an integral representative of norm $o(\sqrt{\mathrm{disc}(K)})$.

[8] One certainly conjectures that the measures μ_K are approaching the Haar measure on $\mathcal{O}_{\mathrm{D}}^{\times}\backslash\mathrm{PGL}_3(\mathbf{R})$; this would be a true rank 3 analogue of Theorem 3.

This conjecture can be seen as a result of the extra freedom that arises from having a large unit group, and is actually a consequence of a stronger conjecture formulated by Margulis [Mar00]. It seems very unlikely that it is true for $d = 2$; this is another manifestation of the difference between rank 1 and rank 2.

It will be convenient to denote by $m(K)$ the maximum, over ideal classes of $\mathcal{O}_K$, of the minimal norm of a representative. Thus the conjecture asserts that

$$\lim_{\mathrm{disc}(K)\to\infty} \frac{m(K)}{\mathrm{disc}(K)^{1/2}} = 0,$$

if K varies through totally real fields of fixed degree $d \geqslant 3$. Note, however, that Minkowski's theorem is rather close to sharp: it may be shown that for any $d \geqslant 2$ there exists a $c' > 0$ such that there is an infinite set of totally real fields of degree d for which $m(K) \geqslant c' \cdot \mathrm{disc}(K)^{1/2}(\log \mathrm{disc}(K))^{1-2d}$. Thus one might speculate that the true bound for $m(K)$ is perhaps $\mathrm{disc}(K)^{1/2}(\log \mathrm{disc}(K))^{-\alpha}$ for some small but positive α.

We will call an ideal class of a field K δ-bad if it does not admit a representative of norm $< \delta \cdot \mathrm{disc}(K)^{1/2}$. Let $h_\delta(K)$ be the number of δ-bad ideal classes and let R_K denote the regulator of the field K.

In [ELMV] it is shown that:

Theorem 12. *Let $d \geqslant 3$, and let K denote a totally real number field of degree d. For all $\varepsilon, \delta > 0$ we have:*

$$\sum_{\mathrm{disc}(K)<X} R_K h_\delta(K) \ll X^\varepsilon \tag{6.3}$$

In particular:

1. *"Conjecture 3 is true for almost all fields": The number of fields K with discriminant $\leqslant X$ for which $m(K) \geqslant \delta \cdot \mathrm{disc}(K)^{1/2}$ is $O_{\epsilon,\delta}(X^\epsilon)$, for any $\epsilon, \delta > 0$;*
2. *"Conjecture 3 is true for fields with large regulator": If K_i is any sequence of fields for which $\liminf_i \frac{\log R_{K_i}}{\log \mathrm{disc}(K_i)} > 0$, then $m(K_i) = o(\mathrm{disc}(K_i)^{1/2})$.*

This is connected to the considerations of Sec 6.4 in the following way: Consider the case $d = 3$. To a real cubic field K and suitable additional data we have associated a collection of compact A-orbits $\Gamma_K \subset \mathrm{PGL}_3(\mathbf{Z})\backslash\mathrm{PGL}_3(\mathbf{R})$, indexed by the class group of K. The key point is the following: the question of the minimal norm of a representative for a given ideal class is closely related to the question of how far the associated A-orbit penetrates into the "cusp" of the noncompact space $\mathrm{PGL}_3(\mathbf{Z})\backslash\mathrm{PGL}_3(\mathbf{R})$. This allows a geometric reformulation of Theorem 12 that is amenable to analysis by the methods of Sec. 6.4.

7. The roles of ergodic theory and harmonic analysis

In this concluding section, we briefly compare dynamical methods and harmonic analysis. We take the period viewpoint (i.e. we shall phrase the aims and results in terms of periods, rather than L-functions; a partial justification for this is that, although all L-functions are periods, there are periods of interest that are not L-functions.)

Fundamentally, the most general context for the type of problem we are considering is the following: let $\mathbf{H} \subset \mathbf{G}$ be a subgroup of a semisimple $\mathbf{Q}$-group $\mathbf{G}$; understand the "distribution" of $\mathbf{H}(\mathbf{Q})\backslash\mathbf{H}(\mathbf{A})$ inside $\mathbf{G}(\mathbf{Q})\backslash\mathbf{G}(\mathbf{A})$. Although we have not discussed it in the present article, such questions arise naturally in a large number of arithmetic questions.

The general approach to such questions have fallen into the following types:

1. Ergodic. Here we choose a suitable finite set of places S and apply results constraining $\mathbf{H}(\mathbf{Q}_S)$-invariant measures on $\mathbf{G}(\mathbf{Q})\backslash\mathbf{G}(\mathbf{A})$.
2. Harmonic-analysis. Here we choose a suitable basis φ_i for functions on $\mathbf{G}(\mathbf{Q})\backslash\mathbf{G}(\mathbf{A})$ and compute the "periods"

$$\int_{\mathbf{H}(\mathbf{Q})\backslash\mathbf{H}(\mathbf{A})} \varphi_i. \tag{7.1}$$

As we have seen in this article, there is considerable interaction between the two methods that has arisen in this context. In [ELMV] harmonic analysis is used to control escape of mass issues, which ergodic theory at present cannot handle; in [Ven05] quantitative ideas from ergodic theory are used to give estimates on periods such as (7.1).

It seems that in general, "good" quantitative upper bound for periods such as (7.1) should be considered as a central goal of the analytic theory of automorphic forms. Part of the difficulty is finding the correct notion of a "good" bound; one clear-cut case is when (7.1) is related to an L-function, in which case a "good" bound should yield subconvexity for the L-function.

At this level of generality, the following principles tend to apply:

1. If $\mathbf{H}$ is "a large enough subgroup" of $\mathbf{G}$ (say if $\mathbf{H}$ acts with an open orbit on the flag variety of $\mathbf{G}$), the periods (7.1) will often have "arithmetic significance", i.e. are often interpretable in terms of quantities of arithmetic interest such as L-functions.

   In this case, one can at least *hope* for complete, quantitative results via harmonic analysis. In addition to "standard harmonic analysis," one has one important trick in this context: one may make use of *equalities between periods on different groups.* That is to say: often there will be another pair $(\mathbf{H}' \subset \mathbf{G}')$ with the property that, for all φ_i in some suitably chosen automorphic basis of functions on $\mathbf{G}(\mathbf{Q})\backslash\mathbf{G}(\mathbf{A})$, one may associate functions φ_i on $\mathbf{G}'(\mathbf{Q})\backslash\mathbf{G}'(\mathbf{A})$ so that

$$\int_{\mathbf{H}(\mathbf{Q})\backslash\mathbf{H}(\mathbf{A})} \varphi_j = \int_{\mathbf{H}'(\mathbf{Q})\backslash\mathbf{H}'(\mathbf{A})} \varphi_j'$$

   The correspondence $\varphi \leftrightarrow \varphi'$ is usually related to functoriality. Thereby one can study the $\mathbf{H}$-periods on $\mathbf{G}$ by switching to $\mathbf{G}'$.

2. If $\mathbf{H}$ is not a torus, one can almost always profitably apply Ratner's theorem in the ergodic method and get strong, *although non-quantitative* results. We have not discussed any examples of this in the present article; a nice instance (not phrased adelically) is [EO03].

3. If $\mathbf{H}$ is a torus, the emerging theory of measure rigidity for torus actions (see in particular [EL06], [EKL06]) can often substitute for Ratner theory. This requires an extra input, positive entropy, and has two further disadvantages (compared to "Ratner theory") that might be noted:

   (a) At present there is no good general way to control escape of mass in the case when $\mathbf{G}(\mathbf{Q})\backslash\mathbf{G}(\mathbf{A})$ is noncompact, or the related phenomenon of concentration on embedded subgroups.

   (b) One needs to have "Linnik's condition," i.e. a fixed set of places S such that $\mathbf{H}(\mathbf{Q}_v)$ is noncompact for $v \in S$.

As a rough rule, then, the strength of ergodic theory is that it can handle "small" orbits – orbits of very small subgroups which at present seem far beyond the reach of traditional harmonic analysis – and the weakness is that it is not quantitative.

On the other hand, the strength of harmonic analysis is that it imports all the rich internal structure of automorphic forms. Let us give one example: "why" is it that harmonic-analysis approaches to Theorem 1 have been able to avoid a Linnik-type condition? We will give (one, probably contentious) attempt at a philosophical answer to this question. The Waldspurger formula (2.5) expresses a period over a non-split torus $\mathbf{T}_d$ in terms of the L-function $L(\pi, 1/2)L(\pi \times \chi_d, 1/2)$; when d varies, the quantity of interest is $L(\pi \times \chi_d, 1/2)$. But, by Hecke theory, this is expressible as a (χ_d-twisted) period of a form in π over a *split!* torus in $\mathrm{GL}_2(\mathbf{Q})\backslash\mathrm{GL}_2(\mathbf{A}_\mathbf{Q})$. Thereby one has an *equality of periods* between a period over a nonsplit torus $\mathbf{T}_d$ and a twisted period over a split torus $\mathbf{T}_{split}$. It seems to be that this equality of periods, which is perhaps a reflection of functoriality, is part of the reason that one is able to sidestep the problem of small split primes that plagues direct analysis of $\mathbf{T}_d$.

References

[BM03] E. M. Baruch and Z. Mao, *Central value of automorphic $L-$functions.*, Preprint (2003).

[BR99] J. Bernstein and A. Reznikov, *Analytic continuation of representations and estimates of automorphic forms*, Ann. of Math. (2) **150** (1999), no. 1, 329–352.

[BR04] ______, *Estimates of automorphic functions*, Mosc. Math. J. **4** (2004), no. 1, 19–37, 310.

[BR05] ______, *Periods, subconvexity and representation theory*, J. of Diff. Geometry **69** (2005), no. 2. (to appear).

[Blo04a] V. Blomer, *Non-vanishing of class group L-functions at the central point*, Ann. Inst. Fourier (Grenoble) **54** (2004), no. 4, 831–847 (English, with English and French summaries).

[Blo04b] ______, *Shifted convolution sums and subconvexity bounds for automorphic L-functions*, Int. Math. Res. Not. (2004), no. 73, 3905–3926.

[BHM06a] V. Blomer, G. Harcos, and P. Michel, *A Burgess-like subconvex bound for twisted L-functions*, Forum Math. (to appear) (2006). with an appendix by Z. Mao.

[BHM06b] ______, *bounds for automorphic L-functions*, in preparation (2006).

[BK03] J. Bourgain and S. V. Konyagin, *Estimates for the number of sums and products and for exponential sums over subgroups in fields of prime order*, C. R. Math. Acad. Sci. Paris **337** (2003), no. 2, 75–80.

[Byk96] V. A. Bykovskiĭ, *A trace formula for the scalar product of Hecke series and its applications*, Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) **226** (1996), no. Anal. Teor. Chisel i Teor. Funktsii. 13, 14–36, 235–236 (Russian, with Russian summary).

[COU01] L. Clozel, H. Oh, and E. Ullmo, *Hecke operators and equidistribution of Hecke points*, Invent. Math. **144** (2001), no. 2, 327–351.

[CU04] L. Clozel and E. Ullmo, *Équidistribution des points de Hecke*, Contributions to automorphic forms, geometry, and number theory, 2004, pp. 193–254 (French).

[CU05a] ______, *Équidistribution de mesures algébriques*, Compos. Math. **141** (2005), no. 5, 1255–1309.

[CU05b] ______, *Equidistribution de sous-variétes spéciales*, Annals of Math. **161** (2005), no. 3, 1571–1588.

[Cog03] J. W. Cogdell, *On sums of three squares*, J. Théor. Nombres Bordeaux **15** (2003), no. 1, 33–44 (English, with English and French summaries). Les XXIIèmes Journées Arithmetiques (Lille, 2001).

[Coh05] P. B. Cohen, *Hyperbolic equidistribution problems on Siegel 3-folds and Hilbert modular varieties*, Duke Math. J. **129** (2005), no. 1, 87–127.

[CI00] J. B. Conrey and H. Iwaniec, *The cubic moment of central values of automorphic L-functions*, Ann. of Math. (2) **151** (2000), no. 3, 1175–1216.

[CV04] C. Cornut and V. Vatsal, *Nontriviality of Rankin-Selberg L-functions and CM points*, Proceedings of the LMS Durham symposium: L-functions and Galois representations (2004). (to appear).

[Duk88] W. Duke, *Hyperbolic distribution problems and half-integral weight Maass forms*, Invent. Math. **92** (1988), no. 1, 73–90.

[DFI94] W. Duke, J. B. Friedlander, and H. Iwaniec, *A quadratic divisor problem*, Invent. Math. **115** (1994), no. 2, 209–217.

[DFI93] W. Duke, J. Friedlander, and H. Iwaniec, *Bounds for automorphic L-functions*, Invent. Math. **112** (1993), no. 1, 1–8.

[DFI94] ______, *Bounds for automorphic L-functions. II*, Invent. Math. **115** (1994), no. 2, 219–239.

[DFI97] ______, *Bilinear forms with Kloosterman fractions*, Invent. Math. **128** (1997), no. 1, 23–43.

[DFI01] W. Duke, J. B. Friedlander, and H. Iwaniec, *Bounds for automorphic L-functions. III*, Invent. Math. **143** (2001), no. 2, 221–248.

[DFI02] ______, *The subconvexity problem for Artin L-functions*, Invent. Math. **149** (2002), no. 3, 489–577.

[DFI95] W. Duke, J. Friedlander, and H. Iwaniec, *Class group L-functions*, Duke Math. J. **79** (1995), no. 1, 1–56.

[DRS93] W. Duke, Z. Rudnick, and P. Sarnak, *Density of integer points on affine homogeneous varieties*, Duke Math. J. **71** (1993), no. 1, 143–179.

[DSP90] W. Duke and R. Schulze-Pillot, *Representation of integers by positive ternary quadratic forms and equidistribution of lattice points on ellipsoids*, Invent. Math. **99** (1990), no. 1, 49–57.

[Edi05] B. Edixhoven, *Special points on products of modular curves*, Duke Math. J. **126** (2005), no. 2, 325–348.

[EKL06] M. Einsiedler, A. Katok, and E. Lindenstrauss, *Invariant measures and the set of exceptions to Littlewoods conjecture,* Annals of. Math. (2006). (to appear).

[EL06] M. Einsiedler and E. Lindenstrauss, *Diagonalizable flows on locally homogeneous spaces and number theory* (2006). these Proceedings.

[ELMVa] M. Einsiedler, E. Lindenstrauss, Ph. Michel, and A. Venkatesh, *Distribution properties of compact orbits on homonegeous spaces*, in preparation.

[ELMVb] ______, *Distribution properties of compact orbits on homonegeous spaces II*, in preparation.

[EO03] A. Eskin and H. Oh, *Representations of integers by an invariant polynomial and unipotent flows.*, Preprint (2003).

[EM93] A. Eskin and C. McMullen, *Mixing, counting, and equidistribution in Lie groups*, Duke Math. J. **71** (1993), no. 1, 181–209.

[EMS96] A. Eskin, Sh. Mozes, and N. Shah, *Unipotent flows and counting lattice points on homogeneous varieties*, Ann. of Math. (2) **143** (1996), no. 2, 253–299.

[Esk98] A. Eskin, *Counting problems and semisimple groups*, Proceedings of the International Congress of Mathematicians, Vol. II (Berlin, 1998), 1998, pp. 539–552 (electronic).

[Fri95] J. B. Friedlander, *Bounds for L-functions*, Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994), 1995, pp. 363–373.

[GGS02] W. T. Gan, B. Gross, and G. Savin, *Fourier coefficients of modular forms on G_2*, Duke Math. J. **115** (2002), no. 1, 105–169.

[GO03] W. T. Gan and H. Oh, *Equidistribution of integer points on a family of homogeneous varieties: a problem of Linnik*, Compositio Math. **136** (2003), no. 3, 323–352.

[GS03] W. T. Gan and G. Savin, *Real and global lifts from* PGL_3 *to* G_2, Int. Math. Res. Not. (2003), no. 50, 2699–2724.

[GS04] ———, *Endoscopic lifts from* PGL_3 *to* G_2, Compos. Math. **140** (2004), no. 3, 793–808.

[Gro87] B. Gross, *Heights and the special values of L-series*, Number theory (Montreal, Que., 1985), 1987, pp. 115–187.

[GZ86] B. Gross and D. Zagier, *Heegner points and derivatives of L-series*, Invent. Math. **84** (1986), no. 2, 225–320.

[Har03] G. Harcos, *An additive problem in the Fourier coefficients of cusp forms*, Math. Ann. **326** (2003), no. 2, 347–365.

[HM06] G. Harcos and P. Michel, *The subconvexity problem for Rankin-Selberg L-functions and equidistribution of Heegner points II*, Invent. Math. (2006). (to appear).

[HK91] M. Harris and S. S. Kudla, *The central critical value of a triple product L-function*, Ann. of Math. (2) **133** (1991), no. 3, 605–672.

[Hat90] R. L. Hatcher, *Heights and L-series*, Canad. J. Math. **42** (1990), no. 3, 533–560.

[Ivi01] A. Ivić, *On sums of Hecke series in short intervals*, J. Théor. Nombres Bordeaux **13** (2001), no. 2, 453–468.

[Iwa87] H. Iwaniec, *Fourier coefficients of modular forms of half-integral weight*, Invent. Math. **87** (1987), no. 2, 385–401.

[Iwa92] ———, *The spectral growth of automorphic L-functions*, J. Reine Angew. Math. **428** (1992), 139–159.

[Iwa99] ———, *Harmonic analysis in number theory*, Prospects in mathematics (Princeton, NJ, 1996), 1999, pp. 51–68.

[IS00] H. Iwaniec and P. Sarnak, *Perspectives on the analytic theory of L-functions*, Geom. Funct. Anal. (2000), no. Special Volume, 705–741. GAFA 2000 (Tel Aviv, 1999).

[Jut99] M. Jutila, *Convolutions of Fourier coefficients of cusp forms*, Publ. Inst. Math. (Beograd) (N.S.) **65(79)** (1999), 31–51.

[JM05] M. Jutila and Y. Motohashi, *Uniform bounds for Hecke L-functions*, Preprint (2005).

[KS93] S. Katok and P. Sarnak, *Heegner points, cycles and Maass forms*, Israel J. Math. **84** (1993), no. 1-2, 193–227.

[Kim03] H. Kim, *Functoriality for the exterior square of* GL_4 *and the symmetric fourth of* GL_2, J. Amer. Math. Soc. **16** (2003), no. 1, 139–183 (electronic). With appendix 1 by Dinakar Ramakrishnan and appendix 2 by Kim and Peter Sarnak.

[KS02] Henry H. Kim and Freydoon Shahidi, *Functorial products for* $\mathrm{GL}_2 \times \mathrm{GL}_3$ *and the symmetric cube for* GL_2, Ann. of Math. (2) **155** (2002), no. 3, 837–893. With an appendix by Colin J. Bushnell and Guy Henniart.

[KZ81] W. Kohnen and D. Zagier, *Values of L-series of modular forms at the center of the critical strip*, Invent. Math. **64** (1981), no. 2, 175–198.

[KMV00] E. Kowalski, P. Michel, and J. VanderKam, *Mollification of the fourth moment of automorphic L-functions and arithmetic applications*, Invent. Math. **142** (2000), no. 1, 95–151.

[KMV02] E. Kowalski, Ph. Michel, and J. VanderKam, *Rankin-Selberg L-functions in the level aspect*, Duke Math. J. **114** (2002), no. 1, 123–191.

[Lin06] E. Lindenstrauss, *Invariant measures and arithmetic quantum unique ergodicity*, Annals of. Math. (2006). (to appear).

[Lin] ———, *Rigidity of multiparameter actions*, `math.DS/0402165`.

[Lin55] Yu. V. Linnik, *The asymptotic distribution of reduced binary quadratic forms in relation to the geometries of Lobačevskiĭ. III*, Vestnik Leningrad. Univ. **10** (1955), no. 8, 15–27 (Russian).

[Lin60] Y. V. Linnik, *Asymptotic-geometric and ergodic properties of sets of lattice points on a sphere*, Amer. Math. Soc. Transl. (2) **13** (1960), 9–27.

[Lin63] Yu. V. Linnik, *Additive problems and eigenvalues of the modular operators*, Proc. Internat. Congr. Mathematicians (Stockholm, 1962), 1963, pp. 270–284.

[Lin68] ———, *Ergodic properties of algebraic fields*, Translated from the Russian by M. S. Keane. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 45, Springer-Verlag New York Inc., New York, 1968.

[LS64] Yu. V. Linnik and B. F. Skubenko, *Asymptotic distribution of integral matrices of third order*, Vestnik Leningrad. Univ. Ser. Mat. Meh. Astronom. **19** (1964), no. 3, 25–36.

[Mar00] G. Margulis, *Problems and conjectures in rigidity theory*, Mathematics: frontiers and perspectives, 2000, pp. 161–174.

[Mic04] Ph. Michel, *The subconvexity problem for Rankin-Selberg L-functions and equidistribution of Heegner points*, Ann. of Math. (2) **160** (2004), no. 1, 185–236.

[MV] Ph. Michel and A. Venkatesh, *Periods, subconvexity and equidistribution*, in preparation.

[MV06] ———, *Heegner points and nonvanishing of Rankin-Selberg L-functions*, Preprint (2006).

[Mol02] G. Molteni, *Upper and lower bounds at $s = 1$ for certain Dirichlet series with Euler product*, Duke Math. J. **111** (2002), no. 1, 133–158.

[Mot04] 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, 28–33.

[Mot97] ———, *Spectral theory of the Riemann zeta-function*, Cambridge Tracts in Mathematics, Cambridge University Press, 1997.

[Pop06] A. Popa, *Central values of Rankin L-series over real quadratic fields*, Compositio Math. (2006). (to appear).

[Rez05] A. Reznikov, *Rankin-Selberg without unfolding*, preprint (2005).

[Sar91] P. Sarnak, *Diophantine problems and linear groups*, Proceedings of the International Congress of Mathematicians, Vol. I, II (Kyoto, 1990), 1991, pp. 459–471.

[Sar94] ———, *Integrals of products of eigenfunctions*, Internat. Math. Res. Notices (1994), no. 6, 251 ff., approx. 10 pp. (electronic).

[Sar01] ———, *Estimates for Rankin-Selberg L-functions and quantum unique ergodicity*, J. Funct. Anal. **184** (2001), no. 2, 419–453.

[Sku62] B. F. Skubenko, *The asymptotic distribution of integers on a hyperboloid of one sheet and ergodic theorems*, Izv. Akad. Nauk SSSR Ser. Mat. **26** (1962), 721–752.

[Ull02] E. Ullmo, *Théorie ergodique et géométrie arithmétique*, Proceedings of the International Congress of Mathematicians, Vol. II (Beijing, 2002), 2002, pp. 197–206.

[Vat02] V. Vatsal, *Uniform distribution of Heegner points*, Invent. Math. **148** (2002), no. 1, 1–46.

[Vat03] ———, *Special values of anticyclotomic L-functions*, Duke Math. J. **116** (2003), no. 2, 219–261.

[Ven05] A. Venkatesh, *Sparse equidistribution problems, period bounds, and subconvexity.*, Preprint (2005).

[Wal85] J.-L. Waldspurger, *Sur les valeurs de certaines fonctions L automorphes en leur centre de symétrie*, Compositio Math. **54** (1985), no. 2, 173–242.

[Wat06] T. Watson, *Rankin triple products and quantum chaos*, Annals of Math. (2006). to appear.

[Xue05a] H. Xue, *Central values of Rankin L-functions*, Preprint (2005).

[Xue05b] ———, *Central values of L-functions over CM fields*, Preprint (2005).

[Yaf03] A. Yafaev, *A conjecture of Yves André*, Preprint (2003).

[Zha01a] S. Zhang, *Heights of Heegner points on Shimura curves*, Ann. of Math. (2) **153** (2001), no. 1, 27–147.

[Zha01b] ———, *Gross-Zagier formula for* GL_2, Asian J. Math. **5** (2001), no. 2, 183–290.

[Zha04] ———, *Equidistribution of CM-points on quaternion Shimura varieties*, Preprint (2004).

Département de Mathématiques, Univ. Montpellier II Place E. Bataillon, F-34095 Montpellier cdx
E-mail: michel@math.univ-montp2.fr

Courant Institute of Mathematical Sciences, 251 Mercer Street, New York, NY 10012-1185, USA
E-mail: venkatesh@math.nyu.edu