Clay Mathematics Proceedings
Volume 10, 2010

Equidistribution on the modular surface and L-functions

Gergely Harcos

ABSTRACT. These are notes for two lectures given at the 2007 summer school "Homogeneous Flows, Moduli Spaces and Arithmetic" in Pisa, Italy. The first lecture introduces Heegner points and closed geodesics on the modular surface $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$ and highlights some of their arithmetic significance. The second lecture discusses how subconvex bounds for certain automorphic L-functions yield quantitative equidistribution results for Heegner points and closed geodesics.

1. Lecture One

Let us start the discussion with the equivalence of integral binary quadratic forms. The concept was introduced by Lagrange [**15**] and studied by Gauss [**9**] in a systematic fashion.

An *integral binary quadratic form* is a homogeneous polynomial

$$\langle a,b,c\rangle := ax^2+bxy+cy^2 \in \mathbb{Z}[x,y]$$

with associated *discriminant*

$$d := b^2-4ac \in \mathbb{Z}.$$

The possible discriminants are the integers congruent to 0 or 1 mod 4. We shall assume that the form $\langle a,b,c\rangle$ is not a product of linear factors in $\mathbb{Z}[x,y]$, then d is not a square, hence $ac \neq 0$. If $d<0$ then $ac>0$ and we shall assume that we are in the *positive definite* case $a,c>0$. Furthermore, we shall assume that d is a *fundamental discriminant* which means that it cannot be written as $d'e^2$ for some smaller discriminant d'. Then $\langle a,b,c\rangle$ is a *primitive* form which means that a,b,c are relatively prime. The possible fundamental discriminants are the square-free numbers congruent to 1 mod 4 and 4 times the square-free numbers congruent to 2 or 3 mod 4.

EXAMPLE 1. The first few negative fundamental discriminants are: -3, -4, -7, -8, -11, -15, -19, -20, -23, -24. The first few positive fundamental discriminants are: 5, 8, 12, 13, 17, 21, 24, 28, 29, 33.

The author was supported by European Community grant MEIF-CT-2006-040371 under the Sixth Framework Programme.

Lagrange [**15**] discovered that every form $\langle a,b,c\rangle$ with a given discriminant d can be reduced by some integral unimodular substitution

$$(x,y)\mapsto(\alpha x+\beta y,\gamma x+\delta y),\qquad \begin{pmatrix}\alpha&\beta\\ \gamma&\delta\end{pmatrix}\in\mathrm{SL}_2(\mathbb{Z}),$$

to some form with the same discriminant that lies in a finite set depending only on d. Forms that are connected by such a substitution are called *equivalent.* It is easiest to understand this reduction by looking at the simple substitutions

$$(x,y)\overset{T}{\mapsto}(x-y,y)\qquad\text{and}\qquad (x,y)\overset{S}{\mapsto}(-y,x). \tag{1}$$

The induced actions on forms are given by

$$\langle a,b,c\rangle\overset{T}{\mapsto}\langle a,b-2a,c+a-b\rangle\qquad\text{and}\qquad \langle a,b,c\rangle\overset{S}{\mapsto}\langle c,-b,a\rangle.$$

Now a given form $\langle a,b,c\rangle$ can always be taken to some $\langle a,b',c'\rangle$ with $|b'|\leqslant|a|$ by applying T or T^{-1} a few times. If $|a|\leqslant|c'|$ then we stop our reduction. Otherwise we apply S to get some $\langle a'',b'',c''\rangle$ with $|a''|<|a|$ and we start over with this form. In this algorithm we cannot apply S infinitely many times because $|a|$ decreases at each such step. Hence in a finite number of steps we arrive at an equivalent form $\langle a,b,c\rangle$ whose coefficients satisfy

$$|b|\leqslant|a|\leqslant|c|,\qquad b^2-4ac=d. \tag{2}$$

These constraints are satisfied by finitely many triples (a,b,c). Indeed, we have

$$|d|=|b^2-4ac|\geqslant4|ac|-b^2\geqslant3b^2, \tag{3}$$

so there are only $\ll|d|^{1/2}$ choices for b and for each such choice there are only $\ll_\varepsilon d^\varepsilon$ choices for a and c since the product ac is determined by b. We have shown that the number of equivalence classes of integral binary quadratic forms of fundamental discriminant d, denoted $h(d)$, satisfies the inequality

$$h(d)\ll_\varepsilon|d|^{1/2+\varepsilon}. \tag{4}$$

In the case $d<0$ it is straightforward to compile a maximal list of inequivalent forms satisfying (2). There is an algorithm for $d>0$ as well but it is less straightforward. In fact the subsequent findings of this lecture can be turned into an algorithm for all d. Note that for $d>0$ (3) implies $4ac=b^2-d<0$, hence by an extra application of S we can always arrange for a reduced form $\langle a,b,c\rangle$ with $a>0$.

Example 2. The equivalence classes for $d=-23$ are represented by the forms $\langle1,1,6\rangle$, $\langle2,\pm1,3\rangle$. Hence $h(-23)=3$. The equivalence classes for $d=21$ are represented by the forms $\langle1,1,-5\rangle$, $\langle-1,1,5\rangle$. Hence $h(21)=2$.

To obtain a geometric picture of equivalence classes of forms we shall think of $\mathbb{Q}(\sqrt{d})$ as embedded in $\mathbb{C}$ such that $\sqrt{d}/i>0$ for $d<0$ and $\sqrt{d}>0$ for $d>0$. For $q_1,q_2\in\mathbb{Q}$ we shall consider the conjugation

$$\overline{q_1+q_2\sqrt{d}}:=q_1-q_2\sqrt{d}.$$

Each form $\langle a,b,c\rangle$ decomposes as

$$ax^2+bxy+cy^2=a(x-zy)(x-\bar{z}y),$$

where

$$z := \frac{-b+\sqrt{d}}{2a}, \qquad \bar{z} := \frac{-b-\sqrt{d}}{2a}. \tag{5}$$

Using (1) we can see that the action of $\mathrm{SL}_2(\mathbb{Z})$ on z and $\bar{z}$ is the usual one given by fractional linear transformations:

$$z \overset{T}{\mapsto} z+1 \qquad \text{and} \qquad z \overset{S}{\mapsto} -1/z.$$

Therefore in fact we are looking at the standard action of $\mathrm{SL}_2(\mathbb{Z})$ on certain conjugate pairs of points of $\mathbb{Q}(\sqrt{d})$ embedded in $\mathbb{C}$. For $d<0$ we consider the points $z \in \mathcal{H}$ and obtain $h(d)$ points on $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$. These are the *Heegner points* of discriminant $d<0$. For $d>0$ we consider the geodesics $G_{\bar{z},z} \subset \mathcal{H}$ connecting the real points $\{\bar{z}, z\}$ and obtain $h(d)$ geodesics on $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$.

It is a remarkable fact that for $d>0$ any geodesic $G_{\bar{z},z}$ as above becomes closed when projected to $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$, and its length is an important arithmetic quantity associated with the number field $\mathbb{Q}(\sqrt{d})$. To see this take any matrix $M \in \mathrm{GL}_2^+(\mathbb{R})$ which takes 0 to $\bar{z}$ and ∞ to z, for example[1]

$$M := \begin{pmatrix} z & \bar{z} \\ 1 & 1 \end{pmatrix}, \tag{6}$$

then M takes the positive real axis (resp. geodesic) connecting $\{0,\infty\}$ to the real segment (resp. geodesic) connecting $\{\bar{z}, z\}$. In particular, using that M is a conformal automorphism of the Riemann sphere, we see that $G_{\bar{z},z}$ is the semicircle above the real segment $[\bar{z}, z]$, parametrized as

$$G_{\bar{z},z} = \{g(\lambda)i : \ \lambda > 0\}, \qquad \text{where} \qquad g(\lambda) := M \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix}.$$

Moreover, the unique isometry of $\mathcal{H}$ fixing the geodesic $G_{\bar{z},z}$ and taking $g(1)i$ to $g(\lambda)i$ is given by the matrix

$$M \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix} M^{-1} \in \mathrm{SL}_2(\mathbb{R}). \tag{7}$$

Therefore we want to see that for some $\lambda > 1$ the matrix

$$M \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix} M^{-1} = \frac{1}{z-\bar{z}} \begin{pmatrix} z\lambda - \bar{z}\lambda^{-1} & z\bar{z}(\lambda^{-1}-\lambda) \\ \lambda - \lambda^{-1} & z\lambda^{-1} - \bar{z}\lambda \end{pmatrix} \tag{8}$$

is in $\mathrm{SL}_2(\mathbb{Z})$, and then the projection of $G_{\bar{z},z}$ to $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$ has length

$$\int_1^{\lambda^2} \frac{dy}{y} = 2\ln(\lambda)$$

for the smallest such $\lambda > 1$. A necessary condition for λ is that the sum and difference of diagonal elements of the matrix (8) are integers and so are the anti-diagonal elements as well. Using that

$$z - \bar{z} = \frac{\sqrt{d}}{a}, \qquad z + \bar{z} = \frac{-b}{a}, \qquad z\bar{z} = \frac{c}{a}$$

this is equivalent to:

$$\lambda + \lambda^{-1} \in \mathbb{Z}, \qquad \{a,b,c\}\frac{\lambda - \lambda^{-1}}{\sqrt{d}} \subset \mathbb{Z}.$$

[1] we assume here that $a>0$ which is legitimate as we have seen

As $\gcd(a,b,c)=1$ we can simplify this to

$$\lambda+\lambda^{-1}\in\mathbb{Z},\qquad\text{and}\qquad\frac{\lambda-\lambda^{-1}}{\sqrt{d}}\in\mathbb{Z}.$$

In other words, there are integers m,n such that

$$\lambda=\frac{m+n\sqrt{d}}{2}\qquad\text{and}\qquad\lambda^{-1}=\frac{m-n\sqrt{d}}{2}. \tag{9}$$

As $\lambda>1$ the integers m,n are positive and they satisfy the diophantine equation

$$m^2-dn^2=4. \tag{10}$$

The equations (9)–(10) are not only necessary but also sufficient for (8) to lie in $\mathrm{SL}_2(\mathbb{Z})$. Namely, (8)–(10) imply that

$$M\begin{pmatrix}\lambda & 0\\ 0 & \lambda^{-1}\end{pmatrix}M^{-1}=\begin{pmatrix}\frac{m-bn}{2} & -nc\\ na & \frac{m+bn}{2}\end{pmatrix}\in\mathrm{SL}_2(\mathbb{Z}) \tag{11}$$

since

$$m\pm bn\equiv m^2-dn^2\equiv 0 \pmod 2.$$

The λ's given by (9)–(10) are exactly the totally positive[2] units in the ring of integers $\mathcal{O}_d$ of $\mathbb{Q}(\sqrt{d})$. These units form a group isomorphic to $\mathbb{Z}$ by Dirichlet's theorem, therefore there is a smallest $\lambda=\lambda_d>1$ among them (which generates the group). In other words, the sought $\lambda=\lambda_d>1$ exists and comes from the smallest positive solution of (10). In classical language, the matrices (11) are the *automorphs* of the form $\langle a,b,c\rangle$.

To summarize, the $\mathrm{SL}_2(\mathbb{Z})$-orbits of forms $\langle a,b,c\rangle$ with given fundamental discriminant d give rise to $h(d)$ Heegner points on $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$ for $d<0$ and $h(d)$ closed geodesics of length $2\ln(\lambda_d)$ for $d>0$ where $\lambda_d=(m+n\sqrt{d})/2$ is the smallest totally positive unit of $\mathcal{O}_d$ greater than 1. This geometric picture is even more interesting in the light of the following refinement of (4) which is a consequence of Dirichlet's class number formula and Siegel's theorem (see [**5**, Chapters 6 and 21]):

$$\begin{aligned} |d|^{1/2-\varepsilon}\ll_\varepsilon h(d)\ll_\varepsilon |d|^{1/2+\varepsilon}, &\qquad d<0,\\ d^{1/2-\varepsilon}\ll_\varepsilon h(d)\ln(\lambda_d)\ll_\varepsilon d^{1/2+\varepsilon}, &\qquad d>0. \end{aligned} \tag{12}$$

This shows that the set of Heegner points of discriminant $d<0$ has cardinality about $|d|^{1/2}$, while the set of closed geodesics of discriminant $d>0$ has total length about $d^{1/2}$.

2. Lecture Two

In the light of (12) the natural question arises if the set Λ_d of Heegner points (resp. closed geodesics) of fundamental discriminant d becomes equidistributed in $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$ as $d\to-\infty$ (resp. $d\to+\infty$). That is, given a smooth and compactly supported weight function $g:\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}\to\mathbb{C}$ do we have

$$\begin{aligned} \frac{1}{h(d)}\sum_{z\in\Lambda_d}g(z)\to\int_{\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}}g(z)\,d\mu(z), &\qquad d\to-\infty,\\ \frac{1}{h(d)\,2\ln(\lambda_d)}\sum_{G\in\Lambda_d}\int_G g(z)\,ds(z)\to\int_{\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}}g(z)\,d\mu(z), &\qquad d\to+\infty, \end{aligned} \tag{13}$$

[2] i.e. positive under both embeddings $\mathbb{Q}(\sqrt{d})\hookrightarrow\mathbb{R}$

where $d\mu(z)$ abbreviates the $\mathrm{SL}_2(\mathbb{R})$-invariant probability measure on $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$ and $ds(z)$ abbreviates the hyperbolic arc length? Duke [**6**] proved that the answer is yes in the sharper form that the difference of the two sides is $\ll_g |d|^{-\delta}$ for some fixed $\delta > 0$. Earlier Linnik [**16**] established the above limits with error term $\ll_g (\log|d|)^{-A}$ for all $A > 0$ under the condition that $\left(\frac{d}{p}\right) = 1$ for a fixed odd prime p.

We shall discuss Duke's quantitative result and a refinement of it from the modern perspective of subconvex bounds for automorphic L-functions. Our first step is to decompose spectrally the weight function considered in (13) as

$$g(z) = \langle g, 1\rangle + \sum_{j=1}^{\infty}\langle g, u_j\rangle u_j(z) + \frac{1}{4\pi}\int_{-\infty}^{\infty}\langle g, E(\cdot, \tfrac{1}{2}+it)\rangle E(z, \tfrac{1}{2}+it)\,dt,$$

where

$$\langle f_1, f_2\rangle := \int_{\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}} f_1(z)\overline{f_2(z)}\,d\mu(z),$$

the $\{u_j\}$ are Hecke–Maass cusp forms on $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$ with $\langle u_j, u_j\rangle = 1$, and the Eisenstein series $E(z, \frac{1}{2}+it)$ are obtained by meromorphic continuation from

$$E(z,s) := \frac{1}{2}\sum_{\substack{m,n\in\mathbb{Z}\\ \gcd(m,n)=1}}\frac{\Im z^s}{|mz+n|^{2s}}, \qquad \Re s > 1.$$

The above decomposition converges in $L^2(\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H})$ and also pointwise absolutely and uniformly on compact sets, see [**14**, Theorem 7.3]. If

$$\Delta := -y^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)$$

denotes the hyperbolic Laplacian and we use the notation and fact

$$\Delta u_j(z) = \left(\tfrac{1}{4}+t_j^2\right)u_j(z), \qquad \Delta E(z, \tfrac{1}{2}+it) = \left(\tfrac{1}{4}+t^2\right)E(z, \tfrac{1}{2}+it),$$

then for any smooth and compactly supported $g(z)$ and for any $B > 0$ we have

$$\langle g, u_j\rangle \ll_{g,B} (1+|t_j|)^{-B}, \qquad \langle g, E(\cdot, \tfrac{1}{2}+it)\rangle \ll_{g,B} (1+|t|)^{-B}. \tag{14}$$

Therefore in order to establish Duke's theorem with an error term $\ll_g |d|^{-\delta'}$ it suffices to show that if g is a Hecke–Maass cusp form with $\langle g, g\rangle = 1$ or a standard Eisenstein series $E(\cdot, \frac{1}{2}+it)$ then for some fixed $\delta > 0$ and $A > 0$ the sums considered in (13) satisfy

$$\sum_{\Lambda_d}\cdots \ll (1+|t|)^A\,|d|^{\frac{1}{2}-\delta}, \tag{15}$$

where $t = t_g$ is the spectral parameter of g, i.e.

$$\Delta g(z) = \left(\tfrac{1}{4}+t^2\right)g(z).$$

At this point we remark that any such g has a Fourier decomposition of the form

$$g(x+iy) = c_1 y^{\frac{1}{2}+it} + c_2 y^{\frac{1}{2}-it} + \sqrt{y}\sum_{n\neq 0}\rho_g(n)K_{it}(2\pi|n|y)e^{2\pi i n x},$$

where $c_{1,2}$ are some constants[3] and K_{it} is a Bessel function. The Fourier coefficients $\rho_g(n)$ are proportional to the Hecke eigenvalues of g, and by a result of Hoffstein–Lockhart [**10**] we have the uniform bound (see also [**14**, (3.25)])

$$|\rho_g(1)| \ll_\varepsilon (1+|t|)^\varepsilon e^{\frac{\pi}{2}|t|}. \tag{16}$$

We note that for t bounded away from zero we have a similar lower bound, with exponent $-\varepsilon$ in place of ε, as proved by Iwaniec [**13**] (see also [**14**, Theorem 8.3]).

Now we state a formula which can be attributed to several people[4] and relates the sums in (15) to central values of automorphic L-functions:

$$\left|\sum_{\Lambda_d}\dots\right|^2 = c_d\,|d|^{\frac12}\,|\rho_g(1)|^2\,\Lambda\left(\tfrac12,g\right)\Lambda\left(\tfrac12,g\otimes\left(\tfrac{d}{\cdot}\right)\right), \tag{17}$$

where the factor c_d is positive and takes only finitely many different values. In this formula $\Lambda(s,\Pi)$ denotes the *completed* L-function; the finite part $L(s,\Pi)$ of the L-function is defined in terms of Hecke eigenvalues; the infinite part of the L-function is a product of exponential and gamma factors whose contribution in (17) is $\ll (1+|t|)e^{-\pi|t|}$ by Stirling's approximation. Using also (16) we conclude that (15) follows by a subconvex bound of the form

$$L\left(\tfrac12,g\otimes\left(\tfrac{d}{\cdot}\right)\right) \ll (1+|t|)^A\,|d|^{\frac12-\delta},$$

where $\delta>0$ and $A>0$ are some fixed constants (different from those in (15)). In the case when g is a cusp form such a bound was proved by Duke–Friedlander–Iwaniec [**7**] for any $\delta<\frac{1}{22}$, by Bykovskiĭ [**3**] for any $\delta<\frac18$, and by Conrey–Iwaniec [**4**] for any $\delta<\frac16$. In the case when g is an Eisenstein series $E(\cdot,\frac12+it)$ the above becomes

$$\left|L\left(\tfrac12+it,\left(\tfrac{d}{\cdot}\right)\right)\right|^2 \ll (1+|t|)^A\,|d|^{\frac12-\delta},$$

and this was established by Burgess [**2**] for any $\delta<\frac18$, and by Conrey–Iwaniec [**4**] for any $\delta<\frac16$.

We shall now formulate a refinement of (13) using the natural action of the narrow ideal class group H_d of $\mathbb{Q}(\sqrt{d})$ on Λ_d. This action comes from the natural bijection $H_d\leftrightarrow\Lambda_d$ which we describe in the Appendix. Note in particular that $|H_d|=h(d)$ by this bijection. Given some $z_0\in\Lambda_d$ when $d<0$ and some $G_0\in\Lambda_d$ when $d>0$, and given some subgroup $H\leqslant H_d$ one can ask if

$$\begin{aligned}\frac{1}{|H|}\sum_{\sigma\in H} g(z_0^\sigma) &\to \int_{\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}} g(z)\,d\mu(z), \qquad d\to-\infty,\\ \frac{1}{|H|\,2\ln(\lambda_d)}\sum_{\sigma\in H}\int_{G_0^\sigma} g(z)\,ds(z) &\to \int_{\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}} g(z)\,d\mu(z), \qquad d\to+\infty.\end{aligned} \tag{18}$$

Using characters of the abelian group H_d we can decompose the sums over H into twisted sums over H_d:

$$\sum_{\sigma\in H}\dots = \sum_{\sigma\in H_d}\frac{1}{(H_d:H)}\sum_{\substack{\psi\in\widehat{H}_d\\ \psi_{|H}\equiv1}}\psi(\sigma)\dots = \frac{|H|}{|H_d|}\sum_{\substack{\psi\in\widehat{H}_d\\ \psi_{|H}\equiv1}}\sum_{\sigma\in H_d}\psi(\sigma)\dots$$

[3] $c_1=c_2=0$ if g is a cusp form, $c_1=|c_2|=1$ if g is an Eisenstein series $E(\cdot,\frac12+it)$

[4] Dirichlet, Hecke, Siegel, Maass, Shimura, Waldspurger, Kohnen–Zagier, Duke, Katok–Sarnak, Guo, Zhang, Popa; see the references for (20) of which (17) is a special case

Note that the number of characters of H_d restricting to the identity character on H is $(H_d : H)$. Therefore if we have, uniformly for all characters $\psi : H_d \to \mathbb{C}^\times$ and for all L^2-normalized Hecke–Maass cusp forms or standard Eisenstein series in the role of g,

$$\begin{aligned} \sum_{\sigma\in H_d} \psi(\sigma) g(z_0^\sigma) &\ll (1+|t|)^A\, |d|^{\frac12-\delta}, \qquad d<0,\\ \sum_{\sigma\in H_d} \psi(\sigma)\int_{G_0^\sigma} g(z)\,ds(z) &\ll (1+|t|)^A\, |d|^{\frac12-\delta}, \qquad d>0, \end{aligned} \tag{19}$$

where $\delta > 0$ and $A > 0$ are fixed constants, then by the same discussion as above, the limits (18) follow with a strong error term $\ll_g |d|^{-\delta'}$ as long as

$$(H_d : H) \ll |d|^\eta$$

for any fixed constant $0 < \eta < \delta$.

The twisted sums in (19) can be related to central automorphic L-values similarly as in (17). The formula is based on the deep work of Waldspurger [**21**] and was carefully derived by Zhang [**22**] when $d < 0$ and by Popa [**19**] when $d > 0$:

$$\left|\sum_{\sigma\in H_d} \overline{\psi(\sigma)}\dots\right|^2 = c_d\, |d|^{\frac12}\, |\rho_g(1)|^2\, \Lambda\left(\tfrac12, g\otimes f_\psi\right). \tag{20}$$

Here f_ψ is the so-called Jacquet–Langlands lift of ψ, discovered by Hecke [**12**] and Maass [**17**] in this special case: it is a modular form on $\mathcal{H}$ of level $|d|$ and nebentypus $\left(\frac{d}{\cdot}\right)$ with the same completed L-function as ψ. In particular, when g is an Eisenstein series $E(\cdot, \frac12 + it)$ the identity (20) follows from [**20**, pp. 70 and 88] and [**14**, (3.25)].

If the character $\psi : H_d \to \mathbb{C}^\times$ is real-valued then it is one of the genus characters discovered by Gauss [**9**]. In this case, as observed by Kronecker [**20**, p. 62],

$$\Lambda(s,\psi) = \Lambda(s, (\tfrac{d_1}{\cdot}))\Lambda(s, (\tfrac{d_2}{\cdot})),$$

where $d = d_1 d_2$ is a factorization of d into fundamental discriminants d_1 and d_2, whence (20) simplifies to

$$\left|\sum_{\sigma\in H_d} \overline{\psi(\sigma)}\dots\right|^2 = c_d\, |d|^{\frac12}\, |\rho_g(1)|^2\, \Lambda\left(\tfrac12, g\otimes(\tfrac{d_1}{\cdot})\right)\Lambda\left(\tfrac12, g\otimes(\tfrac{d_2}{\cdot})\right).$$

In fact (17) is the special case of this formula when ψ is the trivial character ($d_1 = 1$, $d_2 = d$). The necessary estimate (19) follows by the subconvex bounds discussed before:

$$L\left(\tfrac12, g\otimes(\tfrac{d_i}{\cdot})\right) \ll (1+|t|)^A\, |d_i|^{\frac12-\delta}, \qquad i = 1,2.$$

If the character $\psi : H_d \to \mathbb{C}^\times$ is not real-valued then f_ψ is a cusp form of level $|d|$ and nebentypus $\left(\frac{d}{\cdot}\right)$, and we need a subconvex bound of the form

$$L\left(\tfrac12, g\otimes f_\psi\right) \ll (1+|t|)^A\, |d|^{\frac12-\delta}.$$

In the case when g is a cusp form such a bound was proved by Harcos–Michel [**11**] with $\delta = \frac{1}{3000}$. In the case when g is an Eisenstein series $E(\cdot, \frac12 + it)$ the above becomes

$$\left|L\left(\tfrac12 + it, \psi\right)\right|^2 \ll (1+|t|)^A\, |d|^{\frac12-\delta},$$

and this was established by Duke–Friedlander–Iwaniec [**8**] with $\delta = \frac{1}{12000}$ and by Blomer–Harcos–Michel [**1**] with $\delta = \frac{1}{1000}$.

Finally we remark that the above ideas have been greatly extended by several researchers. The interested reader should consult the excellent survey of Michel–Venkatesh [**18**].

3. Appendix

In this Appendix we consider an arbitrary fundamental discriminant d and regard $\sqrt{d}$ as a complex number which lies on the positive real axis or positive imaginary axis depending on the sign of d. We show that the equivalence classes of forms of fundamental discriminant d can be mapped bijectively to narrow ideal classes of the quadratic number field $\mathbb{Q}(\sqrt{d})$ in a natural fashion. As the latter classes form an abelian group under multiplication this will exhibit a natural multiplication law on the equivalence classes of forms. This law, discovered by Gauss [**9**], is called *composition* in the classical theory.

Recall that a fractional ideal of $\mathbb{Q}(\sqrt{d})$ is a finitely generated $\mathcal{O}_d$-module contained in $\mathbb{Q}(\sqrt{d})$ and two nonzero fractional ideals are equivalent (in the narrow sense) if their quotient is a principal fractional ideal generated by a totally positive element of $\mathbb{Q}(\sqrt{d})$. Here "totally positive element" can clearly be changed to "element of positive norm" where the norm of $\mu \in \mathbb{Q}(\sqrt{d})$ is given by $N(\mu) = \mu\bar{\mu}$. Recall also that we can represent equivalence classes of forms of fundamental discriminant d by some

$$Q_i(x,y) = a_i x^2 + b_i xy + c_i y^2 = a_i(x - z_i y)(x - \bar{z}_i y), \qquad i = 1, \ldots, h(d),$$

with

$$a_i > 0, \qquad z_i := \frac{-b_i + \sqrt{d}}{2a_i}, \qquad \bar{z}_i := \frac{-b_i - \sqrt{d}}{2a_i}.$$

It will suffice to show that each fractional ideal I of $\mathbb{Q}(\sqrt{d})$ is equivalent to some fractional ideal

$$I_i := \mathbb{Z} + \mathbb{Z}z_i, \qquad i = 1, \ldots, h(d),$$

and that the fractional ideals I_i are pairwise inequivalent.

Any fractional ideal I can be written as

$$I = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2 \qquad \text{with} \qquad \frac{\bar{\omega}_1\omega_2 - \omega_1\bar{\omega}_2}{\sqrt{d}} > 0.$$

We associate to I (and ω_1, ω_2) the binary quadratic form

$$Q_I(x,y) := \frac{(x\omega_1 - y\omega_2)(x\bar{\omega}_1 - y\bar{\omega}_2)}{N(I)},$$

where $N(I) > 0$ is the absolute norm of I, i.e. the multiplicative function that agrees with $(\mathcal{O}_d : I)$ for integral ideals I. We claim first that $Q_I(x,y)$ has integral coefficients and discriminant d. To see the claim we can assume that I is an integral ideal since $Q_I(x,y)$ does not change if we replace I by nI (and ω_i by $n\omega_i$) for some positive integer n. Then ω_1, ω_2 and their conjugates are in $\mathcal{O}_d$ and the claim amounts to:

- $N(I) \mid \omega_1\bar{\omega}_1,\ \omega_1\bar{\omega}_2 + \bar{\omega}_1\omega_2,\ \omega_2\bar{\omega}_2$;
- $(\omega_1\bar{\omega}_2 - \bar{\omega}_1\omega_2)^2 = N(I)^2 d$.

The first statement follows from the fact that $\omega_1, \omega_2, \omega_1 + \omega_2$ are elements of I, hence their norms are divisible by $N(I)$. The second statement follows by writing $\mathcal{O}_d$ as $\mathbb{Z} + \mathbb{Z}\omega$ and then noting that

$$\begin{vmatrix} \omega_1 & \bar{\omega}_1 \\ \omega_2 & \bar{\omega}_2 \end{vmatrix}^2 = (\mathcal{O}_d : I)^2 \begin{vmatrix} 1 & 1 \\ \omega & \bar{\omega} \end{vmatrix}^2 = N(I)^2 d.$$

The claim implies that there is a unique i and a unique $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$ such that

$$Q_I(\alpha x + \beta y, \gamma x + \delta y) = Q_i(x, y).$$

We can write this as

$$\frac{N(\alpha\omega_1 - \gamma\omega_2)}{N(I)}(x - zy)(x - \bar{z}y) = a_i(x - z_i y)(x - \bar{z}_i y),$$

where

$$z := \frac{-\beta\omega_1 + \delta\omega_2}{\alpha\omega_1 - \gamma\omega_2}. \tag{21}$$

This implies immediately that

$$N(\alpha\omega_1 - \gamma\omega_2) = a_i N(I) > 0. \tag{22}$$

Then a straightforward calculation yields

$$\frac{z - \bar{z}}{\sqrt{d}} = \frac{\alpha\delta - \beta\gamma}{N(\alpha\omega_1 - \gamma\omega_2)} \frac{\bar{\omega}_1\omega_2 - \omega_1\bar{\omega}_2}{\sqrt{d}} > 0$$

which by

$$\frac{z_i - \bar{z}_i}{\sqrt{d}} = \frac{1}{a_i} > 0$$

forces that $z = z_i$. But then (21)–(22) imply that

$$I = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2 = \mathbb{Z}(\alpha\omega_1 - \gamma\omega_2) + \mathbb{Z}(-\beta\omega_1 + \delta\omega_2)$$

is equivalent to

$$\mathbb{Z} + \mathbb{Z}z = \mathbb{Z} + \mathbb{Z}z_i = I_i.$$

Now assume that I_i and I_j are equivalent, i.e. there is some $\mu \in \mathbb{Q}(\sqrt{d})$ such that

$$\mu(\mathbb{Z} + \mathbb{Z}z_i) = \mathbb{Z} + \mathbb{Z}z_j, \qquad N(\mu) > 0.$$

Then we certainly have some $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathrm{GL}_2(\mathbb{Z})$ such that

$$\mu = \alpha + \beta z_j, \qquad \mu z_i = \gamma + \delta z_j.$$

In particular,

$$z_i = \frac{\gamma + \delta z_j}{\alpha + \beta z_j} \qquad \text{with} \qquad N(\alpha + \beta z_j) > 0.$$

By a straightforward calculation as before,

$$\frac{z_i - \bar{z}_i}{\sqrt{d}} = \frac{\alpha\delta - \beta\gamma}{N(\alpha + \beta z_j)} \frac{z_j - \bar{z}_j}{\sqrt{d}},$$

which shows that

$$\alpha\delta - \beta\gamma = 1 \qquad \text{and} \qquad N(\alpha + \beta z_j) = \frac{z_j - \bar{z}_j}{z_i - \bar{z}_i} = \frac{a_i}{a_j}.$$

Now we obtain

$$a_i(x - z_i y)(x - \bar{z}_i y) = a_j\big((\alpha + \beta z_j)x - (\gamma + \delta z_j)y\big)\big((\alpha + \beta \bar{z}_j)x - (\gamma + \delta \bar{z}_j)y\big),$$

i.e.

$$Q_i(x, y) = Q_j(\alpha x - \gamma y, -\beta x + \delta y), \qquad \begin{pmatrix} \alpha & -\gamma \\ -\beta & \delta \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}).$$

This clearly implies that $i = j$, since otherwise the forms Q_i and Q_j are inequivalent.

Incidentally, we see that the equivalence class of the associated form $Q_I(x, y)$ only depends on the narrow class of I (in particular, it is independent of the choice of ordered basis of I) and two fractional ideals I and J are in the same narrow class if and only if $Q_I(x, y)$ and $Q_J(x, y)$ are equivalent.

References

[1] V. Blomer, G. Harcos, P. Michel, *Bounds for modular L-functions in the level aspect*, Ann. Sci. École Norm. Sup. **40** (2007), 697–740.

[2] D. A. Burgess, *On character sums and L-series. II*, Proc. Lond. Math. Soc. **13** (1963), 524–536.

[3] V. A. Bykovskiĭ, *A trace formula for the scalar product of Hecke series and its applications*, translated in J. Math. Sci (New York) **89** (1998), 915–932.

[4] B. Conrey, H. Iwaniec, *The cubic moment of central values of automorphic L-functions*, Ann. of Math. **151** (2000), 1175–1216.

[5] H. Davenport, *Multiplicative number theory [revised and with a preface by H. L. Montgomery]*, 3rd edition, Graduate Texts in Mathematics 74, Springer-Verlag, New York, 2000.

[6] W. Duke, *Hyperbolic distribution problems and half-integral weight Maass forms*, Invent. Math. **92** (1988), 73–90.

[7] W. Duke, J. Friedlander, H. Iwaniec, *Bounds for automorphic L-functions*, Invent. Math. **112** (1993), 1–8.

[8] W. Duke, J. Friedlander, H. Iwaniec, *The subconvexity problem for Artin L-functions*, Invent. Math. **149** (2002), 489–577.

[9] C. F. Gauss, *Disquisitiones arithmeticae [translated by A. A. Clarke and revised by W. C. Waterhouse, C. Greither and A. W. Grootendorst]*, Springer-Verlag, New York, 1986.

[10] J. Hoffstein, P. Lockhart, *Coefficients of Maass forms and the Siegel zero (with an appendix by D. Goldfeld, J. Hoffstein and D. Lieman)*, Ann. of Math. **140** (1994), 161–181.

[11] G. Harcos, P. Michel, *The subconvexity problem for Rankin–Selberg L-functions and equidistribution of Heegner points. II*, Invent. Math. **163** (2006), 581–655.

[12] E. Hecke, *Über Modulfunktionen und die Dirichletschen Reihen mit Eulerscher Produktentwicklung. I.*, Math. Ann. **114** (1937), 1–28; *II.*, ibid. **114** (1937), 316–351.

[13] H. Iwaniec, *Small eigenvalues of Laplacian for* $\Gamma_0(N)$, Acta Arith. **56** (1990), 65–82.

[14] H. Iwaniec, *Spectral methods of automorphic forms*, 2nd edition, Graduate Studies in Mathematics 53, American Mathematical Society, Providence, RI; Revista Matemática Iberoamericana, Madrid, 2002.

[15] J. L. Lagrange, *Recherche d'arithmétique*, Nouv. Mém. Acad. Berlin (1773), 265–312.

[16] Y. V. Linnik, *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, 1968.

[17] H. Maass, *Über eine neue Art von nichtanalytischen automorphen Funktionen und die Bestimmung Dirichletscher Reihen durch Funktionalgleichungen*, Math. Ann. **121** (1949), 141–183.

[18] P. Michel, A. Venkatesh, *Equidistribution, L-functions and ergodic theory: on some problems of Yu. Linnik*, International Congress of Mathematicians. Vol. II, 421–457, Eur. Math. Soc., Zürich, 2006.

[19] A. Popa, *Central values of Rankin L-series over real quadratic fields*, Compos. Math. **142** (2006), 811–866.

[20] C. L. Siegel, *Advanced analytic number theory*, 2nd edition, Tata Institute of Fundamental Research Studies in Mathematics, 9, Tata Institute of Fundamental Research, Bombay, 1980.

[21] J.-L. Waldspurger, *Sur les valeurs de certaines fonctions L automorphes en leur centre de symétrie*, Compos. Math. **54** (1985), 173–242.
[22] S. Zhang, *Gross–Zagier formula for* GL_2, Asian J. Math. **5** (2001), 183–290.

Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences, POB 127, Budapest H-1364, Hungary
E-mail address: gharcos@renyi.hu