Equidistribution on the modular surface and automorphic L-functions

Gergely Harcos

11 February, 2008

Integral binary quadratic forms

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

- discriminant $d := b^2 - 4ac \in \mathbb{Z}$
- possible discriminants are $d \equiv 0, 1 \pmod 4$
- form reducible if and only if d is a square
- form positive definite if $d < 0$ and $a, c > 0$
- form negative definite if $d < 0$ and $a, c < 0$
- form indefinite if $d > 0$

Fundamental discriminants, primitive forms

$$\langle a, b, c\rangle := ax^2 + bxy + cy^2 \in \mathbb{Z}[x, y]$$
$$d := b^2 - 4ac \in \mathbb{Z}$$

- discriminant fundamental if $d \neq d'e^2$ for all discriminants $d' < d$ and $e \in \mathbb{Z}$

- fundamental discriminant implies form $\langle a, b, c\rangle$ is primitive, i.e. $\gcd(a, b, c) = 1$

- possible fundamental discriminants are d square-free $\equiv 1 \pmod 4$ and 4 times square-free $\equiv 2, 3 \pmod 4$; they parametrize the quadratic extensions $\mathbb{Q}(\sqrt{d})$

- first values are −20, −19, −15, −11, −8, −7, −4, −3; 5, 8, 12, 13, 17, 21, 24, 28

Equivalence of integral binary quadratic forms

For $M = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathsf{SL}_2(\mathbb{Z})$ consider the actions

$$(x, y) \overset{M}{\longmapsto} (x', y') \quad \overset{\text{df}}{\Longleftrightarrow} \quad (x', y') = (\alpha x + \beta y, \gamma x + \delta y)$$

$$\langle a, b, c\rangle \overset{M}{\longmapsto} \langle a', b', c'\rangle \quad \overset{\text{df}}{\Longleftrightarrow} \quad a'x'^2 + b'x'y' + c'y'^2 = ax^2 + bxy + cy^2$$

- $\langle a, b, c\rangle$ and $\langle a', b', c'\rangle$ as above are called equivalent

- equivalent forms have the same discriminant

Finiteness of class number

Fix fundamental discriminant d, and consider

$$\langle a,b,c\rangle \overset{S}{\longmapsto} \langle c,-b,a\rangle, \qquad S := \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

$$\langle a,b,c\rangle \overset{T}{\longmapsto} \langle a,b-2a,c+a-b\rangle, \quad T := \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Applying $T^{\pm 1}$, S finitely many times we achieve

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

Then

$$|d| = |b^2 - 4ac| \geqslant 4|ac| - b^2 \geqslant 3b^2$$

shows there are

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

inequivalent forms $\langle a,b,c\rangle$ of discriminant d.

For example, in the case of $d = -23$ we obtain $h(-23) = 3$ different classes represented by the forms $\langle 1,1,6\rangle$ and $\langle 2,\pm 1,3\rangle$.

Geometric picture

Conformal automorphisms of the Riemann sphere $\mathbb{C} \cup \{\infty\}$ fixing $\mathbb{R} \cup \{\infty\}$ are given by fractional linear transformations

$$z \xmapsto{g} \frac{\alpha z + \beta}{\gamma z + \delta}, \qquad g = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathsf{SL}_2(\mathbb{R}).$$

Decompose each form of discriminant d as

$$ax^2 + bxy + cy^2 = a(x - uy)(x - wy),$$

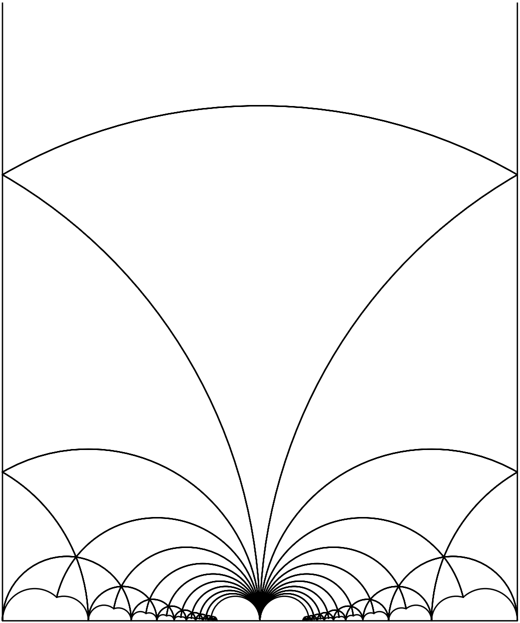
$$u := \frac{-b - \sqrt{d}}{2a}, \qquad w := \frac{-b + \sqrt{d}}{2a},$$

and embed $\mathbb{Q}(\sqrt{d})$ into $\mathbb{C} \cup \{\infty\}$. Then the action of $\mathsf{SL}_2(\mathbb{Z})$ on forms induces on the roots precisely the action given by fractional linear transformations above. In particular,

$$(u, w) \xmapsto{S} (-1/u, -1/w), \quad S := \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

$$(u, w) \xmapsto{T} (u + 1, w + 1), \quad T := \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

T⁻¹
1
T
ρ
i
T⁻¹S
S
TS
STS
ST
ST⁻¹
ST⁻¹S
−1
0
+1



Geometric picture (cont.)

$\mathbb{C} - \mathbb{R}$ is the disjoint union of $\mathcal{H}$ and $\overline{\mathcal{H}}$, where

$$\mathcal{H} := \{z = x + iy \in \mathbb{C} : y > 0\}$$

is the upper half-plane equipped with $\mathrm{SL}_2(\mathbb{R})$-invariant line element and area element

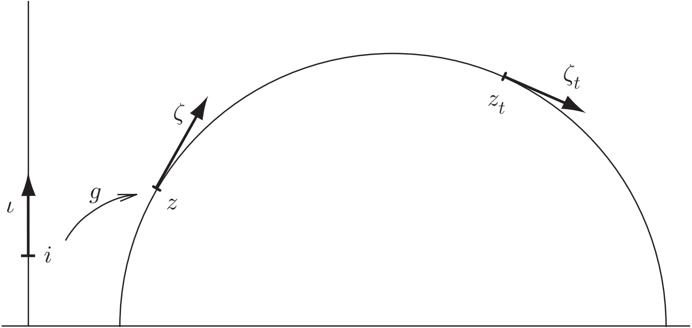
$$d^2s(z) := \frac{dx^2 + dy^2}{y^2} \quad \text{and} \quad d\mu(z) := \frac{3}{\pi}\,\frac{dxdy}{y^2}.$$

Geodesics in $\mathcal{H}$ are the half-lines and semi-circles orthogonal to $\mathbb{R}$. The $\mathrm{SL}_2(\mathbb{Z})$-orbits in $\mathcal{H}$ form a noncompact surface $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$ of curvature -1 and area 1.

Let $\langle a, b, c\rangle$ run through all forms of discriminant d and consider the roots as before,

$$u := \frac{-b - \sqrt{d}}{2a}, \quad w := \frac{-b + \sqrt{d}}{2a}.$$

For $d < 0$ the various roots $w \in \mathcal{H}$ give rise to $h(d)$ points in $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$. For $d > 0$ the geodesics joining the various pairs $\{u, w\} \subset \mathbb{R}$ give rise to $h(d)$ geodesics in $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$.



ι
i
g
z
ζ
z_t
ζ_t

Geometric picture (cont.)

Any geodesic $G_{u,w}$ joining the roots of an indefinite form $\langle a, b, c\rangle$ becomes closed when projected to $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$. Namely, for any $g \in \mathrm{SL}_2(\mathbb{R})$ mapping the pair $(0, \infty)$ to (u, w) the motions in $\mathrm{SL}_2(\mathbb{Z})$ fixing $G_{u,w}$ are given by

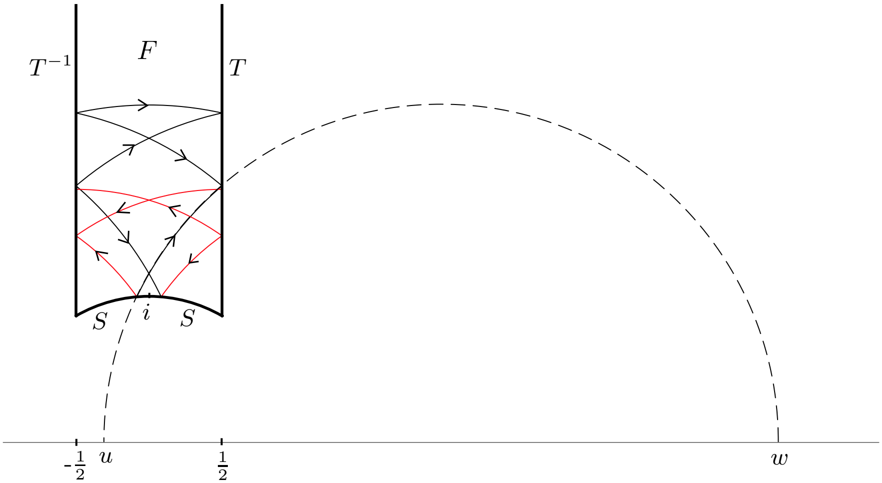
$$g\begin{pmatrix}\lambda & 0\\ 0 & \lambda^{-1}\end{pmatrix}g^{-1} = \begin{pmatrix}\frac{m-bn}{2} & -nc\\ na & \frac{m+bn}{2}\end{pmatrix},$$

where

$$\lambda = \frac{m+n\sqrt{d}}{2}, \quad m, n \in \mathbb{Z}, \quad m^2 - dn^2 = 4,$$

runs through the totally positive units in the ring of integers of $\mathbb{Q}(\sqrt{d})$. If $\lambda_d > 1$ generates the group of totally positive units then the length of the projected geodesic is $2\ln(\lambda_d)$.

For a fixed λ and a fixed closed geodesic in $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$ the above motions for the various $g \in \mathrm{SL}_2(\mathbb{R})$ form a hyperbolic conjugacy class in $\mathrm{SL}_2(\mathbb{Z})$. All hyperbolic conjugacy classes in $\mathrm{SL}_2(\mathbb{Z})$ arise in this way, and primitive classes correspond to $\lambda = \pm\lambda_d^{\pm 1}$.



T^{-1}
F
T
S
i
S
-$\frac{1}{2}$
u
$\frac{1}{2}$
w

Eisenstein series on $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$

$$\theta(s) := 2\pi^{-s}\Gamma(s)\zeta(2s), \qquad \eta_s(n) := \sum_{ab=n} (a/b)^s$$

$$E^*(z,s) := \theta(s)E(z,s) := \frac{\theta(s)}{2} \sum_{\substack{m,n\in\mathbb{Z} \\ \gcd(m,n)=1}} \frac{y^s}{|mz+n|^{2s}}$$

$$= \theta(s)y^s + \theta(1-s)y^{1-s} + 4\sqrt{y}\sum_{n\neq 0} \eta_{s-\frac{1}{2}}(|n|)K_{s-\frac{1}{2}}(2\pi|n|y)e^{2\pi inx}$$

• For any $s \in \mathbb{C} - \{0,1\}$, $E^*(z,s)$ is real-analytic in $z \in \mathcal{H}$ and invariant under $z \mapsto \gamma z$ for any $\gamma \in \mathrm{SL}_2(\mathbb{Z})$.

• For any $z \in \mathcal{H}$, $E^*(z,s)$ is holomorphic in $s \in \mathbb{C} - \{0,1\}$, invariant under $s \mapsto 1-s$, and has a simple pole at $s = 1$ (resp. $s = 0$) with constant residue 1 (resp. -1).

Dirichlet's class number formula via Eisenstein series

• Λ_d: the set of special points or closed geodesics on $\mathsf{SL}_2(\mathbb{Z})\backslash\mathcal{H}$ representing the $h(d)$ classes of forms $\langle a,b,c\rangle$ of discriminant d

• w_d: the number of roots of unity in $\mathbb{Q}(\sqrt{d})$

$$\sum_{z\in\Lambda_d} E^*(z,s) = w_d\,|d|^{\frac{s}{2}}\,(2\pi)^{-s}\Gamma(s)\,\zeta(s)L(s,(\tfrac{d}{\cdot})),\quad d<0,$$

$$\sum_{G\in\Lambda_d}\int_G E^*(z,s)\,ds(z) = w_d\,|d|^{\frac{s}{2}}\;\pi^{-s}\Gamma\left(\frac{s}{2}\right)^2\,\zeta(s)L(s,(\tfrac{d}{\cdot})),\quad d>0.$$

Taking residues at $s=1$ of both sides we obtain

$$h(d) = w_d\,|d|^{\frac{1}{2}}\,(2\pi)^{-1}\,L(1,(\tfrac{d}{\cdot})),\qquad d<0,$$

$$h(d)\,2\ln(\lambda_d) = w_d\,|d|^{\frac{1}{2}}\qquad L(1,(\tfrac{d}{\cdot})),\qquad d<0.$$

Siegel's theorem

- Λ_d: the set of special points or closed geodesics on $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$ representing the $h(d)$ classes of forms $\langle a,b,c\rangle$ of discriminant d
- w_d: the number of roots of unity in $\mathbb{Q}(\sqrt{d})$

$$\begin{aligned} h(d) &= w_d\,|d|^{\frac{1}{2}}\,(2\pi)^{-1}\,L(1,(\tfrac{d}{\cdot})), && d<0,\\ h(d)\,2\ln(\lambda_d) &= w_d\,|d|^{\frac{1}{2}}\qquad\qquad L(1,(\tfrac{d}{\cdot})), && d<0. \end{aligned}$$

Siegel's theorem from 1934 states that

$$|d|^{-\varepsilon} \ll_\varepsilon L(1,(\tfrac{d}{\cdot})) \ll_\varepsilon |d|^{\varepsilon},$$

so that

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

The spectral decomposition of $L^2(\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H})$

The space $L^2(\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H})$ is defined by the inner product

$$\langle g_1, g_2\rangle := \int_{\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}} g_1(z)\,\overline{g_2(z)}\,d\mu(z).$$

Smooth and compactly supported functions $g : \mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H} \to \mathbb{C}$ are dense. They have a decomposition (Selberg, 1956)

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

which converges uniformly on compact sets. The functions u_j here form an orthonormal basis of the so-called cuspidal subspace and possess very nice harmonic properties, along with the functions $E(\cdot, \frac{1}{2}+it)$. Precisely, they are simultaneous eigenfunctions of various "averaging operators" on $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$.

Laplacian and Hecke operators on $L^2(\mathsf{SL}_2(\mathbb{Z})\backslash\mathcal{H})$

- g: some u_j or $E(\cdot, \frac{1}{2}+it)$ with $t \in \mathbb{R}$
- p: any prime number

$$\Delta g \;:=\; -y^2\left(\frac{\partial^2 g}{\partial x^2}+\frac{\partial^2 g}{\partial y^2}\right) \qquad =:\; \left(\tfrac{1}{4}+t_g^2\right)g, \qquad t_g \in \mathbb{R}$$

$$T_p g \;:=\; \frac{1}{\sqrt{p}}\sum_{\substack{ad=p\\0\leqslant b<d}}^{p} g\left(\frac{az+b}{d}\right) \qquad =:\; \Big(\alpha_g(p)+\beta_g(p)\Big)g, \qquad \alpha_g(p)\beta_g(p)=1$$

$$T_{-1}g \;:=\; g(-\overline{z}) \qquad =:\; (-1)^\rho g, \qquad \rho \in \{0,1\}$$

$$\begin{aligned}\Lambda(s,g) \;&:= \pi^{-s}\Gamma\left(\tfrac{s+\rho-it_g}{2}\right)\Gamma\left(\tfrac{s+\rho+it_g}{2}\right)L(s,g)\\ &:= \pi^{-s}\Gamma\left(\tfrac{s+\rho-it_g}{2}\right)\Gamma\left(\tfrac{s+\rho+it_g}{2}\right)\prod_p \frac{1}{(1-\alpha_g(p)p^{-s})(1-\beta_g(p)p^{-s})}\\ &= (-1)^\rho\,\Lambda(1-s,g)\end{aligned}$$

Weyl sums and central twisted L-values

- g: some u_j or $E(\cdot, \frac{1}{2} + it)$ with $t \in \mathbb{R}$
- Λ_d: the set of special points or closed geodesics on $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$ representing the $h(d)$ classes of forms $\langle a, b, c\rangle$ of discriminant d

$$g(x+iy) = g_{\text{const}}(y) + \sqrt{y} \sum_{n \neq 0} \rho_g(n) K_{it_g}(2\pi|n|y) e^{2\pi i n x}$$

The following identity (developed by Waldspurger, Kohnen–Zagier, Katok–Sarnak, Guo, Zhang, Popa from 1985 to 2006) is deep:

$$\left| \sum_{z \in \Lambda_d} g(z) \right|^2 = c_d \, |d|^{\frac{1}{2}} \, |\rho_g(1)|^2 \, \Lambda\left(\tfrac{1}{2}, g\right) \Lambda\left(\tfrac{1}{2}, g \otimes \left(\tfrac{d}{\cdot}\right)\right), \; d < 0,$$

$$\left| \sum_{G \in \Lambda_d} \int_G g(z) \, ds(z) \right|^2 = c_d \, |d|^{\frac{1}{2}} \, |\rho_g(1)|^2 \, \Lambda\left(\tfrac{1}{2}, g\right) \Lambda\left(\tfrac{1}{2}, g \otimes \left(\tfrac{d}{\cdot}\right)\right), \; d > 0.$$

Weyl sums and subconvexity bounds

• g: some u_j or $E(\cdot, \frac{1}{2} + it)$ with $t \in \mathbb{R}$

By work of Burgess (1963) and Duke–Friedlander–Iwaniec (1994),

$$\exists\ \delta > 0, A > 0 :\ L\left(\tfrac{1}{2}, g \otimes \left(\tfrac{d}{\cdot}\right)\right) \ll (1 + |t_g|)^A \, |d|^{\frac{1}{2} - \delta},$$

hence by crude bounds on $\rho_g(1)$ and $\Lambda(s, g)$ we conclude, for some $B > 0$,

$$\left| \sum_{z \in \Lambda_d} g(z) \right|^2 \ll (1 + |t_g|)^B \, |d|^{1-\delta}, \qquad d < 0,$$

$$\left| \sum_{G \in \Lambda_d} \int_G g(z)\, ds(z) \right|^2 \ll (1 + |t_g|)^B \, |d|^{1-\delta}, \qquad d > 0.$$

Equidistribution on the modular surface

• g: any smooth and compactly supported function on $\mathsf{SL}_2(\mathbb{Z})\backslash\mathcal{H}$

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

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

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

$$\frac{1}{h(d)} \sum_{z\in\Lambda_d} g(z) \to \int_{\mathsf{SL}_2(\mathbb{Z})\backslash\mathcal{H}} g(z)\, d\mu(z), \qquad d \to -\infty$$

$$\frac{1}{h(d)\, 2\ln(\lambda_d)} \sum_{G\in\Lambda_d} \int_G g(z)\, ds(z) \to \int_{\mathsf{SL}_2(\mathbb{Z})\backslash\mathcal{H}} g(z)\, d\mu(z), \qquad d \to +\infty$$

Refinement: equidistribution in shorter orbits

There is a natural bijection from Λ_d to the narrow ideal class group H_d of $\mathbb{Q}(\sqrt{d})$ which induces an action of H_d on Λ_d. Equidistribution in orbits of size $\gg_\varepsilon |d|^{1/2-\delta/2+\varepsilon}$ follows from a bound

$$\left|\sum_{\sigma \in H_d} \overline{\psi(\sigma)} g(z_0^\sigma)\right|^2 \ll (1+|t_g|)^B |d|^{1-\delta}, \qquad d<0,$$

$$\left|\sum_{\sigma \in H_d} \overline{\psi(\sigma)} \int_{G_0^\sigma} g(z)\, ds(z)\right|^2 \ll (1+|t_g|)^B |d|^{1-\delta}, \qquad d>0,$$

where g is any u_j or $E(\cdot, \frac{1}{2}+it)$ with $t \in \mathbb{R}$, z_0 (resp. G_0) is any element of Λ_d when $d<0$ (resp. $d>0$), and $\psi : H_d \to \mathbb{C}^\times$ is any ideal class character.

Refinement: equidistribution in shorter orbits (cont.)

By deep formulae of Zhang (2001) and Popa (2006) the left hand side equals

$$\left|\sum_{\sigma\in H_d} \overline{\psi(\sigma)} \ldots\right|^2 = c_d\,|d|^{\frac{1}{2}}\,|\rho_g(1)|^2\,\Lambda\left(\tfrac{1}{2}, g\otimes f_\psi\right),$$

where f_ψ is the Jacquet–Langlands lift of ψ, a modular form on $\mathcal{H}$ of level $|d|$ and nebentypus $\left(\frac{d}{\cdot}\right)$ with the same completed L-function as ψ. If $\psi : H_d \to \mathbb{C}^\times$ is real-valued then there is a factorization $d = d_1 d_2$ into fundamental discriminants such that

$$\Lambda(s, g\otimes f_\psi) = \Lambda(s, g\otimes (\tfrac{d_1}{\cdot}))\,\Lambda(s, g\otimes (\tfrac{d_2}{\cdot})).$$

Otherwise f_ψ is a cusp form and the necessary subconvex bounds were proved by Duke–Friedlander–Iwaniec (2002) and Harcos–Michel (2006). Equidistribution follows in orbits of size $\gg |d|^{0.4997}$.