
Quantum and Arithmetical Chaos

Eugene Bogomolny

Laboratoire de Physique Théorique et Modèles Statistiques
Université de Paris XI, Bât. 100, 91405 Orsay Cedex, France
bogomol@lptms.u-psud.fr

Summary. The lectures are centered around three selected topics of quantum chaos: the Selberg trace formula, the two-point spectral correlation functions of Riemann zeta function zeros, and of the Laplace–Beltrami operator for the modular group. The lectures cover a wide range of quantum chaos applications and can serve as a non-formal introduction to mathematical methods of quantum chaos.

Introduction

Quantum chaos is a nickname for the investigation of quantum systems which do not permit exact solutions. The absence of explicit formulas means that underlying problems are so complicated that they cannot be expressed in terms of known (\simeq simple) functions. The class of non-soluble systems is very large and practically any model (except a small set of completely integrable systems) belongs to it. An extreme case of quantum non-soluble problems appears naturally when one considers the quantization of classically chaotic systems which explains the word ‘chaos’ in the title.

As, by definition, for complex systems exact solutions are not possible, new analytical approaches were developed within the quantum chaos. First, one may find relations between different non-integrable models, hoping that for certain questions a problem will be more tractable than another. Second, one considers, instead of exact quantities, the calculation of their smoothed values. In many cases such coarse graining appears naturally in experimental settings and, usually, it is more easy to treat. Third, one tries to understand statistical properties of quantum quantities by organizing them in suitable ensembles. An advantage of such approach is that many different models may statistically be indistinguishable which leads to the notion of statistical universality.

The ideas and methods of quantum chaos are not restricted only to quantum models. They can equally well be applied to any problem whose analytical solution either is not possible or is very complicated. One of the most spectacular examples of such interrelations is the application of quantum chaos

to number theory, in particular, to the zeros of the Riemann zeta function. Though a hypothetical quantum-like system whose eigenvalues coincide with the imaginary part of Riemann zeta function zeros is not (yet!) found, the Riemann zeta function is, in many aspects, similar to dynamical zeta functions and the investigation of such relations already mutually enriched both quantum chaos and number theory (see e.g. the calculation by Keating and Snaith moments of the Riemann zeta function using the random matrix theory [43]).

The topics of these lectures were chosen specially to emphasize the interplay between physics and mathematics which is typical for quantum chaos.

In Chap. 1 different types of trace formulas are discussed. The main attention is given to the derivation of the Selberg trace formula which relates the spectral density of automorphic Laplacian on hyperbolic surfaces generated by discrete groups with classical periodic orbits for the free motion on these surfaces. This question is rarely discussed in the physical literature but is of general interest because it is the only case where the trace formula is exact and not only a leading semiclassical contribution as for general dynamical systems. Short derivations of trace formulas for dynamical systems and for the Riemann zeta function zeros are also presented in this Chapter.

According to the well-known conjecture [17] statistical properties of eigenvalues of energies of quantum chaotic systems are described by standard random matrix ensembles depending only on system symmetries. In Chap. 2 we discuss analytical methods of confirmation of this conjecture. The largest part of this Chapter is devoted to a heuristic derivation of the ‘exact’ two-point correlation function for the Riemann zeros. The derivation is based on the Hardy–Littlewood conjecture about the distribution of prime pairs which is also reviewed. The resulting formula agrees very well with numerical calculations of Odlyzko.

In Chap. 3 a special class of dynamical systems is considered, namely, hyperbolic surfaces generated by arithmetic groups. Though from viewpoint of classical mechanics these models are the best known examples of classical chaos, their spectral statistics are close to the Poisson statistics typical for integrable models. The reason for this unexpected behavior is found to be related with exponential degeneracies of periodic orbit lengths characteristic for arithmetical systems. The case of the modular group is considered in details and the exact expression for the two-point correlation function for this problem is derived.

To be accessible for physics students the lectures are written in a non-formal manner. In many cases analogies are used instead of theorems and complicated mathematical notions are illustrated by simple examples.

Trace Formulas

Different types of trace formula are the cornerstone of quantum chaos. Trace formulas relate quantum properties of a system with its classical counterparts. In the simplest and widely used case the trace formula expresses the quantum density of states through a sum over periodic orbits and each term in this sum can be calculated from pure classical mechanics.

In general, dynamical trace formulas represent only the leading term of the semiclassical expansion in powers of \hbar . The computation of other terms is possible though quite tedious [1]. The noticeable exception is the free motion on constant negative curvature surfaces generated by discrete groups where the trace formula (called the Selberg trace formula) is exact. The derivation of this formula is the main goal of this Section.

For clarity, in Sect. 1 the simplest case of the rectangular billiard is briefly considered and the trace formula for this system is derived. The derivation is presented in a manner which permits to generalize it to the Selberg case of constant negative curvature surfaces generated by discrete groups which is considered in details in Sect. 2. In Sects. 3 and 4 the derivations of the trace formula for, respectively, classically integrable and chaotic systems are presented. In Sect. 5 it is demonstrated that the density of Riemann zeta function zeros can be written as a sort of trace formula where the role of periodic orbits is played by prime numbers. Section 6 is the summary of this Chapter.

1 Plane Rectangular Billiard

To clarify the derivation of trace formulas let us consider in details a very simple example, namely, the computation of the energy spectrum for the plane rectangular billiard with periodic boundary conditions.

This problem consists of solving the equation

$$(\Delta + E_{\mathbf{n}})\Psi_{\mathbf{n}}(x, y) = 0 \tag{1}$$

where $\Delta = \partial^2/\partial x^2 + \partial^2/\partial y^2$ is the usual two-dimensional Laplacian with periodic boundary conditions

$$\Psi_{\mathbf{n}}(x+a, y) = \Psi_{\mathbf{n}}(x, y+b) = \Psi_{\mathbf{n}}(x, y) \quad (2)$$

where a and b are sizes of the rectangle.

The plane wave

$$\Psi_{\mathbf{n}}(x, y) = e^{ik_1x + ik_2y}$$

is an admissible solution of (1). Boundary conditions (2) determine the allowed values of the momentum \mathbf{k}

$$k_1 = \frac{2\pi}{a}n_1, \quad k_2 = \frac{2\pi}{b}n_2,$$

with $n_1, n_2 = 0, \pm 1, \pm 2, \dots$, and, consequently, energy eigenvalues are

$$E_{n_1 n_2} = \left(\frac{2\pi}{a}n_1\right)^2 + \left(\frac{2\pi}{b}n_2\right)^2. \quad (3)$$

The first step of construction of trace formulas is to consider instead of individual eigenvalues their density defined as the sum over all eigenvalues which explains the word ‘trace’

$$d(E) \equiv \sum_{n_1, n_2 = -\infty}^{+\infty} \delta(E - E_{n_1 n_2}). \quad (4)$$

To transform this and similar expressions into a convenient form one often uses the Poisson summation formula

$$\sum_{n=-\infty}^{+\infty} f(n) = \sum_{m=-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{2\pi i m n} f(n) dn. \quad (5)$$

An informal proof of this identity can, for example, be done as follows.

First

$$\sum_{n=-\infty}^{+\infty} f(n) = \int_{-\infty}^{+\infty} f(x)g(x)dx$$

where $g(x)$ is the periodic δ -function

$$g(x) = \sum_{m=-\infty}^{+\infty} \delta(x - m).$$

As any periodic function with period 1, $g(x)$ can be expanded into the Fourier series

$$g(x) = \sum_{m=-\infty}^{+\infty} e^{2\pi i m x} c_m.$$

Coefficients c_m are obtained by the integration of $g(x)$ over one period

$$c_m = \int_{-1/2}^{+1/2} g(y) e^{-2\pi i m y} dy = 1$$

which gives (5).

By applying the Poisson summation formula (5) to the density of states (4) one gets

$$d(E) = \sum_{m_1, m_2 = -\infty}^{+\infty} \int \int e^{2\pi i(m_1 n_1 + m_2 n_2)} \times \\ \times \delta \left(E - \left(\frac{2\pi}{a} n_1 \right)^2 - \left(\frac{2\pi}{b} n_2 \right)^2 \right) dn_1 dn_2 .$$

Perform the following substitutions: $E = k^2$, $n_1 = ar \cos \varphi / 2\pi$, and $n_2 = br \sin \varphi / 2\pi$. Then $dn_1 dn_2 = ab r dr d\varphi / (2\pi)^2$ and

$$d(E) = \frac{\mu(D)}{(2\pi)^2} \sum_{m_1, m_2 = -\infty}^{+\infty} \int \int e^{i(m_1 a \cos \varphi + m_2 b \sin \varphi)r} \delta(k^2 - r^2) r dr d\varphi \\ = \frac{\mu(D)}{2(2\pi)^2} \sum_{m_1, m_2 = -\infty}^{+\infty} \int_0^{2\pi} e^{ik\sqrt{(m_1 a)^2 + (m_2 b)^2} \cos \varphi} d\varphi \\ = \frac{\mu(D)}{4\pi} \sum_{m_1, m_2 = -\infty}^{+\infty} J_0(kL_p) ,$$

where $\mu(D) = ab$ is the area of the rectangle,

$$J_0(x) = \frac{1}{2\pi} \int_0^{2\pi} e^{ix \cos \varphi} d\varphi$$

is the Bessel function of order zero (see e.g. [32], Vol. 2, Sect. 7), and

$$L_p = \sqrt{(m_1 a)^2 + (m_2 b)^2}$$

is (as it is easy to check) the length of a periodic orbit in the rectangle with periodic boundary conditions.

Separating the term with $m_1 = m_2 = 0$ one concludes that the eigenvalue density of the rectangle with periodic boundary conditions can be written as the sum of two terms

$$d(E) = \bar{d}(E) + d^{(osc)}(E) ,$$

where

$$\bar{d}(E) = \frac{\mu(D)}{4\pi} \tag{6}$$

is the smooth part of the density and

$$d^{(osc)}(E) = \frac{\mu(D)}{4\pi} \sum_{\text{p.o.}} J_0(kL_p), \quad (7)$$

is the oscillating part equal to a sum over all periodic orbits in the rectangle.

As

$$J_0(z) \xrightarrow{z \rightarrow \infty} \sqrt{\frac{2}{\pi z}} \cos\left(z - \frac{\pi}{4}\right)$$

the oscillating part of the level density in the semiclassical limit $k \rightarrow \infty$ takes the form

$$d^{(osc)}(E) = \frac{\mu(D)}{\sqrt{8\pi k}} \sum_{\text{p.o.}} \frac{1}{\sqrt{L_p}} \cos\left(kL_p - \frac{\pi}{4}\right). \quad (8)$$

Let us repeat the main steps which lead to this trace formula. One starts with an explicit formula (like (3)) which expresses eigenvalues as a function of integers. Using the Poisson summation formula (5) the density of states (4) is transformed into a sum over periodic orbits. In Sect. 3 it will be demonstrated that exactly this method can be applied for any integrable systems in the semiclassical limit where eigenvalues can be approximated by the WKB formulas.

More Refined Approach

The above method of deriving the trace formula for the rectangular billiard can be applied only if one knows an explicit expression for eigenvalues. For chaotic systems this is not possible and another method has to be used.

Assume that one has to solve the equation

$$(E_{\mathbf{n}} - \hat{H})\Psi_{\mathbf{n}}(\mathbf{x}) = 0$$

for a certain problem with a Hamiltonian \hat{H} . Under quite general conditions eigenfunctions $\Psi_{\mathbf{n}}(\mathbf{x})$ can be chosen orthogonal

$$\int \Psi_{\mathbf{n}}(\mathbf{x})\Psi_{\mathbf{m}}^*(\mathbf{x})d\mathbf{x} = \delta_{nm}$$

and they form a complete system of functions

$$\sum_{\mathbf{n}} \Psi_{\mathbf{n}}(\mathbf{x})\Psi_{\mathbf{n}}^*(\mathbf{y}) = \delta(\mathbf{x} - \mathbf{y}).$$

The Green function of the problem, by definition, obeys the equation

$$(E - \hat{H})G_E(\mathbf{x}, \mathbf{y}) = \delta(\mathbf{x} - \mathbf{y})$$

and the same boundary conditions as the original eigenfunctions. Its explicit form can formally be written through exact eigenfunctions and eigenvalues as follows

$$G_E(\mathbf{x}, \mathbf{y}) = \sum_{\mathbf{n}} \frac{\Psi_{\mathbf{n}}(\mathbf{x})\Psi_{\mathbf{n}}^*(\mathbf{y})}{E - E_{\mathbf{n}} + i\epsilon}. \quad (9)$$

The $+i\epsilon$ prescription determines the so-called retarded Green function.

Example

To get used to Green functions let us consider in details the calculation of the Green function for the free motion in f -dimensional Euclidean space. This Green function obeys the free equation

$$(E + \hbar^2 \Delta) G_E^{(0)}(\mathbf{x}, \mathbf{y}) = \delta(\mathbf{x} - \mathbf{y}) . \quad (10)$$

Let us look for the solution of the above equation in the form $G_E^{(0)}(\mathbf{x}, \mathbf{y}) = G(r)$ where $r = |\mathbf{x} - \mathbf{y}|$ is the distance between two points.

Simple calculations shows that for $r \neq 0$ $G(r)$ obeys the equation

$$\frac{d^2 G}{dr^2} + \frac{f-1}{r} \frac{dG}{dr} + \frac{k^2}{\hbar^2} G = 0$$

where $E = k^2$.

After the substitution

$$G(r) = r^{1-f/2} g\left(\frac{k}{\hbar} r\right)$$

one gets for $g(z)$ the Bessel equation (see e.g. [32], Vol. 2, Sect. 7)

$$\frac{d^2 g}{dz^2} + \frac{1}{z} \frac{dg}{dz} + \left(1 - \frac{\nu^2}{z^2}\right) g = 0 \quad (11)$$

with $\nu = |f/2 - 1|$.

There are many solutions of this equation. The above $+i\epsilon$ prescription means that when $k \rightarrow k + i\epsilon$ with a positive ϵ the Green function has to decrease at large distances. It is easy to see that $G(r)$ is proportional to $e^{\pm ikr/\hbar}$ at large r . The $+i\epsilon$ prescription selects a solution which behaves at infinity like $e^{+ikr/\hbar}$ with positive k . The required solution of (11) is the first Hankel function (see [32], Vol. 2, Sect. 7)

$$g(z) = C_f H_\nu^{(1)}(z) \quad (12)$$

where C_f is a constant and $H_\nu^{(1)}(z)$ has the following asymptotics for large and small z

$$H_\nu^{(1)}(z) \xrightarrow{z \rightarrow \infty} \sqrt{\frac{2}{\pi z}} e^{i(z - \pi\nu/2 - \pi/4)}$$

and

$$H_\nu^{(1)}(z) \xrightarrow{z \rightarrow 0} \begin{cases} -i2^\nu \Gamma(\nu) z^{-\nu} / \pi, & \nu \neq 2 \\ 2i \ln z / \pi, & \nu = 2 \end{cases} .$$

The overall factor in (12) has to be computed from the requirement that the Green function will give the correct δ -function contribution in the right hand side of (10). This term can appear only in the result of differentiation of the Green function at small r where it has the following behaviour

$$G(r) \xrightarrow{r \rightarrow 0} G_0(r) = A_f r^{2-f}$$

with

$$A_f = C_f \frac{2^\nu \hbar^\nu \Gamma(\nu)}{i\pi k^\nu} .$$

One should have

$$\hbar^2 \Delta G_0(r) = \delta(\mathbf{r}) . \quad (13)$$

Multiplying this equality by a suitable test function $f(r)$ quickly decreasing at infinity one has

$$\hbar^2 \int f(r) \Delta G_0(r) d\mathbf{r} = f(0) .$$

Integrating by parts one obtains

$$\hbar^2 \int \frac{\partial}{\partial x_\mu} f(r) \frac{\partial}{\partial x_\mu} G_0(r) d\mathbf{r} = -f(0) .$$

As both functions $f(r)$ and $G_0(r)$ depend only on the modulus of \mathbf{r} one finally finds

$$\hbar^2 \int_0^\infty \frac{df(r)}{dr} \frac{dG_0(r)}{dr} r^{f-1} dr S_{f-1} = -f(0)$$

where S_{f-1} is the volume of the $(f-1)$ -dimensional sphere $x_1^2 + \dots + x_f^2 = 1$. Using (13) one concludes that in order to give the δ -function term A_f has to obey

$$\hbar^2 A_f (f-2) S_{f-1} = -1 .$$

One of the simplest method of calculation of S_{f-1} is the following identity

$$\int_{-\infty}^\infty e^{-x_1^2} dx_1 \int_{-\infty}^\infty e^{-x_2^2} dx_2 \dots \int_{-\infty}^\infty e^{-x_f^2} dx_f = \pi^{f/2} .$$

By changing Cartesian coordinates in the left hand side to hyper-spherical ones we obtain

$$\int_0^\infty e^{-r^2} r^{f-1} dr S_{f-1} = \pi^{f/2}$$

which gives

$$S_{f-1} = \frac{2\pi^{f/2}}{\Gamma(f/2)}$$

where $\Gamma(x)$ is the usual gamma-function (see e.g. [32], Vol. 1, Sect. 1).

Combining together all terms and using the relation $x\Gamma(x) = \Gamma(x+1)$ one gets the explicit expression for the free Green function in f dimensions

$$G_E^{(0)}(\mathbf{x}, \mathbf{y}) = \frac{k^\nu}{4i\hbar^2 (2\pi\hbar r)^\nu} H_\nu^{(1)} \left(\frac{k}{\hbar} |\mathbf{x} - \mathbf{y}| \right) \quad (14)$$

where $\nu = |f/2 - 1|$. In particular, in the two-dimensional Euclidean space

$$G_E^{(0)}(\mathbf{x}, \mathbf{y}) = \frac{1}{4i\hbar^2} H_0^{(1)}\left(\frac{k}{\hbar}|\mathbf{x} - \mathbf{y}|\right). \quad (15)$$

Another method of calculation of the free Green function is based on (9) which for the free motion is equivalent to the Fourier expansion

$$G_E^{(0)}(\mathbf{x}, \mathbf{y}) = \int \frac{d\mathbf{p}}{(2\pi\hbar)^f} \frac{e^{i\mathbf{p}(\mathbf{x}-\mathbf{y})/\hbar}}{E - p^2 + i\epsilon}. \quad (16)$$

Performing angular integration one obtains the same formulas as above.

The knowledge of the Green function permits to calculate practically all quantum mechanical quantities. In particular, using

$$\text{Im} \frac{1}{x + i\epsilon} \xrightarrow{\epsilon \rightarrow 0} -\pi\delta(x)$$

one gets that the eigenvalue density is expressed through the exact Green function as follows

$$d(E) = -\frac{1}{\pi} \text{Im} \int_D G_E(\mathbf{x}, \mathbf{x}) d\mathbf{x}. \quad (17)$$

This general expression is the starting point of all trace formulas.

For the above model of the rectangle with periodic boundary conditions the exact Green function has to obey

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + E\right)G_E(x, y; x', y') = \delta(x - x')\delta(y - y') \quad (18)$$

and the periodic boundary conditions

$$G_E(x + na, y + mb; x', y') = G_E(x, y; x', y') \quad (19)$$

for all integer m and n .

The fact important for us later is that the rectangular billiard with periodic boundary conditions can be considered as the result of the factorization of the whole plane (x, y) with respect to the group of integer translations

$$x \rightarrow x + na, \quad y \rightarrow y + mb \quad (20)$$

with integer m and n .

The factorization of the plan (x, y) with respect to these transformations means two things. First, any two points connected by a group transformation is considered as one point. Hence (19) fulfilled. Second, inside the rectangle there is no points which are connected by these transformations. In mathematical language the rectangle with sizes (a, b) is the fundamental domain of the group (20).

Correspondingly, the exact Green function for the rectangular billiard with periodic boundary conditions equals the sum of the free Green function over all elements of the group of integer translations (20)

$$G_E(x, y; x', y') = \sum_{n, m=-\infty}^{\infty} G_E^{(0)}(x + na, y + mb; x', y') .$$

Here $G_E^{(0)}(\mathbf{x}, \mathbf{x}')$ is the Green function corresponding to the free motion without periodic boundary conditions. To prove formally that it is really the exact Green function one has to note that (i) it obeys (18) because each term in the sum obeys it, (ii) it obeys boundary conditions (19) by construction (provided the sum converges), and (iii) inside the initial rectangle only identity term can produce a δ -function contribution required in (18) because all other terms will give δ -functions outside the rectangle.

The next steps are straightforward. The free Green function for the two-dimensional Euclidean plane has the form (15). From (17) it follows that the eigenvalue density for the rectangular billiard is

$$\begin{aligned} d(E) &= -\frac{1}{\pi} \text{Im} \int_D G_E(\mathbf{x}, \mathbf{x}) d\mathbf{x} = \\ &= \frac{1}{4\pi} \sum_{mn} \int_D \text{Im} H_0^{(1)} \left(k \sqrt{(ma)^2 + (nb)^2} \right) d\mathbf{x} = \\ &= \frac{\mu(D)}{4\pi} + \frac{\mu(D)}{4\pi} \sum'_{\text{p.o.}} J_0(kL_p) \end{aligned} \quad (21)$$

which coincides exactly with (6) and (7) obtained directly from the knowledge of the eigenvalues.

The principal drawback of all trace formulas is that the sum over periodic orbits does not converge. Even the sum of the squares diverges. The simplest way to treat this problem is to multiply both sides of (21) by a suitable test function $h(E)$ and integrate them over E . In this manner one obtains

$$\sum_n h(E_n) = \frac{\mu(D)}{4\pi} \int_0^\infty h(E) dE + \frac{\mu(D)}{4\pi} \sum'_{\text{p.o.}} \int_0^\infty h(E) J_0(\sqrt{E}L_p) dE .$$

When the Fourier harmonics of $h(E)$ decrease quickly the sum over periodic orbits converges and this expression constitutes a mathematically well defined trace formula. Nevertheless for approximate calculations of eigenvalues of energies one can still use ‘naive’ trace formulas by introducing a cut-off on periodic orbit sum. For example, in Fig. 1 the result of numerical application of the above trace formula is presented. In performing this calculation one uses the asymptotic form of the oscillating part of the density of state (8) with only 250 first periodic orbits. Though additional oscillations are clearly seen, one can read off this figure the positions of first energy levels for the problem considered. In the literature many different methods of resummation of trace formulas were discussed (see e.g. [19] and references therein).

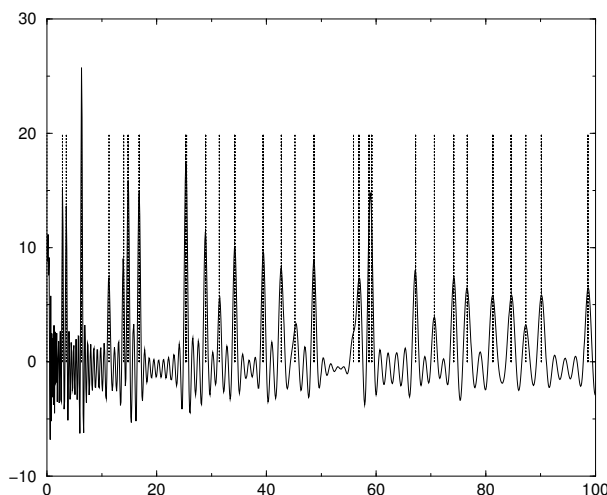


Fig. 1. The trace formula for the rectangular billiard with periodic boundary conditions calculated by taking into account 250 different periodic orbits. Dotted lines indicate the position of exact energy levels.

2 Billiards on Constant Negative Curvature Surfaces

The crucial point in the second method of derivation of the trace formula for the rectangular billiard with periodic boundary conditions was a representation of the exact Green function as a sum of a free Green function over all images of the initial point. This method of images can be applied for any problem which corresponds to a factorization of a space over the action of a discrete group. In the Euclidean plane (i.e. the space of zero curvature) there exist only a few discrete groups. Much more different discrete groups are possible in the constant negative curvature (hyperbolic) space. Correspondingly, one can derive the trace formula (called the Selberg trace formula) for all hyperbolic surfaces generated by discrete groups.

The exposition of this Section follows closely [20]. In Sect. 2.1 hyperbolic geometry is non-formally discussed. The important fact is that on hyperbolic plane there exist an infinite number of discrete groups (see e.g. [42]). Their properties are mentioned in Sect. 2.2. In Sect. 2.3 the classical mechanics on hyperbolic surfaces is considered and in Sect. 2.4 the notion of quantum problems on such surfaces is introduced. The construction of the Selberg trace formula for hyperbolic surfaces generated by discrete groups consists of two steps. The first is the explicit calculation of the free hyperbolic Green function performed in Sect. 2.5. The second step includes the summation over all group transformations. In Sect. 2.6 it is demonstrated that the identity group element gives the mean density of states. Other group elements contribute to the oscillating part of the level density and correspond to classical periodic orbits for the motion on systems considered. The relation between group ele-

ments and periodic orbits is not unique. All conjugated matrices correspond to one periodic orbit. The summation over classes of conjugated elements is done in Sect. 2.7. Performing necessary integrations in Sect. 2.8 one gets the famous Selberg trace formula. Using this formula in Sect. 2.9 we compute the asymptotic density of periodic orbits for discrete groups. In Sect. 2.10 the construction of the Selberg zeta function is presented. The importance of this function follows from the fact that its non-trivial zeros coincide with eigenvalues of the Laplace–Beltrami operator automorphic with respect to a discrete group (see Sect. 2.11). Though the Selberg zeta function is defined formally only in a part of the complex plane, it obeys a functional equation (Sect. 2.12) which permits the analytical continuation to the whole complex plane.

2.1 Hyperbolic Geometry

The standard representation of the constant negative curvature space is the Poincaré upper half plane model (x, y) with $y > 0$ (see e.g. [7] and [42]) with the following metric form

$$ds^2 = \frac{1}{y^2}(dx^2 + dy^2).$$

The geodesic in this space (= the straight line) connecting two points is the arc of circle perpendicular to the abscissa axis which passes through these points (see Fig. 2). The distance $d(\mathbf{x}, \mathbf{y})$ between two points $\mathbf{x} = (x_1, y_1)$ and

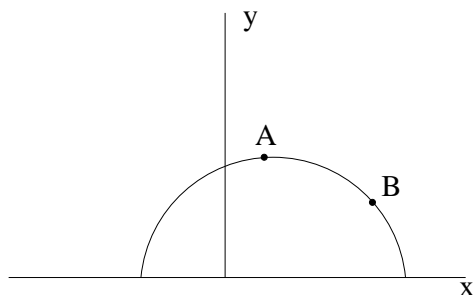


Fig. 2. The Poincaré model of constant negative curvature space. Solid line indicates the geodesic passing through points A and B.

$\mathbf{y} = (x_2, y_2)$ is defined as the length of the geodesic connecting these points. Explicitly

$$\cosh d(\mathbf{x}, \mathbf{y}) = 1 + \frac{(x_1 - x_2)^2 + (y_1 - y_2)^2}{2y_1 y_2} = 1 + \frac{|z_1 - z_2|^2}{2\operatorname{Im} z_1 \operatorname{Im} z_2} \quad (22)$$

where in the last equation one combined coordinates (x, y) into a complex number $z = x + iy$.

In the Euclidean plane the distance between two points remains invariant under 3-parameter group of rotations and translations. For constant negative curvature space the distance (22) is invariant under fractional transformations

$$z \rightarrow z' = g(z) \equiv \frac{az + b}{cz + d} \quad (23)$$

with real parameters a, b, c, d . This invariance follows from the following relations

$$z'_1 - z'_2 = \frac{az_1 + b}{cz_1 + d} - \frac{az_2 + b}{cz_2 + d} = (ad - bc) \frac{z_1 - z_2}{(cz_1 + d)(cz_2 + d)},$$

and

$$y' = \frac{1}{2i}(z' - z'^*) = (ad - bc) \frac{y}{|cz + d|^2}.$$

Substituting these expressions to (22) one concludes that the distance between two transformed points z'_1, z'_2 is the same as between initial points z_1, z_2 .

As fractional transformations are not changed under the multiplication of all elements a, b, c, d by a real factor, one can normalize them by the requirement

$$ad - bc = 1.$$

In this case the distance preserving transformations are described by 2×2 matrices with real elements and unit determinant

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \text{and} \quad \det g \equiv ad - bc = 1.$$

It is easy to check that the result of two successive fractional transformations (23) corresponds to the usual multiplication of the corresponding matrices.

The collection of all such matrices forms a group called the projective special linear group over reals and it is denoted by $\operatorname{PSL}(2, \mathbb{R})$. ‘Linear’ in the name means that it is a matrix group, ‘special’ indicates that the determinant equals 1, and ‘projective’ here has to remind that fractional transformations (23) are not changed when all elements are multiplied by ± 1 which is equivalent that two matrices $\pm \mathbf{1}$ corresponds to the identity element of the group.

The free classical motion on the constant negative curvature surface is defined as the motion along geodesics (i.e. circles perpendicular to the abscissa axis). The measure invariant under fractional transformations is the following differential form

$$d\mu = \frac{dx dy}{y^2}. \quad (24)$$

This measure is invariant in the sense that if two regions, D and D' , are related by a transformation (23), $D' = g(D)$, measures of these two regions are equal, $\mu(D') = \mu(D)$.

The operator invariant with respect to distance preserving transformations (23) is called the Laplace–Beltrami operator and it has the following form

$$\Delta_{LB} = y^2 \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right). \quad (25)$$

Its invariance means that

$$\Delta_{LB} f(g(z)) = g(\Delta_{LB} f(z))$$

for any fractional transformation $g(z)$.

Practically all notions used for the Euclidean space can be translated to the constant negative curvature case (see e.g. [7]).

2.2 Discrete groups

A rectangle (a torus) considered in Sect. 1 was the result of the factorization of the free motion on the plane by a discrete group of translations (20). Exactly in the same way one can construct a finite constant negative surface by factorizing the upper half plane by the action of a discrete group $\in PSL(2, \mathbb{R})$.

A group is discrete if (roughly speaking) there is a finite vicinity of every point of our space such that the results of all the group transformations (except the identity) lie outside this vicinity. The images of a point cannot approach each other too close.

Example

The group of transformation of the unit circle into itself. The group consists of all transformations of the following type

$$z \rightarrow g(n)z, \quad \text{and} \quad g(n) = \exp(2\pi i \alpha n),$$

where α is a constant and n is an integer. If α is a rational number $\alpha = M/N$, $g(n)$ can take only a finite number of values ($g(n)^N = 1$) and the corresponding group is discrete. But if α is an irrational number, the images of any point cover the whole circle uniformly and the group is not discrete.

Modular Group

Mathematical fact: in the upper half plane there exists an infinite number of discrete groups (see e.g. [42]). As an example let us consider the group of 2×2 integer matrices with unit determinant

$$\begin{pmatrix} m & n \\ k & l \end{pmatrix}, \quad m, n, k, l \text{ are integers and } ml - nk = 1.$$

This is evidently a group. It is called the modular group $\text{PSL}(2, \mathbb{Z})$ (\mathbb{Z} means integers) and it is one of the most investigated groups in mathematics.

This group is generated by the translation $T : z \rightarrow z + 1$ and the inversion $S : z \rightarrow -1/z$ (see e.g. [42]) which are represented by the following matrices

$$T : \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad S : \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

These matrices obey defining relations

$$S^2 = -1, \quad (ST)^3 = 1$$

and are generators in the sense that any modular group matrix can be represented as a product of a certain sequence of matrices corresponding to S and T .

Fundamental Region

Similarly to the statement that the rectangular billiard is a fundamental domain of integer translations, one can construct a fundamental domain for any discrete group.

By definition the fundamental domain of a group is defined as a region on the upper half plane such that (i) for all points outside the fundamental domain there exists a group transformation that puts it to fundamental domain and (ii) no two points inside the fundamental domain are connected by group transformations.

The fundamental domain for the modular group is presented in Fig. 3. In general, the fundamental region of a discrete group has a shape of a polygon built from segments of geodesics. Group generators identify corresponding sides of the polygon.

2.3 Classical Mechanics

Assume that we have a discrete group G with corresponding matrices $M \in G \in \text{PSL}(2, \mathbb{R})$

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

The factorization over action of the group means that points z and z' where

$$z' = \frac{az + b}{cz + d} \tag{26}$$

are identified i.e. they are considered as one point. The classical motion on the resulting surface is the motion (with unit velocity) on geodesics (semi-circles

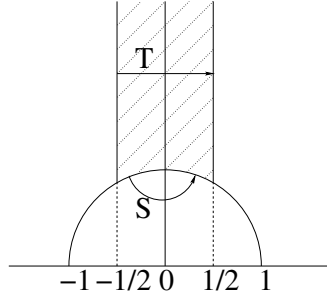


Fig. 3. Fundamental domain of the modular group. The indicated parts are identified by the corresponding generators

perpendicular to the real axis) inside the fundamental domain but when a trajectory hits a boundary it reappears from the opposite side as prescribed by boundary identifications.

For each hyperbolic matrix $M \in G$ with $|\text{Tr } M| > 2$ one can associate a periodic orbit defined as a geodesics which remains invariant under the corresponding transformation. The equation of such invariant geodesics has the form

$$c(x^2 + y^2) + (d - a)x - b = 0 . \tag{27}$$

This equation is the only function which has the following property. If $z = x + iy$ belongs to this curve then

$$z' = \frac{az + b}{cz + d}$$

also belongs to it.

The length of the periodic orbit is the distance along these geodesics between a point and its image. Let z' as above be the result of transformation (26) then the distance between z and z' is

$$\cosh l_p = 1 + \frac{|z - z'|^2}{2yy'}$$

But $y' = y/|cz + d|^2$ and

$$z - \frac{az + b}{cz + d} = \frac{c(x + iy)^2 - (d - a)(x + iy) - b}{cz + d} = y \frac{-2cy + i(d - a + 2cx)}{cz + d}$$

Here we have used the fact that point z belongs to the periodic orbit (i.e. its coordinates obey (27)). Therefore

$$\begin{aligned} \cosh l_p &= 1 + \frac{1}{2} | - 2cy + i(d - a + 2acx) |^2 = \\ &= 1 + \frac{1}{2} [4bc + (d - a)^2] = \frac{1}{2} (a + d)^2 - 1 . \end{aligned}$$

Notice that the length of periodic orbit does not depend on an initial point and is a function only of the trace of the corresponding matrix.

Finally one gets

$$2 \cosh \frac{l_p}{2} = |\text{Tr } M| . \tag{28}$$

Periodic orbits are defined only for hyperbolic matrices with $|\text{Tr } M| > 2$. For discrete groups only a finite number of elliptic matrices with $|\text{Tr } M| < 2$ can exist (see [42]).

To each hyperbolic group matrix one can associate only one periodic orbit but each periodic orbit corresponds to infinite many group matrices. It is due to the fact that z and $g(z)$ for any group transformation has to be considered as one point. Therefore all matrices in the form

$$SMS^{-1}$$

for all $S \in G$ gives one periodic orbit. These matrices form a class of conjugated matrices and periodic orbits of the classical motion are in one-to-one correspondence with classes of conjugated matrices.

2.4 Quantum Problem

The natural ‘quantum’ problem on hyperbolic plane consists in considering the same equation as in (1) but with the substitution of the invariant Laplace–Beltrami operator (25) instead of the usual Laplace operator

$$\left(y^2 \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) + E_n \right) \Psi_n(x, y) = 0$$

for the class of functions invariant (= automorphic) with respect to a given discrete group G

$$\Psi_n(x', y') = \Psi_n(x, y)$$

where $z' = x' + iy'$ is connected with $z = x + iy$ by group transformations

$$z' = \frac{az + b}{cz + d} .$$

It is easy to check that the Laplace–Beltrami operator (25) is self-adjoint with respect to the invariant measure (24), i.e.

$$\int \Psi^* (\Delta \Psi) d\mu = \int (\Delta \Psi^*) \Psi d\mu$$

and all eigenvalues E_n are real and $E_n \geq 0$.

2.5 Construction of the Green Function

As in the case of plain rectangular billiards the construction of the Green function requires two main steps.

- The computation of the exact Green function for the free motion on the whole upper half plane.
- The summation of the free Green function over all images of the initial point under group transformations.

The free hyperbolic Green function obeys the equation

$$(\Delta_{LB} + E)G_E^{(0)}(\mathbf{x}, \mathbf{x}') = \delta(\mathbf{x} - \mathbf{x}')$$

and should depend only on the (hyperbolic) distance between points \mathbf{x}, \mathbf{x}'

$$y = \cosh d(\mathbf{x}, \mathbf{x}') = 1 + \frac{(x - x')^2 + (y - y')^2}{2yy'}.$$

After simple calculations one gets that $G(y)$ with $y \neq 0$ obeys the equation for the Legendre functions (see e.g. [32], Vol.1, Sect. 3)

$$(1 - y^2) \frac{d^2 G}{dy^2} - 2y \frac{dG}{dy} + l(l + 1)G = 0$$

where

$$E = \frac{1}{4} + k^2 = -l(l + 1)$$

and

$$l = -\frac{1}{2} - ik.$$

As for the plane case the required solution of the above equation should grow as e^{ikd} when $d \rightarrow \infty$ and should behave like $\ln d/2\pi$ when $d \rightarrow 0$. From [32], Vol.1, Sect. 3 it follows that

$$G_E^{(0)}(\mathbf{x}, \mathbf{x}') = -\frac{1}{2\pi} Q_{-\frac{1}{2}-ik}(\cosh d(\mathbf{x}, \mathbf{x}')).$$

Here $Q_{-\frac{1}{2}-ik}(\cosh d)$ is the Legendre function of the second kind with the integral representation [32], Vol. 1 (3.7.4)

$$Q_{-\frac{1}{2}-ik}(\cosh d) = \frac{1}{\sqrt{2}} \int_d^\infty \frac{e^{ikr} dr}{\sqrt{\cosh r - \cosh d}}$$

and the following asymptotics

$$Q_{-\frac{1}{2}-ik}(\cosh d) \xrightarrow{d \rightarrow 0} -\log d$$

and

$$Q_{-\frac{1}{2}-ik}(\cosh d) \xrightarrow{d \rightarrow \infty} \sqrt{\frac{\pi}{2k \sinh d}} e^{i(kd - \pi/4)} .$$

The automorphic Green function is the sum over all images of one of the points

$$G_E(\mathbf{x}, \mathbf{x}') = \sum_g G_E^{(0)}(\mathbf{x}, g(\mathbf{x}'))$$

where the summation is performed over all group transformations.

2.6 Density of State

Using the standard formula (17)

$$d(E) = -\frac{1}{\pi} \int_D \text{Im } G_E(\mathbf{x}, \mathbf{x}) d\mu$$

one gets the expression for the density of states as the sum over all group elements

$$d(E) = \frac{1}{2\sqrt{2}\pi^2} \sum_g \int_D \frac{dx dy}{y^2} \left(\int_{d(z, g(z))}^{\infty} \frac{\sin kr dr}{\sqrt{\cosh r - \cosh d(z, g(z))}} \right) .$$

Mean Density of States

The mean density of states corresponds to the identity element of our group. In this case $g(z) = z$ and $d(z, g(z)) = 0$. Therefore

$$\begin{aligned} \bar{d}(E) &= \frac{1}{2\sqrt{2}\pi^2} \int_D \frac{dx dy}{y^2} \int_0^{\infty} \frac{\sin kr}{\sqrt{\cosh r - 1}} dr \\ &= \frac{\mu(D)}{(2\pi)^2} \int_0^{\infty} \frac{\sin kr}{\sinh(r/2)} dr \end{aligned}$$

where

$$\mu(D) = \int_D \frac{dx dy}{y^2}$$

is the (hyperbolic) area of the fundamental domain.

The last integral is

$$\int_0^{\infty} \frac{\sin kr}{\sinh(r/2)} dr = \pi \tanh \pi k$$

and the mean density of states takes the form

$$\bar{d}(E) = \frac{\mu(D)}{4\pi} \tanh \pi k .$$

When $k \rightarrow \infty$ it tends to $\mu(D)/4\pi$ as for the plane case.

2.7 Conjugated Classes

The most tedious step is the computation of the contribution from non-trivial fractional transformations.

Let us divide all group matrices into classes of conjugated elements. It means that all matrices having the form

$$g' = SgS^{-1}$$

where S belong to the group are considered as forming one class.

Two classes either have no common elements or coincide. This statement is a consequence of the fact that if

$$S_1g_1S_1^{-1} = S_2g_2S_2^{-1}$$

then $g_2 = S_3g_1S_3^{-1}$ where $S_3 = S_1^{-1}S_2$. Therefore g_2 belongs to the same class as g_1 and group matrices are splitted into classes of mutually non-conjugated elements.

The summation over group elements can be rewritten as the double sum over classes of conjugated elements and the elements in each class. Let g be a representative of a class. Then the summation over elements in this class is

$$\sum_S \int_D f(z, SgS^{-1}(z))d\mu$$

and the summation is performed over all group matrices S provided there is no double counting in the sum. The latter means that matrices S should be such that they do not contain matrices for which

$$S_1gS_1^{-1} = S_2gS_2^{-1}$$

or the matrix $S_3 = S_1^{-1}S_2$ commutes with matrix g

$$S_3g = gS_3 .$$

Denote the set of matrices commuting with g by S_g . They form a subgroup of the initial group G as their products also commute with g . To ensure the unique decomposition of group matrices into non-overlapping classes of conjugated elements the summation should be performed over matrices S such that no two of them can be represented as

$$S_2 = sS_1$$

and s belongs to S_g . This is equivalent to the statement that we sum over all matrices but the matrices sS are considered as one matrix. It means that we factorize the group over S_g and consider the group G/S_g .

As the distance is invariant under simultaneous transformations of both coordinates

$$d(z, z') = d(S(z), S(z'))$$

one has

$$d(z, g(z)) = d(S(z), Sg(z)) = d(y, SgS^{-1}(y))$$

where $y = S(z)$.

These relations give

$$\int_D f(d(y, SgS^{-1}(y)))d\mu = \int_{S^{-1}(D)} f(z, g(z))d\mu$$

and the last integral is taken over the image of the fundamental domain under the transformation S^{-1} . Therefore

$$\sum_S \int_D f(d(y, SgS^{-1}(y)))d\mu = \sum_S \int_{S^{-1}(D)} f(d(z, g(z)))d\mu .$$

For different S images $S^{-1}(D)$ are different and do not overlap. The integrand does not depend on S and

$$\sum_S \int_D f(d(y, SgS^{-1}(y)))d\mu = \int_{D_g} f(d(z, g(z)))d\mu$$

where

$$D_g = \sum_S S^{-1}(D) .$$

The sum of all images $S^{-1}(D)$ will cover the whole upper half plane but we have to sum not over all S but only over S factorized by the action the group of matrices commuting with a fixed matrix g . Therefore the sum will be a smaller region.

Any matrix g can be written as a power of a primitive element

$$g = g_0^n$$

and it is (almost) evident that matrices commuting with g are precisely the group of matrices generated by g_0 . This is a cyclic abelian group consisting of all (positive, negative, and zero) powers of g_0

$$S_g = g_0^m, \quad m = 0, \pm 1, \pm 2, \dots$$

and as a discrete group it has a fundamental domain FD_g .

Therefore

$$\sum_{S \in G/S_g} \int_D f(d(y, SgS^{-1}(y)))d\mu = \int_{FD_g} f(d(z, g(z)))d\mu .$$

In the left hand side the integration is taken over the fundamental domain of the whole group G and the summation is done over matrices from G factorized by the subgroup S_g of matrices which commutes with a fixed matrix g . In the right hand side there is no summation but the integration is performed over the (large) fundamental domain of the subgroup S_g .

2.8 Selberg Trace Formula

We have demonstrated that the density of states of the hyperbolic Laplace–Beltrami operator automorphic over a discrete group can be represented as

$$d(E) = \bar{d}(E) + \sum_g d_g(E)$$

where

$$d_g(E) = \frac{1}{2\sqrt{2}\pi^2} \int_{FD_g} d\mu \int_{d(z,g(z))}^{\infty} \frac{\sin kr}{\sqrt{\cosh r - \cosh d(z,g(z))}} dr$$

and the summation is performed over classes of conjugated matrices.

Let us consider the case of hyperbolic matrices $g = g_0^m$ (i.e. $|\text{Tr } g_0| > 2$). By a suitable matrix B such matrix can be transform to the diagonal form

$$Bg_0B^{-1} = \begin{pmatrix} \lambda_0 & 0 \\ 0 & \lambda_0^{-1} \end{pmatrix}.$$

For hyperbolic matrices λ_0 is real and $|\lambda_0| > 1$. By the same transformation the matrix g will be transformed to

$$BgB^{-1} = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix}$$

and $\lambda = \lambda_0^m$.

Assume that g is in the diagonal form. Then $g(z) = \lambda^2 z$ and

$$\cosh d(z, g(z)) = 1 + \frac{(\lambda^2 - 1)^2(x^2 + y^2)}{2\lambda^2 y^2}.$$

Because λ_0 is real the transformation $z' = \lambda_0^2 z$ gives $y' = \lambda_0^2 y$ and the fundamental domain of $S_g = \lambda_0^{2m} z$ has the form of a horizontal strip $1 < y < \lambda_0^2$ indicated in Fig. 4. Now

$$d_g(E) = \int_{-\infty}^{\infty} dx \int_1^{\lambda_0^2} F\left(\frac{(\lambda^2 - 1)^2(x^2 + y^2)}{\lambda^2 y^2}\right) \frac{dy}{y^2}.$$

Introducing a new variable $\xi = xy$ one gets

$$\begin{aligned} d_g(E) &= \int_1^{\lambda_0^2} \frac{dy}{y} \int_{-\infty}^{\infty} F\left((1 + \xi^2) \frac{(\lambda^2 - 1)^2}{\lambda^2}\right) d\xi = \\ &= \ln \lambda_0^2 \int_{-\infty}^{\infty} F\left((1 + \xi^2) \frac{(\lambda^2 - 1)^2}{\lambda^2}\right) d\xi. \end{aligned}$$

After the substitution

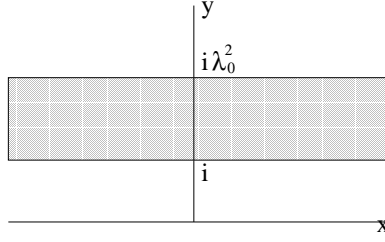


Fig. 4. Fundamental domain of multiplication group

$$u = (1 + \xi^2) \frac{(\lambda^2 - 1)^2}{\lambda^2}$$

one obtains

$$d_g(E) = \frac{\ln \lambda_0^2}{\sqrt{u_0}} \int_{u_0}^{\infty} \frac{F(u)}{\sqrt{u - u_0}} du$$

where

$$u_0 = \frac{(\lambda^2 - 1)^2}{\lambda^2} = \lambda^2 + \frac{1}{\lambda^2} - 2.$$

The variable u is connected with the distance by $\cosh d = 1 + u/2$ and the function $F(\cosh d)$ has the form

$$F(\cosh d) = \frac{1}{2\sqrt{2}\pi^2} \int_d^{\infty} \frac{\sin kr}{\sqrt{\cosh r - \cosh d}} dr.$$

Introduce a variable τ connected with r as u is connected with d

$$\cosh \tau = 1 + \frac{r}{2}, \quad \frac{dr}{d\tau} = \frac{1}{\sqrt{\tau^2 + 4\tau}}.$$

It gives

$$F(u) = \frac{1}{2\pi^2} \int_u^{\infty} \frac{\sin kr(\tau)}{\sqrt{(\tau - u)(\tau^2 + 4\tau)}} d\tau$$

and

$$d_g(E) = \frac{\ln \lambda_0^2}{2\pi^2 \sqrt{u_0}} f(u_0)$$

where

$$f(w) = \int_w^{\infty} \frac{du}{\sqrt{u - w}} \int_u^{\infty} \frac{\sin kr(\tau)}{\sqrt{(\tau - u)(\tau^2 + 4\tau)}} d\tau.$$

Changing the order of integration one obtains

$$f(w) = \int_w^\infty \frac{\sin kr(\tau)}{\sqrt{\tau^2 + 4\tau}} d\tau \int_w^\tau \frac{du}{\sqrt{(u-w)(\tau-u)}}.$$

The last integral is a half of the residue at infinity

$$\int_w^\tau \frac{du}{\sqrt{(u-w)(\tau-u)}} = \pi$$

and

$$f(w) = \pi \int_w^\infty \frac{\sin kr(\tau)}{\sqrt{\tau^2 + 4\tau}} d\tau = \pi \int_{l_p}^\infty \sin(kr) dr = \frac{\pi}{k} \cos kl_p.$$

Here l_p is the minimal value of r corresponding to u_0

$$\cosh l_p = 1 + \frac{u_0}{2} = 1 + \frac{1}{2}(\lambda^2 + \frac{1}{\lambda^2} - 2) = \frac{1}{2}(\lambda + \frac{1}{\lambda})^2 - 1$$

or

$$2 \cosh l_p = \lambda + \frac{1}{\lambda} \equiv \text{Tr } g$$

i.e. l_p is the length of periodic orbit associated with the matrix g .

Therefore

$$d_g(E) = \frac{\ln \lambda_0^2}{2\pi k \sqrt{\lambda + \lambda^{-1} - 2}} \cos kl_p = \frac{l_p^{(0)}}{4\pi k \sinh l_p/2} \cos kl_p$$

where $l_p^{(0)}$ is the length of the primitive periodic orbit associated with g_0 .

Combining all terms together one finds that the eigenvalues density of the Laplace–Beltrami operator automorphic with respect to a discrete group with only hyperbolic matrices has the form

$$d(E) = \frac{\mu(D)}{4\pi} \tanh \pi k + \sum_{\text{p.p.o.}} \frac{l_p}{4\pi k} \sum_{n=1}^{\infty} \frac{\cos(knl_p)}{\sinh(nl_p/2)}.$$

The oscillating part of the density is given by the double sum. The first summation is done over all primitive periodic orbits (p.p.o.) and the second sum is performed over all repetitions of these orbits. Here k is the momentum related with the energy by $E = k^2 + 1/4$.

To obtain mathematically sound formula and to avoid problems with convergence it is common to multiply both parts of the above equality by a test function $h(k)$ and to integrate over $dE = 2kdk$. To assume the convergence the test function $h(r)$ should have the following properties

- The function $h(r)$ is a function analytical in the region $|\text{Im } r| \leq 1/2 + \delta$ with certain $\delta > 0$.
- $h(-r) = h(r)$.
- $|h(r)| \leq A(1 + |r|)^{-2-\delta}$.

The left hand side of the above equation is

$$\int d(E)h(k)dE = \sum_n \delta(E - E_n)h(k)dE = \sum_n h(k_n) .$$

In the right hand side one obtains

$$\int h(k) \frac{\cos kl}{2\pi k} k dk = \frac{1}{2\pi} \int_{-\infty}^{\infty} h(k) e^{-ikl} dk .$$

The final formula takes the form

$$\begin{aligned} \sum_n h(k_n) &= \frac{\mu(D)}{2\pi} \int_{-\infty}^{\infty} kh(k) \tanh(\pi k) dk + \\ &+ \sum_{\text{p.p.o.}} l_p \sum_{n=1}^{\infty} \frac{1}{2 \sinh(nl_p/2)} g(nl_p) \end{aligned} \tag{29}$$

where k_n is related with eigenvalue E_n as follows

$$E_n = k_n^2 + \frac{1}{4}$$

and $g(l)$ is the Fourier transform of $h(k)$

$$g(l) = \frac{1}{2\pi} \int_{-\infty}^{\infty} h(k) e^{-ikl} dk .$$

This is the famous Selberg trace formula. It connects eigenvalues of the Laplace–Beltrami operator for functions automorphic with respect to a discrete group having only hyperbolic elements with classical periodic orbits.

2.9 Density of Periodic Orbits

To find the density of periodic orbits for a discrete group let us choose the test function $h(r)$ in (29) as

$$h(r) = e^{-(r^2+1/4)T} \equiv e^{-ET}$$

with a parameter $T > 0$. Its Fourier transforms is

$$g(u) = \frac{1}{2\pi} \int_{-\infty}^{\infty} h(k) e^{-iku} dk = \frac{e^{-T/4}}{2\sqrt{\pi T}} e^{-u^2/4T} .$$

In the left hand side of the Selberg trace formula one obtains

$$\sum_n e^{-E_n T} = 1 + \sum_{E_n > 0} e^{-E_n T}$$

where we take into account that for any discrete group there is one zero eigenvalue corresponding to a constant eigenfunction. Therefore when $T \rightarrow \infty$ the above sum tends to one

$$\sum_n e^{-E_n T} \xrightarrow{T \rightarrow \infty} 1.$$

One can easily check that in the right hand side of (29) the contribution of the smooth part of the density goes to zero at large T and the contribution of periodic orbits is important only for primitive periodic orbits with $n = 1$. The latter is

$$\frac{e^{-T/4}}{2\sqrt{\pi T}} \sum_p l_p e^{-l_p^2/4T - l_p/2} = \frac{e^{-T/4}}{2\sqrt{\pi T}} \int_0^\infty l e^{-l^2/4T - l/2} \rho(l) dl$$

where $\rho(l)$ is the density of periodic orbits. Hence the Selberg trace formula states that

$$\lim_{T \rightarrow \infty} \frac{e^{-T/4}}{2\sqrt{\pi T}} \int_0^\infty l e^{-l^2/4T - l/2} \rho(l) dl = 1.$$

Assume that $\rho(l) = be^{al}/l$ with certain constants a and b . Then from the above limit it follows that $a = b = 1$ which demonstrates that the density of periodic orbits for a discrete group increases exponentially with the length

$$\rho(l) = \frac{e^l}{l}.$$

2.10 Selberg Zeta Function

Amongst many applications of the Selberg trace formula let us consider the construction of the Selberg zeta function.

Choose as test function $h(k)$ the function

$$h(k) = \frac{1}{k^2 + \alpha^2} - \frac{1}{k^2 + \beta^2}.$$

Its Fourier transform is

$$g(l) = \frac{1}{2\alpha} e^{-\alpha|l|} - \frac{1}{2\beta} e^{-\beta|l|}.$$

The Selberg trace formula gives

$$\begin{aligned} & \sum_n \left(\frac{1}{k_n^2 + \alpha^2} - \frac{1}{k_n^2 + \beta^2} \right) = \\ & = \frac{\mu(D)}{2\pi} \int_{-\infty}^\infty k \tanh \pi k \left(\frac{1}{k^2 + \alpha^2} - \frac{1}{k^2 + \beta^2} \right) dk + \\ & + \sum_{\text{p.p.o.}} \sum_{n=1}^\infty \frac{l_p}{2 \sinh nl_p/2} \left(\frac{e^{-\alpha l_p}}{2\alpha} - \frac{e^{-\beta l_p}}{2\beta} \right). \end{aligned}$$

The Selberg zeta function is defined as the following formal product

$$Z(s) = \prod_{\text{p.p.o.}} \prod_{m=0}^{\infty} (1 - e^{-l_p(s+m)}). \quad (30)$$

One has

$$\begin{aligned} \frac{1}{Z} \frac{dZ}{ds} &= \sum_{\text{p.p.o.}} \sum_{m=0}^{\infty} \frac{l_p e^{-l_p(s+m)}}{1 - e^{-l_p(s+m)}} = \sum_{\text{p.p.o.}} l_p \sum_{n=1}^{\infty} \sum_{m=0}^{\infty} e^{-l_p(s+m)n} = \\ &= \sum_{\text{p.p.o.}} l_p \sum_{n=1}^{\infty} \frac{e^{-l_p n s}}{1 - e^{-l_p n}} = \sum_{\text{p.p.o.}} l_p \sum_{n=1}^{\infty} \frac{1}{2 \sinh n l_p / 2} e^{-l_p n (s-1/2)}. \end{aligned}$$

Choose $\alpha = s - 1/2$ and $\beta = s' - 1/2$ then

$$\begin{aligned} &\sum_n \left(\frac{1}{k_n^2 + (s-1/2)^2} - \frac{1}{k_n^2 + (s'-1/2)^2} \right) = \\ &= \frac{\mu(D)}{4\pi} \int_{-\infty}^{\infty} k \tanh \pi k \left(\frac{1}{k^2 + (s-1/2)^2} - \frac{1}{k^2 + (s'-1/2)^2} \right) dk + \\ &+ \frac{1}{2s-1} \frac{Z(s)'}{Z(s)} - \frac{1}{2s'-1} \frac{Z(s')'}{Z(s')}. \end{aligned}$$

The integral in the right hand side can be computed by the residues

$$\int_{-\infty}^{\infty} k \tanh \pi k \left(\frac{1}{k^2 + (s-1/2)^2} - \frac{1}{k^2 + (s'-1/2)^2} \right) dk = f(s) - f(s')$$

where $f(s)$ is the sum over residues from one pole $k = i(s - 1/2)$ and from poles $k_n = i(n + 1/2)$ of $\tanh \pi k$

$$\begin{aligned} f(s) &= 2\pi i \left[\frac{1}{2} \tanh[i\pi(s-1/2)] + \frac{i}{\pi} \sum_{n=0}^{\infty} \frac{n+1/2}{(s-1/2)^2 - (n+1/2)^2} \right] = \\ &= \pi \cot \pi s - \sum_{n=1}^{\infty} \frac{1}{s-n} + \sum_{n=1}^{\infty} \frac{1}{s+n}. \end{aligned}$$

But

$$\pi \cot \pi s = \sum_{n=1}^{\infty} \frac{1}{s-n} + \sum_{n=1}^{\infty} \frac{1}{s+n},$$

therefore

$$f(s) = 2 \sum_{n=1}^{\infty} \frac{1}{s+n}.$$

Using these relations one gets the identity valid for all values of s and s'

$$\frac{1}{2s-1} \frac{Z'(s)}{Z(s)} = \frac{1}{2s'-1} \frac{Z'(s')}{Z(s')} - \frac{\mu(D)}{2\pi} \sum_{n=0}^{\infty} \left(\frac{1}{s+n} - \frac{1}{s'+n} \right) + \sum_n \left(\frac{1}{k_n^2 + (s-1/2)^2} - \frac{1}{k_n^2 + (s'-1/2)^2} \right). \quad (31)$$

The right hand side of this identity has poles at $s = 1/2 + ik_n$ and $s = -n$. The same poles have to be present in the left hand side. If

$$\frac{Z'(s)}{Z(s)} \rightarrow \frac{\nu_k}{s - s_k}$$

then

$$Z(s) \rightarrow (s - s_k)^{\nu_k} \quad \text{when } s \rightarrow s_k.$$

When $\nu_k > 0$ (resp. $\nu_k < 0$) point s_k is a zero (resp. a pole) of the Selberg zeta function $Z(s)$.

2.11 Zeros of the Selberg Zeta Function

Combining all poles one concludes that the Selberg zeta function for a group with only hyperbolic elements have two different sets of zero. The first consists of non-trivial zeros

$$s = 1/2 \pm ik_n,$$

coming from eigenvalues of the Laplace–Beltrami operator for automorphic functions. The second set includes a zero from $E = 0$ eigenvalue and zeros from the smooth term. These zeros are called trivial zeros and they are located at points

$$s = -m \quad (m = 1, 2, \dots)$$

with multiplicity $\nu_m = (2m+1)\mu(D)/2\pi$, at point $s = 0$ with multiplicity $\nu_0 = \mu(D)/2\pi$ and a single zero at $s = 1$. These multiplicities are integers because the area of a compact fundamental domain $\mu(D) = 4\pi(g-1)$ where g is the genus of the surface.

The structure of these zeros is presented schematically at Fig. 5.

2.12 Functional Equation

The infinite product defining the Selberg zeta function (30) converges only when $\text{Re } s > 1/2$. Nevertheless the Selberg zeta function can be analytically continued to the whole complex plane s with the aid of (31).

Put $s' = 1 - s$ in (31). The sum over eigenvalues cancels and $f(s) - f(1-s) = 2\pi \cot \pi s$. Therefore

$$\frac{1}{2s-1} \left(\frac{Z'(s)}{Z(s)} + \frac{Z'(1-s)}{Z(1-s)} \right) = -\frac{\mu(D)}{2} \cot \pi s$$

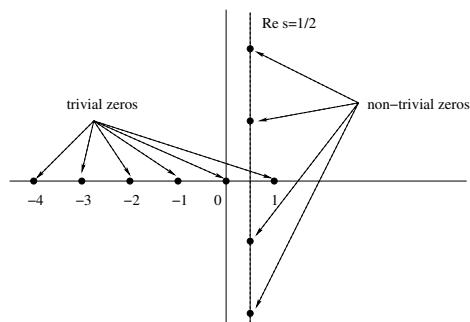


Fig. 5. Zeros of the Selberg zeta function

which is equivalent to the following relation (called functional equation)

$$Z(s) = \varphi(s)Z(1-s) \quad (32)$$

where

$$\frac{\varphi'(s)}{\varphi(s)} = -\mu(D)\left(s - \frac{1}{2}\right) \cot \pi s$$

and $\varphi(1/2) = 1$.

Explicitly

$$\varphi(s) = \exp\left(\mu(D) \int_0^{s-1/2} u \tan \pi u du\right).$$

Therefore if one knows the Selberg zeta function when $\text{Re } s > 1$ (32) gives its continuation to the mirror region $\text{Re } s < 0$.

3 Trace Formulas for Integrable Dynamical Systems

A f -dimensional system is called integrable if its classical Hamiltonian can be written as a function of action variables only

$$H(\mathbf{I}) = H(I_1, \dots, I_f).$$

In this representation the classical equations of motion take especially simple form

$$\dot{\mathbf{I}} = -\frac{\partial H}{\partial \varphi} = 0, \quad \dot{\varphi} = \frac{\partial H}{\partial \mathbf{I}} = \boldsymbol{\omega}.$$

The semiclassical quantization consists of fixing the values of the action variables

$$I_j = \hbar(n_j + \frac{\mu_j}{4})$$

where n_j are integers and μ_j are called the Maslov indices.

In this approximation eigenvalues of energy of the system are a function of these integers

$$E(\mathbf{n}) = H\left(\hbar(n_1 + \frac{\mu_1}{4}), \dots, \hbar(n_f + \frac{\mu_f}{4})\right).$$

The eigenvalue density is the sum over all integers n_j

$$d(E) = \sum_{\mathbf{n}} \delta(E - H(\hbar(\mathbf{n} + \frac{1}{4}\boldsymbol{\mu}))).$$

Using the Poisson summation formula (5) one transforms this expression as follows

$$\begin{aligned} d(E) &= \sum_{\mathbf{N}} \int e^{2\pi i \mathbf{N} \mathbf{n}} \delta(E - H(\hbar(\mathbf{n} + \frac{1}{4}\boldsymbol{\mu}))) d\mathbf{n} = \\ &= \frac{1}{\hbar^f} \sum_{\mathbf{N}} e^{-i\pi \mathbf{N} \boldsymbol{\mu} / 2} \int e^{2\pi i \mathbf{N} \mathbf{I} / \hbar} \delta(E - H(\mathbf{I})) d\mathbf{I} \end{aligned} \quad (33)$$

where the summation is taken over f integers N_j .

3.1 Smooth Part of the Density

The term with $\mathbf{N} = 0$ in (33) corresponds to the smooth part of the density

$$\bar{d}(E) = \frac{1}{\hbar^f} \int \delta(E - H(\mathbf{I})) d\mathbf{I}.$$

As $d\mathbf{I}d\boldsymbol{\varphi}$ is the canonical invariant, $d\mathbf{I}d\boldsymbol{\varphi} = d\mathbf{p}d\mathbf{q}$ where \mathbf{p} and \mathbf{q} are the momenta and coordinates and, because $\int d\boldsymbol{\varphi} = (2\pi)^f$, the formula for the smooth part of the level density can be rewritten in the Thomas-Fermi form

$$\bar{d}(E) = \int \delta(E - H(\mathbf{p}, \mathbf{q})) \frac{d\mathbf{p}d\mathbf{q}}{(2\pi\hbar)^f}. \quad (34)$$

The usual interpretation of this formula is that each quantum state occupies $(2\pi\hbar)^f$ volume on the constant energy surface. For general systems (34) represents the leading term of the expansion of the smooth part of the level density when $\hbar \rightarrow 0$. Other terms can be found e.g. in [5]. See also [14] for the resummation of such series for certain models.

3.2 Oscillating Part of the Density

In the semiclassical approximation $\hbar \rightarrow 0$ terms with $\mathbf{N} \neq 0$ in (33) can be calculated by the saddle point method. Our derivation differs slightly from the one given in [9]. First it is convenient to represent δ -function as follows

$$\delta(x) = \frac{1}{2\pi\hbar} \int_{-\infty}^{\infty} e^{i\alpha x/\hbar} d\alpha .$$

Then

$$d^{(osc)}(E) = \frac{1}{2\pi\hbar^{f+1}} \sum_{\mathbf{N}} e^{-i\pi\mathbf{N}\mu/2} \int_{-\infty}^{\infty} d\alpha \int e^{iS(\mathbf{I},\alpha)/\hbar} d\mathbf{I}$$

where the effective action, $S(\mathbf{I}, \alpha)$, is

$$S(\mathbf{I}, \alpha) = 2\pi\mathbf{N}\mathbf{I} + \alpha(E - H(\mathbf{I})) .$$

The integration over \mathbf{I} and α can be performed by the saddle point method. The saddle point values, \mathbf{I}_{sp} and α_{sp} , are determined from equations

$$\frac{\partial S}{\partial \alpha} = E - H(\mathbf{I}_{sp}) = 0 , \quad \frac{\partial S}{\partial \mathbf{I}} = 2\pi\mathbf{N} - \alpha_{sp}\boldsymbol{\omega}_{sp} = 0 .$$

The first equation shows that in the leading approximation \mathbf{I}_{sp} belongs to the constant energy surface and the second equation selects special values of \mathbf{I}_{sp} for which frequencies ω_j are commensurable

$$\boldsymbol{\omega}_{sp} = \frac{2\pi}{\alpha_{sp}} \mathbf{N} .$$

Together the saddle point conditions demonstrate that in the limit $\hbar \rightarrow 0$ the dominant contribution to the term with fixed integer vector \mathbf{N} comes from the classical periodic orbit with period

$$T_p = \alpha_{s.p}$$

and the saddle point action coincides with the classical action along this trajectory

$$S_{sp} = 2\pi\mathbf{N}\mathbf{I}_{sp} .$$

To compute remaining integrals it is necessary to expand the full action up to quadratic terms on deviations from the saddle point values. One has

$$S(\mathbf{I}_{sp} + \delta\mathbf{I}, \alpha_{s.p} + \delta\alpha) = S_{sp} + \frac{T_p}{2} (\delta I_i H_{ij} \delta I_j) - \delta\alpha (\omega_j \delta I_j)$$

where the summation over repeating indexes is assumed. H_{ij} is the matrix of the second derivatives of the Hamiltonian computed at the saddle point

$$H_{ij} \equiv \left. \frac{\partial^2 H}{\partial I_i \partial I_j} \right|_{\mathbf{I}=\mathbf{I}_{sp}} .$$

The following steps are straightforward

$$\begin{aligned}
& \int d\delta\mathbf{I}d\delta\alpha \exp\left(\frac{i}{\hbar}S(\mathbf{I}, \alpha)\right) = \\
& = e^{iS_{sp}/\hbar} \int d\delta\alpha \int d\delta\mathbf{I} \exp\left(\frac{i}{2\hbar}T_p(\delta I_i H_{ij} \delta I_j) - \frac{\delta\alpha}{\hbar}(\omega_j \delta I_j)\right) = \\
& = \left(\frac{2\pi\hbar}{T_p}\right)^{f/2} \frac{e^{iS_{sp}/\hbar}}{\sqrt{|\det H_{ij}|}} \int \delta\alpha \exp\left(-\frac{i}{2\hbar T_p}(\delta\alpha)^2(\omega_i H_{ij}^{-1} \omega_j) + \frac{i}{4}\pi\beta'\right) = \\
& = \left(\frac{2\pi\hbar}{T_p}\right)^{f/2} \frac{\sqrt{2\pi\hbar T_p}}{\sqrt{|\det H_{ij}|(\omega_k H_{kl}^{-1} \omega_l)}} \exp\left(\frac{i}{\hbar}S_{sp} + \frac{i}{4}\pi\beta\right) = \\
& = \frac{(2\pi)^{(f-1)/2} \hbar^{(f+1)/2}}{T_p^{(f-3)/2} |(N_i Q_{ij} N_j)|^{1/2}} \exp\left(\frac{i}{\hbar}S_{sp} + \frac{i}{4}\pi\beta\right)
\end{aligned}$$

where $Q_{ij} = H_{ij}^{-1} \det H$ called the co-matrix of H_{ij} is the determinant obtained from H_{ij} by omitting the i -th row and the j -th column. The phase β is the signature of H_{ij} minus the sign of $(\omega H^{-1} \omega)$.

The final expression for the oscillating part of the level density of an integrable system with a Hamiltonian $H(\mathbf{I})$ is

$$d^{(osc)}(E) = \sum_{\mathbf{N}} P_{\mathbf{N}} \exp\left(i\frac{S_p}{\hbar} - i\frac{\pi}{4}\mathbf{N}\boldsymbol{\mu} + i\frac{\pi}{4}\beta\right)$$

where $S_p = 2\pi\mathbf{N}\mathbf{I}$ is the action over a classical periodic orbit with fixed winding numbers and

$$P_{\mathbf{N}} = \left(\frac{2\pi}{\hbar T_p}\right)^{(f-3)/2} \frac{1}{\hbar^2 |(N_i Q_{ij} N_j)|^{1/2}}.$$

The summation over integer vectors \mathbf{N} is equivalent to the summation over all classical periodic orbit families of the system.

4 Trace Formula for Chaotic Systems

To compute the eigenvalue density for a chaotic system one has to start with general expression (17)

$$d(E) = -\frac{1}{\pi} \int \text{Im } G_E(\mathbf{x}, \mathbf{x}) d\mathbf{x}$$

which relates the quantum density with the Green function of the system, $G_E(\mathbf{x}, \mathbf{y})$, obeying the Schroedinger equation with a δ -function term in the right hand side

$$(E - \hat{H})G_E(\mathbf{x}, \mathbf{y}) = \delta(\mathbf{x} - \mathbf{y}) .$$

For concreteness let us consider the usual case

$$\hat{H} = -\hbar^2 \Delta + V(\mathbf{x}) .$$

The exact Green function can be computed exactly only in very limited cases. For generic systems the best which can be achieved is the calculation of the Green function in the semiclassical limit $\hbar \rightarrow 0$.

4.1 Semiclassical Green Function

Let us try to obey the Schroedinger equation in the following form (see [33])

$$G_E(\mathbf{x}, \mathbf{y}) = A(\mathbf{x}, \mathbf{y})e^{iS(\mathbf{x}, \mathbf{y})/\hbar} \quad (35)$$

where the prefactor $A(\mathbf{x}, \mathbf{y})$ can be expanded into a power series of \hbar .

Separating the real and imaginary parts of the Schroedinger equation one gets two equations

$$(E - (\nabla S)^2 - V(\mathbf{x})) + \hbar^2 \Delta A = 0$$

and

$$2\nabla S \nabla A + \Delta S A = 0 .$$

In the leading order in \hbar the first equation reduces to the Hamilton-Jacobi equation for the classical action $S(\mathbf{x}, \mathbf{y})$

$$E = (\nabla S)^2 + V(\mathbf{x}) .$$

It is well known that the solution of this equation can be obtained in the following way.

Find the solution of the usual classical equations of motion

$$\ddot{\mathbf{x}} = -\frac{\partial V}{\partial \mathbf{x}}$$

with energy E which starts at a fixed point \mathbf{y} and ends at a point \mathbf{x} . Then

$$S(\mathbf{x}, \mathbf{y}) = \int_{\mathbf{y}}^{\mathbf{x}} \mathbf{p} d\mathbf{x}$$

where \mathbf{p} is the momentum and the integral is taken over this trajectory.

Instead of proving this fact we illustrate it on an example of the free motion. The free motion equations $\ddot{\mathbf{x}} = 0$ have a general solution

$$\mathbf{x} = \mathbf{k}t + \mathbf{y}$$

with a fixed vector \mathbf{k} . One has

$$\mathbf{k} = \frac{\mathbf{x} - \mathbf{y}}{t}$$

and the conservation of energy $|\mathbf{k}|^2 = E$ determines the time of motion

$$t = \frac{|\mathbf{x} - \mathbf{y}|}{\sqrt{E}}.$$

Therefore

$$S(\mathbf{x}, \mathbf{y}) = \sqrt{E}|\mathbf{x} - \mathbf{y}|$$

which, evidently, is the solution of the free Hamilton–Jacobi equation.

The next order equation

$$2\nabla S \nabla A + \Delta S A = 0$$

is equivalent to the conservation of current. Indeed, for the semiclassical wave function (35)

$$\mathbf{J} = \frac{1}{2i}(\Psi^* \nabla \Psi - \Psi \nabla \Psi^*) = A^2 \nabla S$$

and

$$\nabla \mathbf{J} = A(2\nabla A \nabla S + A \Delta S) = 0.$$

The solution of the above transport equation has the form

$$A(\mathbf{x}, \mathbf{y}) = \frac{\pi}{(2\pi\hbar)^{(f+1)/2}} \left| \frac{1}{k_i k_f} \det \left(-\frac{\partial^2 S}{\partial t_{i\perp} \partial t_{f\perp}} \right) \right|^{1/2}$$

where $t_{i\perp}$ and $t_{f\perp}$ are coordinates perpendicular to the trajectory in the initial, \mathbf{y} , and final, \mathbf{x} , points respectively and k_i , k_f are the initial and final momenta. The derivation of this formula can be found e.g. in [33]. The overall prefactor in this formula can be fixed by comparing with the asymptotics of the free Green function (14) at large distances.

The final formula for the semiclassical Green function takes the form

$$G_E(\mathbf{x}, \mathbf{y}) = \sum_{\substack{\text{classical} \\ \text{trajectories}}} \frac{\pi}{(2\pi\hbar)^{(f+1)/2}} \left| \frac{1}{k_i k_f} \det \left(-\frac{\partial^2 S}{\partial t_{i\perp} \partial t_{f\perp}} \right) \right|^{1/2} \times \\ \times \exp \left(\frac{i}{\hbar} S_{cl}(\mathbf{x}, \mathbf{y}) - \frac{i}{4} \pi \mu \right)$$

where the sum is taken over all classical trajectories with energy E which connect points \mathbf{y} and \mathbf{x} . μ is the Maslov index which, roughly speaking, counts the number of points along the trajectory where semiclassical approximation cannot be applied.

4.2 Gutzwiller Trace Formula

The knowledge of the Green function permits the calculation of the density of eigenstates by the usual formula (17)

$$d(E) = -\frac{1}{\pi} \int \text{Im } G_E(\mathbf{x}, \mathbf{x}) d\mathbf{x} .$$

The Green function $G_E(\mathbf{x}, \mathbf{y})$ at points \mathbf{x} and \mathbf{y} very close to each other has two different contributions (see Fig. 6). The first comes from very short

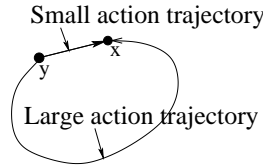


Fig. 6. Small and large action contributions to the Green function for nearby points

trajectories where semiclassical approximation cannot, in general, be applied. The second is related with long trajectories. The first contribution can be computed by using the Thomas–Fermi (local) approximation for the Green function. In this approximation one uses the local formula (cf. (16))

$$G_E(\mathbf{x}, \mathbf{y}) \xrightarrow{\mathbf{y} \rightarrow \mathbf{x}} \int \frac{d\mathbf{p}}{(2\pi\hbar)^f} \frac{e^{i\mathbf{p}(\mathbf{x}-\mathbf{y})/\hbar}}{(E - H(\mathbf{p}, \mathbf{x}) + i\epsilon)} .$$

Therefore

$$\text{Im } G_E(\mathbf{x}, \mathbf{x}) = -\pi \int \frac{d\mathbf{p}}{(2\pi\hbar)^f} \delta(E - H(\mathbf{p}, \mathbf{x}))$$

and the smooth part of the level density in the leading approximation equals the phase-space volume of the constant energy surface divided by $(2\pi\hbar)^f$

$$\bar{d}(E) = \int \frac{d\mathbf{p}d\mathbf{x}}{(2\pi\hbar)^f} \delta(E - H(\mathbf{p}, \mathbf{x})) .$$

The contributions from long classical trajectories with finite actions corresponds to the oscillating part of the density and can be calculated using the semiclassical approximation of the Green function (35).

One has

$$d^{(osc)}(E) = -\frac{1}{\pi} \operatorname{Im} \sum_{\substack{\text{classical} \\ \text{trajectories}}} \int A(\mathbf{x}, \mathbf{x}) e^{iS(\mathbf{x}, \mathbf{x})/\hbar} d\mathbf{x}.$$

When $\hbar \rightarrow 0$ the integration can be performed in the saddle point approximation. The saddles are solutions of the equation

$$\left[\frac{\partial S(\mathbf{x}, \mathbf{y})}{\partial \mathbf{x}} + \frac{\partial S(\mathbf{x}, \mathbf{y})}{\partial \mathbf{y}} \right]_{\mathbf{y}=\mathbf{x}} = 0.$$

But

$$\frac{\partial S(\mathbf{x}, \mathbf{y})}{\partial \mathbf{x}} = \mathbf{k}_f, \quad \frac{\partial S(\mathbf{x}, \mathbf{y})}{\partial \mathbf{y}} = -\mathbf{k}_i$$

where \mathbf{k}_f and \mathbf{k}_i are the momenta in the final and initial points respectively.

Hence the saddle point equations select special classical orbits which start and end in the same point with the same momentum. It means that the saddles are classical periodic orbits of the system and

$$S_{sp} = S_p.$$

To calculate the integral around one particular periodic orbit it is convenient to split the integration over the whole space to one integration along the orbit and $(f-1)$ integrations in directions perpendicular to the orbit. For simplicity we consider the two-dimensional case.

The change of the action when a point is at the distance y from the periodic orbit is

$$\delta S = \frac{1}{2} y^2 \frac{\partial^2 S(y, y)}{\partial y^2} \Big|_{y=0}$$

where $S(y, y)$ is the classical action for a classical orbit in a vicinity of the periodic orbit (see Fig. 7). To compute such derivatives it is useful to use

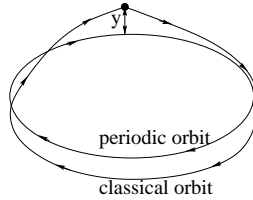


Fig. 7. A periodic orbit and a closed classical orbit in its vicinity

the monodromy matrix, m_{ij} , which relates initial and final coordinates and momenta in a vicinity of periodic orbit in the linear approximation

$$\begin{pmatrix} \delta y_f \\ \delta p_f \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \begin{pmatrix} \delta y_i \\ \delta p_i \end{pmatrix}.$$

As the classical motion preserves the canonical invariant $dpdq$ it follows that $\det M = 1$.

One has

$$\begin{aligned} \delta y_f &= m_{11}\delta y_i + m_{12}\delta p_i, \\ \delta p_f &= m_{21}\delta y_i + m_{22}\delta p_i. \end{aligned}$$

But

$$p_i = -\frac{\partial S}{\partial y_i}, \quad p_f = \frac{\partial S}{\partial y_f}.$$

Therefore

$$\delta p_i = -\frac{\partial^2 S}{\partial y_i^2}\delta y_i - \frac{\partial^2 S}{\partial y_i \partial y_f}\delta y_f, \quad \delta p_f = \frac{\partial^2 S}{\partial y_i \partial y_f}\delta y_i + \frac{\partial^2 S}{\partial y_f^2}\delta y_f.$$

From comparison of these two expressions one obtains the expressions of the second derivatives of the action through monodromy matrix elements

$$\frac{\partial^2 S}{\partial y_i \partial y_f} = -\frac{1}{m_{12}}, \quad \frac{\partial^2 S}{\partial y_i^2} = \frac{m_{11}}{m_{12}}, \quad \frac{\partial^2 S}{\partial y_f^2} = \frac{m_{22}}{m_{12}}.$$

Substituting these expressions to the contribution to the trace formula from one periodic orbit one gets (in two dimensions)

$$d_p^{(osc)}(E) = \frac{1}{i(2\pi i \hbar)^{3/2}} \int |m_{12}|^{-1/2} \exp\left(\frac{i}{\hbar} S_p + i \frac{m_{11} + m_{22} - 2}{2\hbar m_{12}} y^2\right) dy \frac{dx}{k(x)}$$

where x and y are respectively coordinates parallel and perpendicular to the trajectory.

Computing the resulting integrals one obtains

$$d_p^{(osc)}(E) = \frac{T_p}{\pi \hbar} \frac{e^{iS_p/\hbar - i\pi\mu_p/2}}{\sqrt{|m_{11} + m_{22} - 2|}}$$

where $T_p = \int dx/k(x)$ is the geometrical period of the trajectory.

Finally the Gutzwiller trace formula takes the form (valid in arbitrary dimensions)

$$d^{(osc)}(E) = \sum_{\substack{\text{primitive} \\ \text{periodic} \\ \text{orbits}}} \frac{T_p}{\pi \hbar} \sum_{n=1}^{\infty} \frac{1}{|\det(M_p^n - 1)|^{1/2}} \cos \left[n \left(\frac{S_p}{\hbar} - \frac{\pi}{2} \mu_p \right) \right].$$

In the derivation of this formula we assumed that all periodic orbits are unstable and M_p is the monodromy matrix for a primitive periodic orbit.

5 Riemann Zeta Function

The trace-like formulas exist not only for dynamical systems but also for the Riemann zeta function (and others number-theoretical zeta functions as well).

The Riemann zeta function is a function of complex variable s defined as follows

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_p (1 - p^{-s})^{-1} \quad (36)$$

where the product is taken over prime numbers. The second equality (called the Euler product) is a consequence of the unique factorization of integers into a product of prime numbers.

This function converges only when $\text{Re } s > 1$ but can analytically be continued in the whole complex s -plane.

5.1 Functional Equation

The possibility of this continuation is connected with the fact that the Riemann zeta function satisfies the important functional equation

$$\zeta(s) = \varphi(s)\zeta(1-s) \quad (37)$$

where

$$\varphi(s) = 2^s \pi^{s-1} \sin\left(\frac{\pi s}{2}\right) \Gamma(1-s). \quad (38)$$

We present one of numerous method of proving this relation (see e.g. [55]).

When $\text{Re } s > 0$ one has the equality

$$\int_0^{\infty} x^{s/2-1} e^{-\pi n^2 x} dx = \frac{\Gamma(s/2)}{n^s \pi^{s/2}}$$

where $\Gamma(x)$ is the Gamma function (see e.g. [32], Vol. 1, Sect. 1). Therefore if $\text{Re } s > 1$

$$\frac{\Gamma(s/2)\zeta(s)}{\pi^{s/2}} = \int_0^{\infty} x^{s/2-1} \Psi(x) dx$$

where $\Psi(x)$ is given by the following series

$$\Psi(x) = \sum_{n=1}^{\infty} e^{-\pi n^2 x}.$$

Using the Poisson summation formula (5) one obtains

$$\sum_{n=-\infty}^{\infty} e^{-\pi n^2 x} = \frac{1}{\sqrt{x}} \sum_{n=-\infty}^{\infty} e^{-\pi n^2/x}$$

which leads to the identity

$$2\Psi(x) + 1 = \frac{1}{\sqrt{x}} \left(2\Psi\left(\frac{1}{x}\right) + 1 \right) .$$

Hence

$$\begin{aligned} \xi(s) &\equiv \pi^{-s/2} \Gamma\left(\frac{1}{2}s\right) \zeta(s) = \int_0^1 x^{s/2} \Psi(x) dx + \int_1^\infty x^{s/2} \Psi(x) dx = \\ &= \int_0^1 x^{s/2} \left(\frac{1}{\sqrt{x}} \Psi\left(\frac{1}{x}\right) + \frac{1}{2\sqrt{x}} - \frac{1}{2} \right) dx + \int_1^\infty x^{s/2} \Psi(x) dx = \\ &= \frac{1}{s-1} - \frac{1}{s} + \int_0^1 x^{s/2-3/2} \Psi\left(\frac{1}{x}\right) dx + \int_1^\infty x^{s/2} \Psi(x) dx = \\ &= \frac{1}{s(s-1)} + \int_1^\infty \left(x^{-s/2-1/2} + x^{s/2-1} \right) \Psi(x) dx . \end{aligned}$$

The last integral is convergent for all values of s and gives the analytical continuation of the Riemann zeta function to the whole complex s -plane, the only singularity being the pole at $s = 1$ with unit residue

$$\zeta(s) \xrightarrow{s \rightarrow 1} \frac{1}{s-1} .$$

(The pole at $s = 0$ is canceled by the pole of $\Gamma(s/2)$ giving $\zeta(0) = -1/2$.)

One of important consequences of the above formula of analytical continuation is that it does not change under the substitution $s \rightarrow 1 - s$. Therefore for all values of s

$$\xi(s) = \xi(1 - s)$$

or

$$\zeta(s) = \varphi(s) \zeta(1 - s)$$

where

$$\varphi(s) = \pi^{s-1/2} \frac{\Gamma(1/2 - s/2)}{\Gamma(s/2)} \tag{39}$$

By standard formulas (see e.g. [32], Vol. 1, 1.2.5, 1.2.15)

$$\Gamma(x) \Gamma(1 - x) = \frac{\pi}{\sin \pi x} , \quad \Gamma(2x) = 2^{2x-1} \pi^{-1/2} \Gamma(x) \Gamma\left(x + \frac{1}{2}\right)$$

the last expression can be transformed to (38) which proves the functional equation (37).

From the functional equation (37) it follows that $\zeta(s)$ has 'trivial' zeros at negative even integers (except zero) $s = -2, -4, \dots$ which appear from $\sin(\pi s/2)$ in (38). All other non-trivial zeros, $\zeta(s_n) = 0$, are situated in the so-called critical strip $0 < \text{Re } s < 1$. If one denotes these zeros as $s_n = 1/2 + i\gamma_n$ then functional equation together with the fact that $\zeta(s)^* = \zeta(s^*)$ state that in general there exist 4 sets of zeros: $\gamma_n, -\gamma_n, \gamma_n^*, -\gamma_n^*$.

According to the famous *Riemann conjecture* (see e.g. [55]) all nontrivial zeros of $\zeta(s)$ lie at the symmetry line $\text{Re } s = 1/2$ or γ_n are all real quantities.

Numerical calculations confirms this conjecture for exceptionally large number of zeros (see e.g. [47] and the web site of Odlyzko [48]) but a mathematical proof is still absent.

5.2 Trace Formula for the Riemann Zeros

Let us fix a test function $h(r)$ exactly as it was done for the Selberg trace formula in Sect. 2.8 i.e.

- $h(r)$ is a function analytical in the region $|\text{Im } r| \leq 1/2 + \delta$,
- $h(-r) = h(r)$,
- $|h(r)| \leq A(1 + |r|)^{-2-\delta}$.

Denote as in that Section

$$g(u) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} h(r) e^{-iru} dr$$

and define

$$H(s) = \int_{-\infty}^{+\infty} g(u) e^{(s-1/2)u} du .$$

Now let us compute the integral

$$\frac{1}{2\pi i} \oint ds H(s) \frac{\zeta'(s)}{\zeta(s)}$$

where the contour of integration is taken over the rectangle $-\eta \leq \text{Re } s \leq 1+\eta$ and $-T \leq \text{Im } s \leq T$ with $0 < \eta < \delta$ and $T \rightarrow +\infty$. Inside this rectangle there are poles of $\zeta'(s)/\zeta(s)$ coming from non-trivial zeros of the Riemann zeta function, $s_n = 1/2 + i\gamma_n$, and the one from the pole of $\zeta(s)$ at $s = 1$. The total contribution from these poles is

$$\sum_n h(\gamma_n) - h\left(-\frac{i}{2}\right) .$$

One can check that the limit $T \rightarrow \infty$ exists and, consequently, one has the identity

$$\sum_n h(\gamma_n) - h\left(-\frac{i}{2}\right) = \frac{1}{2\pi i} \int_{1+\eta-i\infty}^{1+\eta+i\infty} ds H(s) \frac{\zeta'(s)}{\zeta(s)} - \frac{1}{2\pi i} \int_{-\eta-i\infty}^{-\eta+i\infty} ds H(s) \frac{\zeta'(s)}{\zeta(s)} .$$

Let us substitute in the second integral the functional equation (37) with $\varphi(s)$ from (39). One has

$$\frac{\zeta'(s)}{\zeta(s)} = \ln \pi - \frac{\zeta'(1-s)}{\zeta(1-s)} - \frac{1}{2} \left[\frac{\Gamma'}{\Gamma} \left(\frac{s}{2} \right) + \frac{\Gamma'}{\Gamma} \left(\frac{1-s}{2} \right) \right] .$$

Now all integrals converge and one can move the integration contour till $s = 1/2 + ir$ with real r . In this manner one obtains

$$\begin{aligned} & \frac{1}{4\pi i} \int_{-\eta-i\infty}^{-\eta+i\infty} ds H(s) \left[\frac{\Gamma'}{\Gamma} \left(\frac{s}{2} \right) + \frac{\Gamma'}{\Gamma} \left(\frac{1-s}{2} \right) \right] = \\ & = h\left(\frac{i}{2}\right) + \frac{1}{2\pi} \int_{-\infty}^{+\infty} h(r) \frac{\Gamma'}{\Gamma} \left(\frac{1}{4} + \frac{i}{2}r \right) dr . \end{aligned}$$

The first term in the right hand side of this equality is due to the appearance of the pole of $\Gamma(s/2)$ at $s = 0$ when the integration contour shifted till $s = 1/2 + ir$. Also we have used that $h(-r) = h(r)$.

For terms with the Riemann zeta function one can use the expansion which follows from (36)

$$\frac{\zeta'(s)}{\zeta(s)} = - \sum_p \ln p \sum_{n=1}^{\infty} p^{-ns} .$$

Shifting the integration contour as above (i.e. till $s = 1/2 + ir$), using that $g(-u) = g(u)$, and combining all terms together one gets the following Weil explicit formula for the Riemann zeros

$$\begin{aligned} \sum_{\substack{\text{non-trivial} \\ \text{zeros}}} h(\gamma_n) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} h(r) \frac{\Gamma'}{\Gamma} \left(\frac{1}{4} + \frac{i}{2}r \right) dr + h\left(\frac{i}{2}\right) + h\left(-\frac{i}{2}\right) - \\ & - g(0) \ln \pi - 2 \sum_{\text{primes}} \ln p \sum_{n=1}^{\infty} \frac{1}{p^{n/2}} g(n \ln p) . \end{aligned}$$

Here γ_n are related with non-trivial zeros of the Riemann zeta function, s_n , as follows

$$s_n = \frac{1}{2} + i\gamma_n .$$

This formula is an analog of usual trace formulas as it relates zeros of the Riemann zeta function defined in a quite complicated manner with prime numbers which are a common notion.

The similarity with dynamical trace formulas is more striking if one assumes the validity of the Riemann conjecture which states that γ_n are real quantities (which in a certain sense can be considered as energy levels of a quantum system). In 'semiclassical' limit $r \rightarrow \infty$ using the Stirling formula (see e.g. [32], Vol. 1, 1.9.4)

$$\ln \Gamma(z) \xrightarrow{|z| \rightarrow \infty} \left(z - \frac{1}{2}\right) \ln z - z + \frac{1}{2} \ln 2\pi$$

one obtains that the density of Riemann zeros

$$d(E) = \sum_n \delta(E - \gamma_n)$$

can be expressed by the following ‘physical’ trace formula valid at large E

$$d(E) = \bar{d}(E) + d^{(osc)}(E)$$

where

$$\bar{d}(E) = \frac{1}{2\pi} \ln \frac{E}{2\pi} + \text{corrections},$$

and

$$d^{(osc)}(E) = -\frac{1}{\pi} \sum_p \sum_{n=1}^{\infty} \frac{\ln p}{p^{n/2}} \cos(En \ln p)$$

where the summation is performed over all prime numbers.

5.3 Chaotic Systems and the Riemann Zeta Function

By comparing the above equations with the trace formulas of chaotic systems one observes (see e.g. [38], [12], [13]) a remarkable correspondence between different quantities in these trace formulas

- periodic orbits of chaotic systems \leftrightarrow primes,
- periodic orbit period $T_p \leftrightarrow \ln p$,
- convergence properties of both formulas are also quite similar.

The number of periodic orbits with period less than T for chaotic systems is asymptotically

$$N(T_p < T) = \frac{e^{hT}}{hT},$$

where a constant h is called the topological entropy.

The number of prime numbers less than x is given by the prime number theorem (see e.g. [55])

$$N(p < x) = \frac{x}{\ln x}.$$

As $\ln p \equiv T_p$ this expression has the form similar to number of periodic orbits of chaotic systems with $h = 1$

$$N(T_p < T) = \frac{e^T}{T}.$$

Due to these similarities number-theoretical zeta functions play the role of a simple (but by far non-trivial) model of quantum chaos.

Notice that the overall signs of the oscillating part of trace formulas for the Riemann zeta function and dynamical systems are different. According to Connes [30] it may be interpreted as Riemann zeros belong not to a spectrum of a certain self-adjoint operator but to an ‘absorption’ spectrum. Roughly speaking it means the following. Let us assume that the spectrum of a ‘Riemann Hamiltonian’ is continuous and it covers the whole axis. But exactly when eigenvalues equal Riemann zeros corresponding eigenfunctions of this

Hamiltonian vanish. Therefore these eigenvalues do not belong to the spectrum and Riemann zeros correspond to such missing points similarly to black lines (forming absorption spectra) which are visible when light passes through an absorption media. In Connes' approach the 'Riemann Hamiltonian' may be very simple (see also [15]) but the functional space where it has to be defined is extremely intricate.

6 Summary

Trace formulas can be constructed for all 'reasonable' systems. They express the quantum density of states (and other quantity as well) as a sum over classical periodic orbits. All quantities which enter trace formulas can be computed within pure classical mechanics.

Trace formulas consist of two terms

$$d(E) = \bar{d}(E) + d^{(osc)}(E).$$

The smooth part of the density, $\bar{d}(E)$, for all dynamical systems is given by the Thomas-Fermi formula (plus corrections if necessary)

$$\bar{d}(E) = \int \frac{d\mathbf{p}d\mathbf{x}}{(2\pi\hbar)^f} \delta(E - H(\mathbf{p}, \mathbf{x})) .$$

For integrable systems the oscillating part of the density, $d^{(osc)}(E)$, is

$$d^{(osc)}(E) = \sum_{\mathbf{N}} \left(\frac{2\pi}{\hbar T_p} \right)^{(f-3)/2} \frac{1}{\hbar^2 \sqrt{|(N_i Q_{ij} N_j)|}} \exp \left(i \frac{S_p}{\hbar} - i \frac{\pi}{4} \mathbf{N} \boldsymbol{\mu} + i \frac{\pi}{4} \beta \right)$$

where $S_p = 2\pi \mathbf{N} \mathbf{I}$ is the action over a classical periodic orbit with fixed winding numbers \mathbf{N} and Q_{ij} is the co-matrix of the matrix of the second derivatives of the Hamiltonian.

For chaotic systems $d^{(osc)}(E)$ is represented as a sum over all classical periodic orbits

$$d^{(osc)}(E) = \sum_{\text{p.p.o.}} \frac{T_p}{\pi \hbar} \sum_{n=1}^{\infty} \frac{1}{|\det(M_p^n - 1)|^{1/2}} \cos \left(n \frac{S_p}{\hbar} - n \frac{\pi}{2} \mu_p \right)$$

where S_p is the classical action along a primitive periodic trajectory and M_p is its monodromy matrix.

Usually trace formulas represent the dominant contribution when $\hbar \rightarrow 0$. They are exact only in very special cases as for constant negative curvature surfaces generated by discrete groups where they coincide with the Selberg trace formula. For a group with only hyperbolic elements

$$\bar{d}(E) = \frac{\mu(D)}{4\pi} \tanh \pi k$$

where $\mu(D)$ is the area of the fundamental domain of the group and

$$d^{(osc)}(E) = \sum_{\text{p.p.o.}} \frac{l_p}{4\pi k} \sum_{n=1}^{\infty} \frac{\cos(knl_p)}{\sinh(nl_p/2)}$$

where l_p are lengths of periodic orbits.

The formulas similar to trace formulas exist also for number-theoretical zeta functions (assuming the generalized Riemann conjecture). In particular, for the Riemann zeta function

$$\bar{d}(E) = \frac{1}{2\pi} \ln \frac{E}{2\pi}$$

and

$$d^{(osc)}(E) = -\frac{1}{\pi} \sum_{\text{prime } n=1}^{\infty} \frac{\ln p}{p^{n/2}} \cos(En \ln p).$$

The principal difficulty of all trace formulas is the divergence of the sums over periodic orbits. To obtain a mathematically meaningful formula one considers instead of the singular density of states its smoothed version defined as a sum over all eigenvalues of a suitable chosen smooth test-function. When its Fourier harmonics decrease quickly the resulting formula represent a well defined object.

Suggestions for Further Readings

- A very detailed account of trace formulas derived by multiple scattering method can be found in a series of papers by Balian and Bloch [8].
- A short but concise mathematical review of hyperbolic geometry is given in [42].
- Explicit forms of the Selberg trace formula for general discrete groups with elliptic and parabolic elements are presented in two volumes of Hejhal's monumental work [39] which contains practically all known information about the Selberg trace formula.
- In [38] one can find a mathematical discussion about different relations between number-theoretical zeta functions and dynamical systems.

Statistical Distribution of Quantum Eigenvalues

Wigner and Dyson in the fifties had proposed to describe complicated (and mostly unknown) Hamiltonian of heavy nuclei by a member of an ensemble of random matrices and they argued that the type of this ensemble depends only on the symmetry of the Hamiltonian. For systems without time-reversal invariance the relevant ensemble is the Gaussian Unitary Ensemble (GUE), for systems invariant with respect to time-reversal the ensemble is the Gaussian Orthogonal Ensemble (GOE) and for systems with time-reversal invariance but with half-integer spin energy levels have to be described according to the Gaussian Symplectic Ensemble (GSE) of random matrices.

For these classical ensembles all correlation functions which determines statistical properties of eigenvalues E_n can be written explicitly (see e.g. [46], [16]). The simplest of them is the one-point correlation function or the mean level density, $\bar{d}(E)$, which is the probability density of finding a level in the interval $(E, E + dE)$. When $\bar{d}(E)$ is known one can construct a new sequence of levels, e_n , called unfolded spectrum as follows

$$e_n = \int^{E_n} \bar{d}(E) dE .$$

This artificially constructed sequence has automatically unit local mean density which signifies that the mean level density (provided it is a smooth function of E) plays a minor role in describing statistical properties of a spectrum at small intervals.

The two-point correlation function, $R_2(\epsilon)$, is the probability density of finding two levels separated by a distance in the interval $(\epsilon, \epsilon + d\epsilon)$. The characteristic properties of the above ensembles is the phenomenon of level repulsion which manifest itself in the vanishing of the two-point correlation function at small values of argument

$$R_2(\epsilon) \xrightarrow{\epsilon \rightarrow 0} \epsilon^\beta$$

where the parameter $\beta = 1, 2$, and 4 for, respectively, GOE, GUE, and GSE. This behaviour is in contrast with the case of the Poisson statistics of inde-

pendent random variables where

$$R_2(\epsilon) \xrightarrow{\epsilon \rightarrow 0} \bar{d}(E) \neq 0 .$$

For later use we present the explicit form of the two-point correlation function for GUE with mean density \bar{d}

$$\tilde{R}_2(\epsilon) = \bar{d}^2 + \bar{d}\delta(\epsilon) + \bar{R}_2(\epsilon) + R_2^{(osc)}(\epsilon) \quad (40)$$

where the smooth part of the connected two-point correlation function is given by

$$\bar{R}_2(\epsilon) = -\frac{1}{2\pi^2\epsilon^2} \quad (41)$$

and its oscillating part is

$$R_2^{(osc)}(\epsilon) = \frac{e^{2\pi i\bar{d}\epsilon} + e^{-2\pi i\bar{d}\epsilon}}{4\pi^2\epsilon^2} . \quad (42)$$

The term $\bar{d}\delta(\epsilon)$ in (40) corresponds to taking into account two identical levels and it is universal for all systems without spectral degeneracy. It is a matter of convenience to include it to $R_2(\epsilon)$ or not. When one adopts the definition (45) the appearance of such terms is inevitable.

Another useful quantity is the two-point correlation form factor defined as the Fourier transform of the two-point correlation function (unfolded to the unit density)

$$K(t) = \int_{-\infty}^{\infty} R_2(x) e^{2\pi i t x} dx . \quad (43)$$

For convenience one introduces a factor 2π in the definition of time.

In Fig. 8 the two-point correlation form factors for usual random matrix ensembles are presented. Their explicit formulas can be found in [46], [16]. For these classical ensembles small- t behaviour of the form factors is

$$K(t) \xrightarrow{t \rightarrow 0} \frac{2}{\beta} t \quad (44)$$

with the same β as above.

The nearest-neighbor distribution, $p(s)$, is defined as the probability density of finding two levels separated by distance s but, contrary to the two-point correlation function, no levels inside this interval are allowed. For classical ensembles the nearest-neighbor distributions can be expressed through solutions of certain integral equations and numerically they are close to the Wigner surmise (see e.g. [16])

$$p(s) = a s^\beta e^{-b s^2}$$

where β is the same as above and constants a and b are determined from normalization conditions

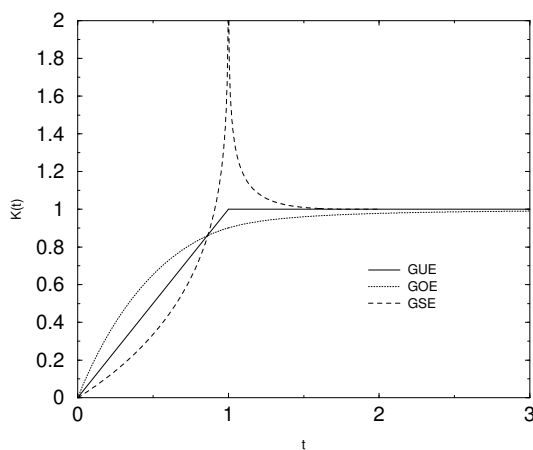


Fig. 8. Two point correlation form factor of classical random matrix ensembles.

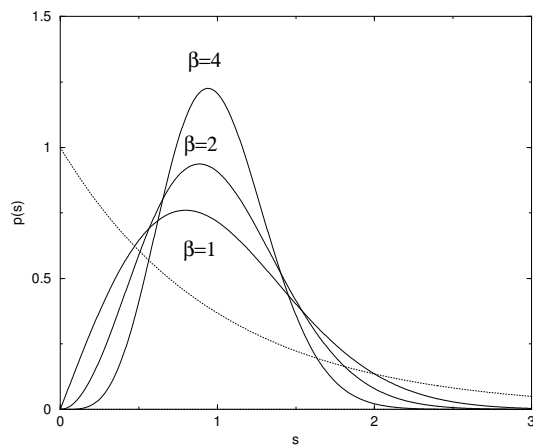


Fig. 9. Nearest-neighbor distribution for the standard random matrix ensembles. Dotted line – the Poisson prediction

$$\int_0^\infty p(s)ds = \int_0^\infty sp(s)ds = 1 .$$

These functions are presented at Fig. 9 together with the Poisson prediction for this quantity $p(s) = e^{-s}$.

Though random matrix ensembles were first introduced to describe spectral statistics of heavy nuclei later it was understood that the same conjectures can be applied also for simple dynamical systems and to-day standard accepted conjectures are the following

- The energy levels of classically integrable systems on the scale of the mean level density behave as independent random variables and their distribution is close to the Poisson distribution [10].
- The energy levels of classically chaotic systems are not independent but on the scale of the mean level density they are distributed as eigenvalues of random matrix ensembles depending only on symmetry properties of the system considered [17].
 - For systems without time-reversal invariance the distribution of energy levels should be close to the distribution of the Gaussian Unitary Ensemble (GUE) characterized by quadratic level repulsion.
 - For systems with time-reversal invariance the corresponding distribution should be close to that of the Gaussian Orthogonal Ensemble (GOE) with linear level repulsion.
 - For systems with time-reversal invariance but with half-integer spin energy levels should be described according to the Gaussian Symplectic Ensemble (GSE) of random matrices with quartic level repulsion.

These conjectures are well confirmed by numerical calculations.

The purpose of this Chapter is to investigate methods which permit to obtain spectral statistics analytically. For a large part of the Section we follow [25]. In Sect. 1 a formal expression is obtained which relates correlation functions with products of trace formulas. In Sect. 1.1 the simplest approximation to compute such products is discussed. It is called the diagonal approximation and it consists of taking into account only terms with exactly the same actions. Unfortunately, for chaotic systems this approximation can be used, strictly speaking, only for very small time estimated in Sect. 1.2. To understand the behaviour of the correlation functions for longer time more complicated methods of calculation of non-diagonal terms have to be developed. In Sect. 2 this goal is achieved for the Riemann zeta function. To obtain the information about correlations of prime pairs we use the Hardy–Littlewood conjecture which is reviewed in Sect. 2.1. The explicit form of the two-point correlation function for the Riemann zeros is obtained in Sec. 2.2. In Sect. 3 it is demonstrated that the obtained expression very well agrees with numerical calculations of spectral statistics for Riemann zeros.

1 Correlation Functions

Formally n -point correlation functions of energy levels are defined as the probability density of having n energy levels at given positions. Because the density of states, $d(E)$, is the probability density of finding one level at point E , correlation functions are connected to the density of states as follows

$$R_n(\epsilon_1, \epsilon_2, \dots, \epsilon_n) = \langle d(E + \epsilon_1)d(E + \epsilon_2) \dots d(E + \epsilon_n) \rangle . \quad (45)$$

The brackets $\langle \dots \rangle$ denote a smoothing over an appropriate energy window

$$\langle f(E) \rangle = \int f(E') \sigma(E - E') dE'$$

with a certain function $\sigma(E)$. Such smoothing means that one considers eigenvalues of quantum dynamical systems at different intervals of energy as forming a statistical ensemble.

The function $\sigma(E)$ is assumed to fulfill the normalization condition

$$\int \sigma(E) dE = 1$$

and to be centered around zero with a width ΔE obeying inequalities

$$\Delta E_q \ll \Delta E \ll \Delta E_{cl} \ll E. \quad (46)$$

Here ΔE_q has to be of the order of the mean level spacing, $\Delta E_q \approx 1/\bar{d}$, and ΔE_{cl} denotes the energy scale at which classical dynamics changes noticeably. A schematic picture of $\sigma(E)$ is represented at Fig. 10.

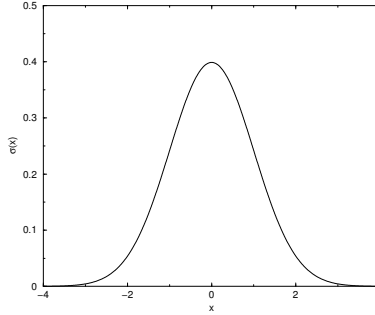


Fig. 10. Schematic form of smoothing function.

The trace formula for the density of states of chaotic systems was discussed in Chap. 1 and it has the form

$$d(E) = \bar{d}(E) + \sum_{p,n} A_{p,n} e^{inS_p(E)/\hbar} + \text{c.c.}$$

where the summation is performed over all primitive periodic orbits and its repetitions, and

$$A_{p,n} = \frac{T_p}{2\pi\hbar |\det(M_p^n - 1)|^{1/2}} e^{-\pi i n \mu_p / 2}. \quad (47)$$

Substituting this expression in the formula for the two-point correlation function one gets

$$R_2(\epsilon_1, \epsilon_2) = \bar{d}^2 + \sum_{p_i, n_i} A_{p_1, n_1} A_{p_2, n_2}^* \left\langle \exp \frac{i}{\hbar} (n_1 S_{p_1}(E + \epsilon_1) - n_2 S_{p_2}(E + \epsilon_2)) \right\rangle + \text{c.c.}$$

and the terms with the sum of actions are assumed to be washed out by the smoothing procedure.

Expanding the actions and taking into account that $\partial S(E)/\partial E = T(E)$ where $T(E)$ is the classical period of motion one finds

$$R_2^{(c)}(\epsilon_1, \epsilon_2) = \sum_{p_i, n_i} A_{p_1, n_1} A_{p_2, n_2}^* \left\langle \exp \frac{i}{\hbar} (n_1 S_{p_1}(E) - n_2 S_{p_2}(E)) \right\rangle \times \exp \frac{i}{\hbar} (n_1 T_{p_1}(E) \epsilon_1 - n_2 T_{p_2}(E) \epsilon_2) + \text{c.c.} .$$

Here $R_2^{(c)}(\epsilon_1, \epsilon_2)$ is the connected part of the two-point correlation function $R_2(\epsilon_1, \epsilon_2) = \bar{d}^2 + R_2^{(c)}(\epsilon_1, \epsilon_2)$.

The most difficult part is the computation of the mean value of terms with the difference of actions

$$\left\langle \exp \frac{i}{\hbar} (n_1 S_{p_1}(E) - n_2 S_{p_2}(E)) \right\rangle .$$

1.1 Diagonal Approximation

Berry [11] proposed to estimate such sums in an approximation (called the diagonal approximation) by taking into account only terms with *exactly* the same actions having in mind that terms with different values of actions will be small after the smoothing.

Let g be the mean multiplicity of periodic orbit actions. Then the connected part of the two-point correlation function in the diagonal approximation is

$$R_2^{(diag)}(\epsilon) = g \sum_{p, n \geq 1} |A_{p, n}|^2 e^{inT_p(E)\epsilon/\hbar} + \text{c.c.} . \quad (48)$$

Here $\epsilon = \epsilon_1 - \epsilon_2$ and the sum is taken over all primitive periodic orbits.

From (48) it follows that the two-point correlation form factor

$$K(t) = \int_{-\infty}^{+\infty} R_2(\epsilon) e^{2\pi i t \epsilon} d\epsilon .$$

in the diagonal approximation equals the following sum over classical periodic orbits

$$K^{(diag)}(t) = 2\pi g \sum_{p, n} |A_{p, n}|^2 \delta \left(2\pi t - \frac{nT_p(E)}{\hbar} \right) + \text{c.c.} . \quad (49)$$

According to the Hannay-Ozorio de Almeida sum rule [34] sums over periodic orbits of a chaotic systems can be calculated by using the local density of periodic orbits related with the monodromy matrix, M_p , as follows

$$d\rho_p = \frac{dT_p}{T_p} |\det(M_p - 1)| .$$

Using (47) one gets

$$K^{(diag)}(t) = \frac{g}{2\pi\hbar} \int T_p \delta(2\pi t - \frac{T_p}{\hbar}) dT_p = gt$$

where g is the mean multiplicity of periodic orbits (i.e. the mean proportion of periodic orbits with exactly the same action). For generic systems without time-reversal invariance there is no reasons for equality of actions for different periodic orbits and $g = 1$ but for systems with time-reversal invariance each orbit can be traversed in two directions therefore in general for such systems $g = 2$. Comparing these expressions one concludes that the diagonal approximation reproduces the correct small- t behavior of form-factors of classical ensembles (cf. (44)).

Unfortunately, $K^{(diag)}(t)$ grows with increasing of t but the exact form-factor for systems without spectral degeneracy should tends to \bar{d} for large t . This is a consequence of the following arguments. According to (45)

$$\begin{aligned} R_2(\epsilon) &= \left\langle \sum_{m,n} \delta(E - E_n) \delta(E + \epsilon - E_m) \right\rangle = \\ &= \left\langle \sum_n \delta(E - E_n) \delta(\epsilon - E_m + E_n) \right\rangle . \end{aligned}$$

If there is no levels with exactly the same energy the second δ -function in the right hand side of this equation tends to $\delta(\epsilon)$ when $\epsilon \rightarrow 0$ and the first one gives \bar{d} . Therefore

$$R_2(\epsilon) \rightarrow \bar{d}\delta(\epsilon) , \quad \text{when } \epsilon \rightarrow 0$$

which is equivalent to the following asymptotics of the form factor

$$K(t) \rightarrow \bar{d} , \quad \text{when } t \rightarrow \infty .$$

This evident contradiction clearly indicates that the diagonal approximation for chaotic systems cannot be correct for all values of t and more complicated tools are needed to obtain the full form factor.

1.2 Criterion of Applicability of Diagonal Approximation

One can give a (pessimistic) estimate till what time the diagonal approximation can be valid by the following method. The main ingredient of the diagonal

approximation is the assumption that after smoothing all off-diagonal terms give negligible contribution. This condition is almost the same as the condition of the absence of quantum interference. But it is known that the quantum interference is not important for times smaller than the Ehrenfest time which is of the order of

$$t_E \approx \frac{1}{\lambda_0} \ln(1/\hbar),$$

where λ_0 is a (classical) constant of the order of the Lyapunov exponent defined in such a way that the mean splitting of two nearby trajectories at time t grows as $\exp(\lambda_0 t)$. For billiards $(ka)^{-1}$, where a is of the order of system size, plays the role of \hbar and $\lambda_0 = k\lambda$ where k is the momentum and λ determines the deviation of two trajectories with length $L = kt$. The constant λ which we also called the Lyapunov exponent is independent on k for billiards and

$$t_E \approx \frac{1}{\lambda k} \ln(ka).$$

In the semiclassical limit $k \rightarrow \infty$ the Ehrenfest time and, consequently, the time during which one can use the diagonal approximation tends to zero as $\ln k/k$.

More careful argumentation can be done as follows. The off-diagonal terms can be neglected if

$$\left| \left\langle \exp \frac{i}{\hbar} (S_{p_1}(E) - S_{p_2}(E)) \right\rangle \right| \ll 1.$$

But this quantity is small provided the difference of periods of two orbits $\Delta T = T_{p_1} - T_{p_2}$ times the energy window ΔE used in the definition of smoothing procedure is large

$$\frac{1}{\hbar} (T_{p_1} - T_{p_2}) \Delta E \gg 1. \quad (50)$$

For billiards $T_p = L_p/k$ and this condition means that one has to consider all periodic orbits such that their difference of lengths is

$$L_{p_1} - L_{p_2} \gg \frac{\hbar k}{\Delta E}.$$

But the number of periodic orbits with the length L for chaotic systems grows exponentially

$$N(L_p < L) = \frac{e^{\lambda L}}{\lambda L}$$

where λ is a constant of the order of the Lyapunov exponent. Therefore in the interval $L, L + \delta l$ there is $e^{\lambda L} \delta l / L$ orbits and the mean difference of lengths between orbits with the lengths less than L is of the order of

$$\Delta L = L \exp(-\lambda L).$$

To fulfilled the above condition one has to restrict the maximum length of periodic orbits, L_m , by

$$L_m \exp(-\lambda L_m) \approx \frac{k\hbar}{\Delta E}.$$

In the limit of large L_m with logarithmic accuracy this relation gives

$$L_m \approx \frac{1}{\lambda} \ln \frac{\Delta E}{k\hbar\lambda} \quad (51)$$

which corresponds to the maximal time till the diagonal approximation can be applied

$$t_m = \frac{L_m}{k} \sim \frac{1}{\lambda k} \ln \frac{\Delta E}{\lambda k}.$$

As $\Delta E \ll E = k^2$, $t_m < t_E$.

Another important time scale for bounded quantum systems is called the Heisenberg time, t_H . It is the time during which one can see the discreteness of the spectrum

$$t_H = 2\pi\bar{d}.$$

As for billiards \bar{d} is a constant

$$t_E \ll t_H.$$

For the Riemann zeta function the situation is better because (i) in this case 'momentum' plays the role of 'energy' (the 'action' $E \ln p$ is linear in E and not quadratic as for dynamical systems) and (ii) the density of states for the Riemann zeta function is $(\ln(E/2\pi))/(2\pi)$.

The analog of (50) in this case is

$$(\ln p_1 - \ln p_2)\Delta E \gg 1.$$

It means that to apply the diagonal approximation prime numbers have to be such that the difference between any two of them obeys

$$\frac{\delta p}{p} \Delta E \gg 1.$$

The difference between primes near p is of the order of $\ln p$. Hence from the above inequalities it follows that diagonal approximation can be used till time $t_m = \ln p_m$ where p_m is such that

$$\frac{\ln p_m}{p_m} \geq \frac{1}{\Delta E}.$$

Or with logarithmic precision $p_m \leq \Delta E$. As $\Delta E \leq E$ (see (46)), $p_m \sim E$ and the maximum time

$$t_m \sim \ln E = 2\pi\bar{d}(E)$$

i.e. the diagonal approximation for the Riemann zeta function is valid till the Heisenberg time which agrees with the Montgomery theorem [45].

This type of estimates clearly indicates that the diagonal approximation for chaotic dynamical systems can not, strictly speaking, be used to obtain an information about the form-factor for large value of t . Only the short-time behaviour of correlation functions can be calculated by this method. (Notice that for GUE systems the diagonal approximation gives the expected answer till the Heisenberg time but it just signifies that one has to find special reasons why all other terms cancel.)

2 Beyond the Diagonal Approximation

The simplest and the most natural way of semi-classical computation of the two-point correlation functions is to find a method of calculating off-diagonal terms. We shall discuss here this type of computation on the example of the Riemann zeta function where much more information than for dynamical systems is available (for the latter see [21] and [25]).

The trace formula for the Riemann zeta function may be rewritten in the form

$$d^{(osc)}(E) = -\frac{1}{\pi} \sum_{n=1}^{\infty} \frac{1}{\sqrt{n}} \Lambda(n) \cos(E \ln n)$$

where

$$\Lambda(n) = \begin{cases} \ln p, & \text{if } n = p^k \\ 0, & \text{otherwise} \end{cases} .$$

The connected two-point correlation function of the Riemann zeros, $R_2^{(c)} = R_2 - \bar{d}^2$, is

$$R_2^{(c)}(\epsilon_1, \epsilon_2) = \frac{1}{4\pi^2} \sum_{n_1, n_2} \frac{\Lambda(n_1)\Lambda(n_2)}{\sqrt{n_1 n_2}} \left\langle e^{i(E+\epsilon_1) \ln n_1 - i(E+\epsilon_2) \ln n_2} \right\rangle + \text{c.c.} .$$

The diagonal approximation corresponds to taking into account terms with $n_1 = n_2$

$$\begin{aligned} R_2^{(diag)}(\epsilon_1, \epsilon_2) &= \frac{1}{4\pi^2} \sum_n \frac{\Lambda^2(n)}{n} e^{i(\epsilon_1 - \epsilon_2) \ln n} + \text{c.c.} = \\ &= \frac{1}{4\pi^2} \sum_{p, m} \frac{\ln^2 p}{p^m} e^{i(\epsilon_1 - \epsilon_2) m \ln p} + \text{c.c.} . \end{aligned}$$

This expression may be transform as follows (cf. [2])

$$R_2^{(diag)}(\epsilon) = -\frac{1}{4\pi^2} \frac{\partial^2}{\partial \epsilon^2} \ln \Delta(\epsilon)$$

where

$$\Delta(\epsilon) = |\zeta(1 + i\epsilon)|^2 \Phi^{(diag)}(\epsilon),$$

and function $\Phi^{(diag)}(\epsilon)$ is given by a convergent sum over prime numbers

$$\Phi^{(diag)}(\epsilon) = \exp\left(2 \sum_p \sum_{m=1}^{\infty} \frac{1-m}{m^2 p^m} \cos(m\epsilon \ln p)\right).$$

In the limit $\epsilon \rightarrow 0$, $\zeta(1 + i\epsilon) \rightarrow (i\epsilon)^{-1}$ and $\Phi^{(diag)}(\epsilon) \rightarrow \text{const.}$ Therefore in this limit

$$R_2^{(diag)}(\epsilon) \rightarrow -\frac{1}{2\pi^2 \epsilon^2}$$

which agrees with the smooth part of the GUE result (41).

The off-diagonal contribution takes the form

$$R_2^{(off)}(\epsilon_1, \epsilon_2) = \sum_{n_1 \neq n_2} \frac{\Lambda(n_1)\Lambda(n_2)}{4\pi^2 \sqrt{n_1 n_2}} \left\langle e^{iE \ln(n_1/n_2) + i(\epsilon_1 \ln n_1 - \epsilon_2 \ln n_2)} \right\rangle + \text{c.c.} .$$

The term $\exp(iE \ln(n_1/n_2))$ oscillates quickly if n_1 is not close to n_2 . Denoting

$$n_1 = n_2 + r$$

and expanding all smooth functions on r one gets

$$R_2^{(off)}(\epsilon) = \frac{1}{4\pi^2} \sum_{n,r} \frac{\Lambda(n)\Lambda(n+r)}{n} \left\langle e^{iEr/n + i\epsilon \ln n} \right\rangle + \text{c.c.}$$

where $\epsilon = \epsilon_1 - \epsilon_2$.

The main problem is clearly seen here. The function

$$F(n, r) = \Lambda(n)\Lambda(n+r)$$

is quite a wild function as it is nonzero only when both n and $n+r$ are powers of prime numbers. As we have assumed that $r \ll n$, the dominant contribution to the two-point correlation function will come from the mean value of this function over all n , i.e. one has to substitute into $R_2^{(off)}(\epsilon)$ instead of $F(n, r)$ its mean value

$$\alpha(r) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \Lambda(n)\Lambda(n+r).$$

2.1 The Hardy-Littlewood Conjecture

Fortunately the explicit expression for this function comes from the famous Hardy–Littlewood conjecture. There are two different methods which permit to ‘find’ this conjecture. We start with the original Hardy-Littlewood derivation [35].

First, let us remind two known facts. The number of prime numbers less than a given number $N(p < x)$ is asymptotically (see e.g. [55])

$$N(p < x) = \frac{x}{\ln x} .$$

Conveniently it can also be expressed in the following form

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \Lambda(n) = 1 .$$

The number of prime number $N_{q,r}(p < x)$ in arithmetic progression of the form $mq + r$ with $(r, q) = 1$ and $r < q$ is given by the following asymptotic formula (see e.g. [31])

$$N_{q,r}(p < x) = \frac{x}{\varphi(q) \ln x}$$

where $\varphi(n)$ is the Euler function which counts integers less than n and co-prime with n

$$\varphi(n) = n \prod_{p|n} \left(1 - \frac{1}{p}\right) .$$

As above, this relation can be rewritten in the equivalent form

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{m=1}^N \Lambda(mq + r) = \frac{1}{\varphi(q)} . \quad (52)$$

In the Hardy-Littlewood method [35] one introduces the function

$$f(x) = \sum_{n=1}^{\infty} \Lambda(n) x^n$$

which converges for all complex x such that $|x| < 1$.

In the circle method of Hardy and Littlewood [35] one considers the behaviour of this function close to the unit circle when the phase of x is near a rational number $2\pi p/q$ with co-prime integers p and q . One gets

$$f(e^{-u} e^{2\pi i p/q + i\delta}) = \sum_{n=1}^{\infty} \Lambda(n) e^{-nu} e^{2\pi i n p/q + i n \delta}$$

with $u, \delta \rightarrow 0$.

In the exponent there is a quickly changing function $2\pi n p/q$. It is quite natural to consider n from the arithmetic progression

$$n = mq + r$$

with fixed q and $r < q$. In this case

$$f(e^{-u}e^{2\pi ip/q+i\delta}) = \sum_{m, r} \Lambda(mq + r)e^{-(mq+r)(u-i\delta)}e^{2\pi irp/q}.$$

Substituting instead of $\Lambda(mq + r)$ its mean value (52) one gets

$$f(e^{-u}e^{i2\pi p/q+i\delta}) \approx \frac{1}{\varphi(q)} \sum_{(r, q)=1} e^{2\pi irp/q} \int_0^\infty e^{-n(u-i\delta)}dn = \frac{\mu(q)}{\varphi(q)(u-i\delta)}.$$

In the last step we use that fact that [36]

$$\sum_{(r, q)=1} e^{2\pi ir/q} = \mu(q)$$

where $\mu(q)$ is the Möbius function defined through the factorization of q on prime factors

$$\mu(q) = \begin{cases} 1 & \text{if } q = 1 \\ (-1)^k & \text{if } q = p_1 \dots p_k \\ 0 & \text{if } q \text{ is divisible on } p^2 \end{cases}.$$

The final expression means that function $f(x)$ has a pole singularity at the unit circle at every rational point.

The knowledge of $f(x)$ permits formally to compute the mean value of the product of two Λ -functions.

Let

$$J_r(R) = \frac{1}{2\pi} \int_0^{2\pi} f(Re^{i\varphi})f(Re^{-i\varphi})e^{-ir\varphi}d\varphi = R^r \sum_m \Lambda(m+r)\Lambda(m)R^{2m}.$$

As the function $f(x)$ has a pole singularity at the unit circle at every rational point one can try to approximate this integral by the sum over singularities

$$\begin{aligned} J_r(e^{-u}) &= \frac{1}{2\pi} \int_0^{2\pi} f(Re^{i\varphi})f(Re^{-i\varphi})e^{-ir\varphi}d\varphi = \\ &= \frac{1}{2\pi} \sum_{(p, q)=1} \int f(e^{-u+i2\pi p/q+i\delta})f(e^{-u-2\pi ip/q-i\delta})e^{-ir(2\pi p/q+i\delta)}d\delta = \\ &= \frac{1}{2\pi} \sum_{(p, q)=1} e^{2\pi irp/q} \left(\frac{\mu(q)}{\varphi(q)}\right)^2 \int \frac{d\delta}{u^2 + \delta^2} = \\ &= \frac{1}{2u} \sum_{(p, q)=1} e^{2\pi ipr/q} \left(\frac{\mu(q)}{\varphi(q)}\right)^2. \end{aligned}$$

Therefore

$$\sum_{n=1}^\infty \Lambda(n)\Lambda(n+r)e^{-2nu} \xrightarrow{u \rightarrow 0} \frac{1}{2u} \sum_{(p, q)=1} e^{2\pi irp/q} \left(\frac{\mu(q)}{\varphi(q)}\right)^2$$

from which it follows that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \Lambda(n) \Lambda(n+r) = \alpha(r)$$

where

$$\alpha(r) = \sum_{q=1}^{\infty} \left(\frac{\mu(q)}{\varphi(q)} \right)^2 \sum_{(p, q)=1} e^{2\pi i r p / q} . \quad (53)$$

Using properties of such singular series one can prove [35] that for even r $\alpha(r) = 0$ and for odd r it can be represented as the following product over prime numbers

$$\alpha(r) = C_2 \prod_{p|r} \frac{p-1}{p-2} \quad (54)$$

where the product is taken over all prime divisors of r bigger than 2 and C_2 is the so-called twin prime constant

$$C_2 = 2 \prod_{p>2} \left(1 - \frac{1}{(p-1)^2} \right) \approx 1.32032 \dots \quad (55)$$

Instead of demonstration the formal equivalence of (53) and (54) we present another heuristic 'derivation' based on the probabilistic interpretation of prime numbers which gives directly (54) and (55).

The argumentation consists on the following steps.

- Probability that a given number is divisible by a prime p is

$$\lim_{N \rightarrow \infty} \frac{1}{N} [\text{number of integers divisible by } p \leq N] = \frac{1}{p} .$$

In general to find such probabilities it is necessary to consider only the residues modulo p and find how many of them obey the requirement.

- Probability that a given number is not divisible by a prime p is

$$1 - \frac{1}{p} .$$

- Probability that a number is not divisible by primes p_1, p_2, \dots, p_k is

$$\prod_{j=1}^k \left(1 - \frac{1}{p_j} \right) . \quad (56)$$

The above formula is correct for any finite collection of primes but for computations with infinite number of primes it may be wrong.

For example, when used naively it gives that

- probability that a number x is a prime is

$$\prod_{p < \sqrt{x}} \left(1 - \frac{1}{p}\right).$$

This prime number 'theorem' is false because from it it follows that the number of primes less than x is [55]

$$\Pi(x) = x \prod_{p < \sqrt{x}} \left(1 - \frac{1}{p}\right) \xrightarrow{x \rightarrow \infty} \frac{x}{\ln x} 2e^{-\gamma}$$

which differs from the true prime number theorem by a factor $2e^{-\gamma} \approx 1.123$ where γ is the Euler constant. The origin of this discrepancy is related with the approximation frequently used above: $[x/p] = x/p$ where $[x]$ is the integer part of x . Instead of (56) one should have $\prod_p (1 - [x/p]/x)$. For a finite number of primes and $x \rightarrow \infty$ it tends to (56). But when the number of primes considered increases with x errors are accumulated giving a constant factor.

Nevertheless one could try to use probabilistic arguments by forming artificially convergent quantities. One has

- Probability that x and $x + r$ are primes is

$$\lim_{N \rightarrow \infty} \frac{1}{N} [\text{number of integers } x < N \text{ such that } x \text{ and } x + r \text{ are primes}].$$

Let consider a prime p . Two cases are possible. Either $p|r$ or $p \nmid r$. In the first case the probability that both number x and $x + r$ are not divisible by p is the same as the probability that only number x is not divisible by p which is

$$\prod_{p|r} \left(1 - \frac{1}{p}\right).$$

When $p \nmid r$ one has to remove two numbers from the set of residues as $x = 0, 1, \dots, p-1 \pmod{p}$ and $x+r = 0, 1, \dots, p-1 \pmod{p}$. Therefore the probability that both numbers x and $x + r$ are not divisible by a prime p is

$$\prod_{p \nmid r} \left(1 - \frac{2}{p}\right).$$

Finally

- Probability that both x and $x + r$ are primes = $\prod_{p|r} \frac{p-1}{p} \prod_{p \nmid r} \frac{p-2}{p}$.

To find a convergent expression we divide both sides by the probability that numbers x and $x + r$ are independently prime numbers computed also in the probabilistic approximation. The latter quantity is

$$[\text{Probability that } x \text{ is prime and } x \leq N] = \prod_p \frac{p-1}{p}.$$

Therefore

$$\frac{[\text{Probability that both } x \text{ and } x+r \text{ are primes with } x, x+r \leq N]}{[\text{Probability that } x \text{ is prime}]^2} \approx \\ \approx \prod_{p|r} \frac{p-1}{p} \prod_{p \nmid r} \frac{p-2}{p} \prod_p \left(\frac{p}{p-1}\right)^2 = 2 \prod_{p>2} \left(1 - \frac{1}{(p-1)^2}\right) \prod_{p|r} \frac{p-1}{p-2}.$$

As the denominator in the above expression is $1/\ln^2 N$ it follows that the probability that both x and $x+r$ are primes with $x \leq N$, and $x+r \leq N$ is asymptotically

$$\frac{\alpha(r)}{\ln^2 N}$$

with the same function $\alpha(r)$ as in (54).

We stress that the Hardy–Littlewood conjecture is still not proved. Even the existence of infinite number of twin primes (primes separated by 2) is not yet proved while the Hardy–Littlewood conjecture states that their density is $C_2/\ln^2 N$.

2.2 Two-Point Correlation Function of Riemann Zeros

Taking the above expression of the Hardy–Littlewood conjecture as granted we get

$$R_2^{(off)}(\epsilon) = \frac{1}{4\pi^2} \sum_{n \geq 1} \frac{1}{n} e^{i\epsilon \ln n} \sum_r \alpha(r) e^{iEr/n} + \text{c.c.}.$$

After substitution the formula for $\alpha(r)$ and performing the sum over all r one obtains

$$R_2^{(off)}(\epsilon) = \frac{1}{4\pi^2} \sum_n \frac{1}{n} e^{i\epsilon \ln n} \sum_{(p,q)=1} \left(\frac{\mu(q)}{\varphi(q)}\right)^2 \delta\left(\frac{p}{q} - \frac{E}{2\pi n}\right) + \text{c.c.}$$

where the summation is taken over all pairs of mutually co-prime positive integers p and q (without the restriction $p < q$).

Changing the summation over n to the integration permits to transform this expression to contributions of values of n where

$$\frac{p}{q} - \frac{E}{2\pi n} = 0.$$

In this approximation

$$R_2^{(off)}(\epsilon) = \frac{1}{4\pi^2} e^{i\epsilon \ln E/2\pi} \sum_{(p,q)=1} \left(\frac{\mu(q)}{\varphi(q)}\right)^2 \left(\frac{q}{p}\right)^{1+i\epsilon} + \text{c.c.}.$$

Using the formula (which is a mathematical expression of the inclusion–exclusion principle)

$$\sum_{(p,q)=1} f(p) = \sum_{k=1}^{\infty} \sum_{\delta|q} f(k\delta)\mu(\delta)$$

and taking into account that $2\pi\bar{d} = \ln(E/2\pi)$ one obtains

$$R_2^{(off)}(\epsilon) = \frac{1}{4\pi^2} |\zeta(1+i\epsilon)|^2 e^{2\pi i \bar{d} \epsilon} \Phi^{(off)}(\epsilon) + c.c. \tag{57}$$

where function $\Phi^{(off)}(\epsilon)$ is given by a convergent product over primes

$$\Phi^{(off)}(\epsilon) = \prod_p \left(1 - \frac{(1-p^{i\epsilon})^2}{(p-1)^2} \right)$$

and $\Phi^{(off)}(0) = 1$.

In the limit of small ϵ

$$R_2^{(off)}(\epsilon) = \frac{1}{(2\pi\epsilon)^2} \left(e^{2\pi i \bar{d} \epsilon} + e^{-2\pi i \bar{d} \epsilon} \right)$$

which exactly corresponds to the GUE results for the oscillating part of the two-point correlation function (42).

The above calculations demonstrate how one can compute the two-point correlation function through the knowledge of correlation function of periodic orbit pairs. For the Riemann case one can prove under the same conjectures that all n -point correlation functions of Riemann zeros tend to corresponding GUE results [22].

3 Summary

Trace formulas can formally be used to calculate spectral correlation functions for dynamical systems. In particular, the two-point correlation function is the product of two densities of states

$$R_2(\epsilon) \equiv \langle d(E+\epsilon)d(E) \rangle .$$

The diagonal approximation consists of taking into account in such products only terms with exactly the same action. For chaotic systems this approximation is valid only for very small time. In particular, it permits to obtain the short-time behaviour of correlation form factors which agrees with predictions of standard random matrix ensembles.

The main difficulty in such approach to spectral statistics is the necessity to compute contributions from non-diagonal terms which requires the knowledge of correlation functions of periodic orbits with nearby actions.

For the Riemann zeta function zeros it can be done using the Hardy–Littlewood conjecture which claims that the number of prime pairs p and $p+r$ such that $p < N$ for large N is asymptotically

$$\alpha(r) \frac{N}{\ln^2 N}$$

where $\alpha(r)$ (with even r) is given by the product over all odd prime divisors of r

$$\alpha(r) = C_2 \prod_{p|r} \frac{p-1}{p-2}$$

and

$$C_2 = 2 \prod_{p>2} \left(1 - \frac{1}{(p-1)^2} \right).$$

Using this formula one gets that the two-point correlation function of Riemann zeros is

$$R_2(\epsilon) = \bar{d}^2(E) + R_2^{(diag)}(\epsilon) + R_2^{(off)}(\epsilon)$$

where the diagonal part

$$R_2^{(diag)}(\epsilon) = -\frac{1}{4\pi^2} \frac{\partial^2}{\partial \epsilon^2} \ln \left[|\zeta(1+i\epsilon)|^2 \Phi^{(diag)}(\epsilon) \right]$$

and non-diagonal part

$$R_2^{(off)}(\epsilon) = \frac{1}{4\pi^2} |\zeta(1+i\epsilon)|^2 e^{2\pi i \bar{d}\epsilon} \Phi^{(off)}(\epsilon) + \text{c.c.} .$$

The functions $\Phi^{(diag)}(\epsilon)$ and $\Phi^{(off)}(\epsilon)$ are given by convergent products over all primes

$$\Phi^{(diag)}(\epsilon) = \exp \left(2 \sum_p \sum_{m=1}^{\infty} \frac{1-m}{m^2 p^m} \cos(m\epsilon \ln p) \right)$$

and

$$\Phi^{(off)}(\epsilon) = \prod_p \left(1 - \frac{(1-p^{i\epsilon})^2}{(p-1)^2} \right).$$

In [25] a few other methods were developed to 'obtain' the two-point correlation function for Riemann zeros. These methods were based on different ideas and certain of them can be generalized for dynamical systems. Though neither of the methods can be considered as a strict mathematical proof, all lead to the same expression (57).

It is also of interest to check numerically the above formulas. When numerical calculations are performed one considers usually correlation functions for the unfolded spectrum. For the two-point correlation function this procedure corresponds to the following transformation

$$R_2^{(\text{unfolded})}(\epsilon) = \frac{1}{\bar{d}^2(E)} R_2 \left(\frac{\epsilon}{\bar{d}(E)} \right).$$

At Fig 11 we present the two-point correlation function for $2 \cdot 10^8$ zeros near the 10^{23} -th zero computed numerically by Odlyzko [49] together with the GUE prediction for this quantity

$$R_2^{\text{GUE}}(\epsilon) = 1 - \left(\frac{\sin \pi \epsilon}{\pi \epsilon} \right)^2 .$$

At Figs. 12-15 we present the difference between the two-point correlation

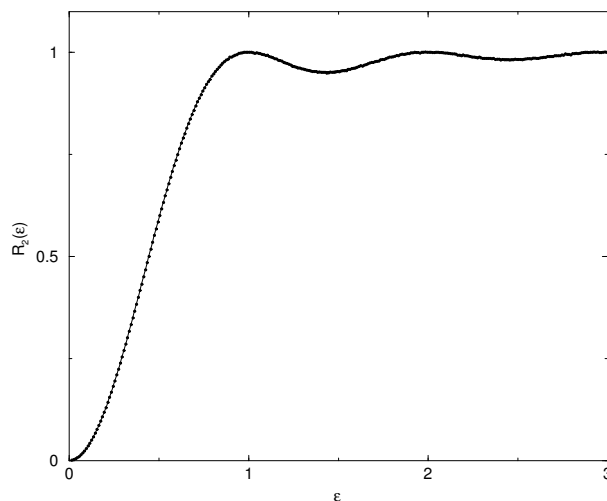


Fig. 11. Two point correlation function of the Riemann zeros near the 10^{23} -th zero (dots) and the GUE prediction (solid line).

function computed numerically and the GUE prediction. At Fig. 16 we present the difference between numerically computed two-point correlation function and the ‘exact’ function and at Fig. 17 the histogram of differences is given. Notice that these differences are structure less and the histogram corresponds practically exactly to statistical errors inherent in the calculation of the two-point correlation functions which signifies that the obtained formula agrees very well with the numerics.

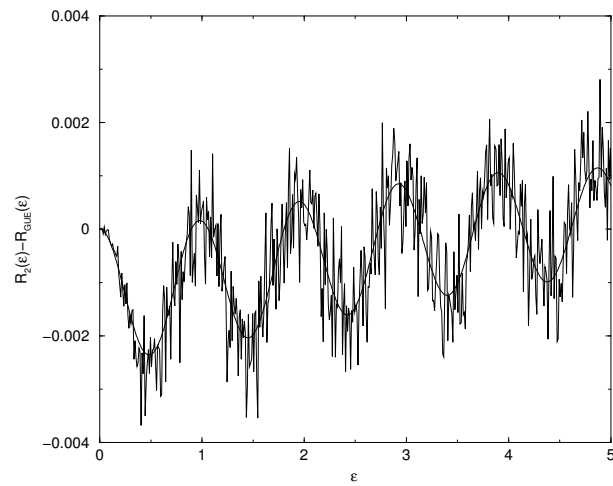


Fig. 12. The difference between the two point correlation function of the Riemann zeros and the GUE prediction in the interval $0 < \epsilon < 5$. The solid line is the difference between the 'exact' correlation function and the GUE prediction in this interval.

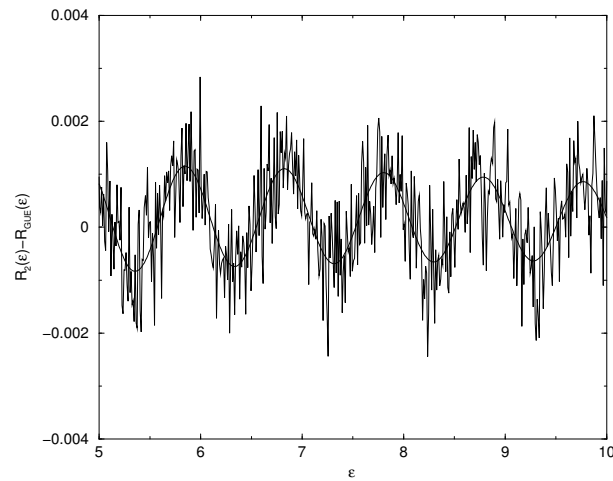


Fig. 13. The same as at Fig. 12 but in the interval $5 < \epsilon < 10$.

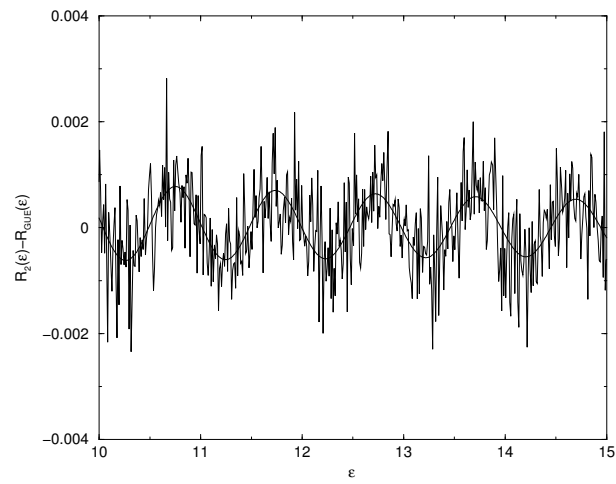


Fig. 14. The same as at Fig. 12 but in the interval $10 < \epsilon < 15$.

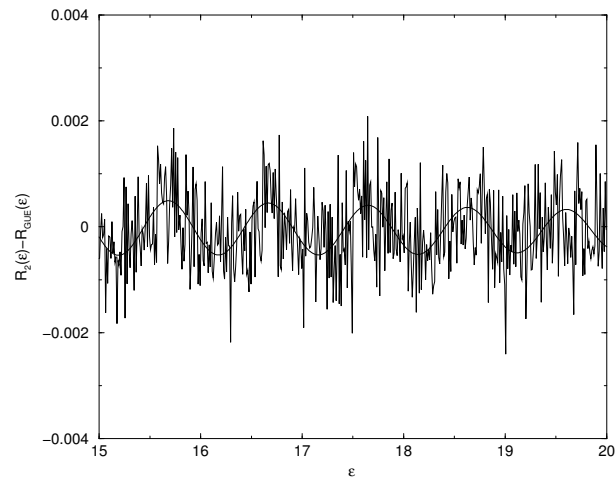


Fig. 15. The same as at Fig. 12 but in the interval $15 < \epsilon < 20$.

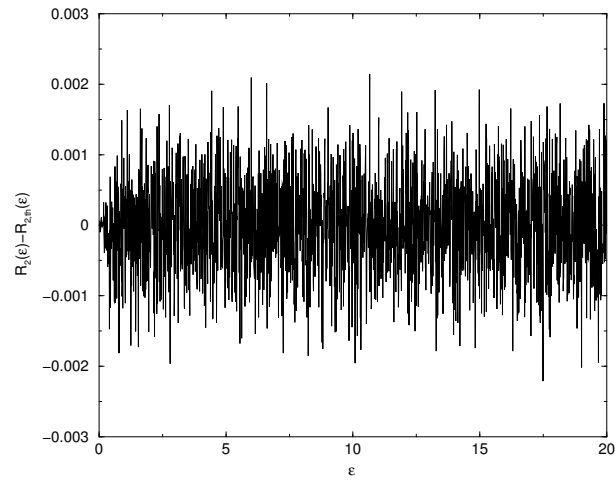


Fig. 16. The difference between numerically computed two-point correlation function and the 'exact' function

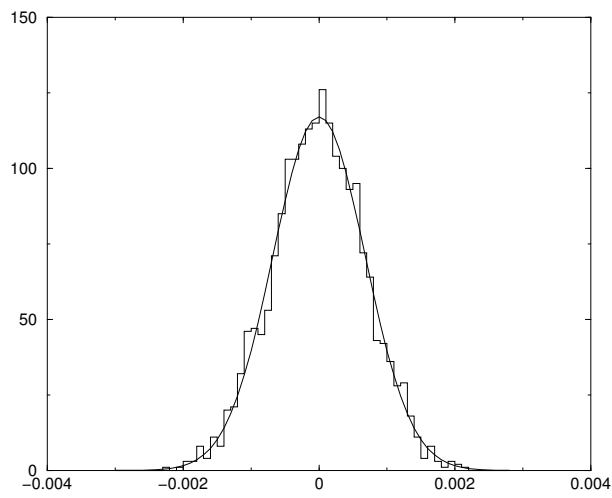


Fig. 17. The histogram of the deviations of the numerically computed two-point correlation function of Riemann zeros and the 'exact' formula. Solid line is the Gaussian fit to the histogram.

Arithmetic Systems

As was discussed above it is well accepted that spectral statistics of classically chaotic systems in the universal limit coincides with spectral statistics of the usual random matrix ensembles. But it is also known (see e.g. [7], [29] and references therein) that the motion on constant negative curvature surfaces generated by discrete groups (considered in Chap. 1) is the best example of classical chaos. Consequently, models on constant negative curvature seem to be ideal tools to check the conjecture on spectral fluctuations of classically chaotic systems. Their classical motion is extremely chaotic and time-reversal invariant and *a priori* assumption was that all of them should have energy levels distributions close to predictions of the Gaussian orthogonal ensemble (GOE) of random matrices.

Nevertheless when the first large scale numerical calculations were performed [3], [52] they clearly indicated that for certain hyperbolic models the spectral statistics were quite close to Poisson statistics typical for integrable systems.

As an example we present in Fig. 18 the nearest-neighbor distribution for the hyperbolic triangle with angles $(\pi/2, \pi/3, 0)$ corresponding to the well-known modular triangle with Dirichlet boundary conditions. The agreement with Poisson prediction is striking though classical motion for this system is perfectly chaotic.

The purpose of this Chapter is to show that this strange behaviour is the consequence of exponentially large exact degeneracy of periodic orbit lengths in systems considered [18]. In all hyperbolic surfaces where the Poisson-like statistics was observed there is on average $\exp(l/2)$ classical periodic orbits with exactly the same length l . It will be demonstrated that this is the characteristic property of models generated by the so-called arithmetic groups. As classical mechanics is not sensitive to lengths of periodic orbits all these models remain completely chaotic. But the cumulative effect of interference of degenerated periodic orbits changes drastically the quantum mechanical properties.

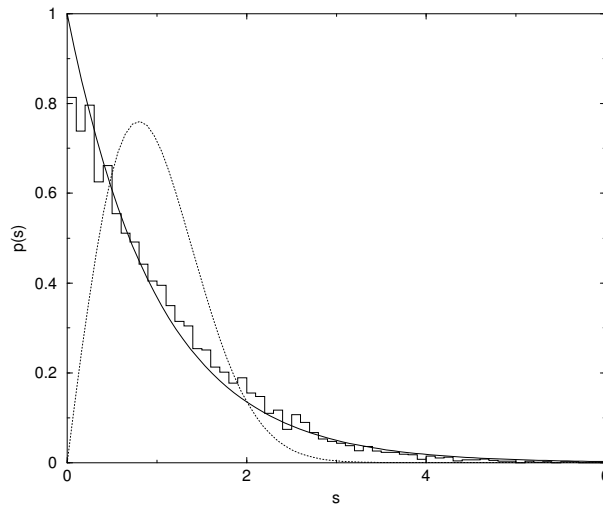


Fig. 18. The nearest neighbor distribution for 10000 first levels of the triangle $(\pi/2, \pi/3, 0)$ (the modular triangle). Solid line - the Poisson distribution. Dotted line - the GOE distribution.

This Chapter is based mostly on [24]. In Sect. 1 simple calculations prove exponential degeneracy of periodic orbit lengths for the modular group. The main peculiarity of the modular group matrices is that their traces are integers. Therefore if one considers all matrices with $|\text{Tr } M| < X$ the number of different traces increases at most linearly with X . In Sect. 2 it is shown that this property is typical for all arithmetic groups. An informal mini-review of such groups is given in this Section and it is demonstrated that for all these groups exponentially many periodic orbits have exactly the same length. From the results of [53] it follows that there is exactly 85 triangles generated by discrete arithmetic groups. All triangular models where the Poisson-like spectral statistics was numerically observed are in this list. In Sect. 3 it is shown that in the diagonal approximation the two-point correlation form factor of arithmetic systems jumps very quickly to the Poisson value thus confirming unusual nature of arithmetic systems. In Sect. 4 the exact two-point correlation function for the modular domain is calculated. The correlations of multiplicities are obtained by a generalization of the Hardy–Littlewood method discussed in Sect. 2.1. The resulting formula proves that in the universal limit the two-point correlation function of eigenvalues of the Laplace–Beltrami operator automorphic with respect to the modular group tends to the Poisson prediction. Arithmetic groups have many other interesting properties. In particular, for all arithmetic groups it is possible to construct an infinite number of mutually commuting operators which commute also with the Laplace–Beltrami operator. Properties of these operators called the Hecke operators are discussed in Sect. 5. The Jacquet–Langlands correspondence between different arithmetic

groups is mentioned in Sect. 6. In Sect. 7 non-arithmetic models are briefly discussed.

1 Modular group

The modular group is the group of all 2×2 matrices

$$M = \begin{pmatrix} m & n \\ k & l \end{pmatrix}$$

with integer m, n, k, l and the unit determinant $ml - nk = 1$.

The periodic orbits correspond in a unique way to the conjugacy classes of hyperbolic elements of the group (see Sect. 2.3). The length of periodic orbit l_p is related with the trace of a representative matrix of the conjugacy class M as follows

$$2 \cosh \frac{l_p}{2} = |\text{Tr } M| .$$

As all elements of modular group matrices are integers, the trace is also an integer

$$|\text{Tr}M| = n . \tag{58}$$

Here the arithmetical nature of the group clearly comes into the play. This simple property is very important. It signifies that for the modular group there is just a quite restrictive set of all possible traces and, consequently, of periodic orbit lengths. For modular group the number of possible different lengths is the number of different integers less than $2 \cosh L/2$ (see (58)), hence

$$N_{\text{dif. lengths}} = 2 \cosh \frac{L}{2} \xrightarrow{L \rightarrow \infty} e^{L/2} .$$

On the other hand, for any discrete group the number of periodic orbits of length less than L grows asymptotically as

$$N(l_p < L) = \frac{e^L}{L} .$$

Let $g(l)$ be the multiplicity of periodic orbits with length l . One has obvious relations valid for large L

$$\sum_{l < L} g(l) = \frac{e^L}{L} , \quad \sum_{l < L} 1 = e^{L/2}$$

where the summation extends over different lengths of periodic orbits counted without taking multiplicity into account.

Let us define the *mean* multiplicity $\langle g(l) \rangle$ as the following ratio

$$\langle g(l) \rangle = \frac{\text{Number of periodic orbits with } l < l_p < l + \Delta l}{\text{Number of different lengths with } l < l_p < l + \Delta l} . \tag{59}$$

Asymptotically for large L the previous formulas gives

$$\langle g(l) \rangle = 2 \frac{e^{l/2}}{l}$$

which demonstrates that periodic orbit lengths for the modular group are exponentially degenerated.

2 Arithmetic Groups

The crucial feature which led to the exponential degeneracy of periodic orbit lengths for the modular group was the fact that traces of modular group matrices were integers which was a direct consequence of the arithmetic nature of modular group. But 2×2 matrix groups with integer elements are exhausted by the modular group and its subgroups.

Nevertheless, one can construct a quite large class of discrete groups with strong arithmetic properties by considering groups which are not equal to 2×2 integer matrices but which permit a *representation* by $n \times n$ integer matrices ($n > 2$).

The existence of such representation means that for each 2×2 group matrix, g , one can associate a $n \times n$ matrix with integer entries, $M(g)$, in such a way that the matrix associated to the product of two groups matrices equals the product of two matrices associated to the corresponding factors

$$M(ab) = M(a) \times M(b)$$

for all a and b from the group considered and $M(\mathbf{1}) = \mathbf{1}$.

To define general arithmetic groups we need a few definitions.

- A subset of a group Γ is called a subgroup if it forms itself a group.
- A subgroup g of a group Γ is called a subgroup of finite index if Γ can be represented as a finite union

$$\Gamma = g + g\gamma_1 + \dots + g\gamma_k$$

with $\gamma_k \in \Gamma$.

- Two groups are called commensurable if they have a common subgroup which is of finite index in both of them.

Groups which have a representation by integer matrices and all groups commensurable with them are called arithmetic groups. This Section is devoted to the investigation of their properties.

In Sect. 2.1 a non-formal review of algebraic fields is given and in Sect. 2.2 the construction of quaternion algebras over algebraic fields is shortly discussed. It appears that all arithmetic groups can be obtained from quaternion algebras with division and in Sect. 2.3 the necessary and sufficient conditions

that a given group will be an arithmetic group is presented. Using these conditions in Sect. 2.4 it is proved that periodic orbit lengths for all arithmetic groups have the same exponential degeneracy (up to a constant factor) as for the modular group.

2.1 Algebraic Fields

Everybody is familiar with usual rational numbers

$$u = \frac{p}{q}$$

with integer p and q . Their important properties are (i) that the sum and the product of any two rational numbers also have the same form and (ii) all elements except 0 have an inverse (i.e. the division is always possible). From mathematical viewpoint rational numbers form a field called \mathbb{Q} .

Algebraic fields of finite degree, \mathbb{F} , are a generalization of this reference field obtaining by adding to the set of rational numbers a root α of an irreducible polynomial

$$\sum_{k=0}^n c_k \alpha^k = 0 \tag{60}$$

with integer coefficients c_k . This field is denoted $\mathbb{F} = \mathbb{Q}(\alpha)$.

Each element $u \in \mathbb{Q}(\alpha)$ can be represented by the sum

$$u = \sum_{i=0}^{n-1} b_i \alpha^i$$

where the b_i are usual rationals. The summation and the multiplication of these elements are done as with usual numbers except that all powers of α larger than $n - 1$ have to be reduced using the defining equation (60).

Integers of the field $\mathbb{Q}(\alpha)$ are its elements which obey a polynomial equation with integer coefficients with an additional condition that the highest power coefficients equals one (such polynomials are called monic polynomials).

In general, algebraic integers, ω , of a field of degree n are freely generated by n linearly independent elements of the field β_k with integer coefficients (in mathematical language it means that they form a free \mathbb{Z} -module of rank n). Explicitly

$$\omega = \sum_{k=0}^{n-1} m_k \beta_k \tag{61}$$

where all m_k are usual integers. In simple cases $\beta_k = \alpha^k$ and

$$\omega = \sum_{k=0}^{n-1} m_k \alpha^k$$

with integer or demi-integer coefficients m_k . Algebraic integers like usual integers form a ring (not a field) because the division is not always possible.

The polynomial equation defining the field (60) has n different roots $\alpha_i, i = 0, 1, \dots, n-1$ with $\alpha_0 = \alpha$. Any relation between elements of the field remains unchanged under the transformations

$$\phi_i : \alpha \rightarrow \alpha_i$$

where one substitutes in all expressions instead of one root α another root α_i . These transformations are called isomorphisms or embeddings of this field into \mathbb{C} and they are the only transformations respecting the laws of the field.

Example

Add to the field of rational numbers \mathbb{Q} one root of the quadratic equation

$$x^2 = d$$

where d is a square-free integer. Elements of this field $\mathbb{Q}(\sqrt{d})$ can be written as

$$u = p + q\sqrt{2}$$

with p and q rationals. Let

$$\omega = a + b\sqrt{2}$$

be integers of this field. To find values of a and b one notes that ω obeys the quadratic equation

$$\omega^2 - 2a\omega + a^2 - db^2 = 0.$$

To describe integers $2a$ and $a^2 - db^2$ have to be usual integers. Depending on d two types of solutions are possible.

- If $d \equiv 2$ or $d \equiv 3 \pmod{4}$ then a and b have to be integers and

$$\omega = m + n\sqrt{d}$$

with integers m and n .

- If $d \equiv 1 \pmod{4}$ then both a and b can be demi-integers and

$$\omega = \frac{m}{2} + \frac{n}{2}\sqrt{d}$$

with integers $m \equiv n \pmod{2}$.

To avoid the last restriction this expression can be rewritten in the form (61)

$$\omega = m + n \frac{1 + \sqrt{d}}{2} \tag{62}$$

with arbitrary integers m and n .

As this field is defined by an equation of the second degree it has two isomorphisms

$$\begin{aligned}\phi_0 : p + q\sqrt{d} &\longrightarrow p + q\sqrt{d}, \\ \phi_1 : p + q\sqrt{d} &\longrightarrow p - q\sqrt{d}.\end{aligned}$$

Because the product of two algebraic integers is also an algebraic integer from (61) it follows that all algebraic integers permit a representation by matrices with integer elements in such a way that the matrix representing a product of two integers equals the product of matrices representing each factor.

For example, for the above considered case of $\mathbb{Q}(\sqrt{d})$ with $d \not\equiv 1 \pmod{4}$ one can associate with an integer of this field, $\omega = m + n\sqrt{2}$, a 2×2 matrix

$$M(\omega) = \begin{pmatrix} m & n \\ dn & m \end{pmatrix}. \tag{63}$$

It is easy to check that this is the true representation of field integers because $M(\omega_1\omega_2) = M(\omega_1)M(\omega_2)$ and $M(1) = \mathbf{1}$.

When $d \equiv 1 \pmod{4}$ the integers have the form (62) and one can check that the matrix representation can be chosen as follows

$$M(\omega) = \begin{pmatrix} m & n \\ \frac{d-1}{4}n & m+n \end{pmatrix}.$$

2.2 Quaternion Algebras

Algebras are more general objects than fields. A (vector) algebra of finite dimension d is defined as formal sum

$$\gamma = x_1\mathbf{i}_2 + x_2\mathbf{i}_2 + \dots + x_d\mathbf{i}_d.$$

Here x_j belong to a basis field \mathbb{F} and \mathbf{i}_j are formal objects (vectors) with a prescribed multiplication table

$$\mathbf{i}_j\mathbf{i}_k = \sum_{p=1}^d C_{jk}^p\mathbf{i}_p$$

where C_{jk}^p are from the basis field. The sum and the product of any two elements of an algebra belong to it. General algebras should be neither commutative, nor associative.

An algebra is called a normed algebra if there exists a function, $N(\gamma)$, which associates to any element of the algebra a number from the basis field such that the norm of the product equals the product of the norms of both factors

$$N(\gamma_1\gamma_2) = N(\gamma_1)N(\gamma_2).$$

An algebra is called a division algebra if the division is always possible (except a zero element).

Finite dimensional normed division algebras over real numbers are exhausted by the following three possibilities (the Frobenius theorem [44]).

- Commutative and associative division algebras are isomorphed either to the usual field of real numbers \mathbb{R} or to the field of complex numbers \mathbb{C} . In the latter case the algebra is given by

$$\gamma = x_1 + x_2\mathbf{i}$$

with $\mathbf{i}^2 = -1$. The norm in this case is

$$N(\gamma) = x_1^2 + x_2^2 .$$

- Non-commutative but associative division algebras are isomorphed to the quaternion algebra

$$\gamma = x_1 + x_2\mathbf{i} + x_3\mathbf{j} + x_4\mathbf{k} \quad (64)$$

where

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -1 , \quad \mathbf{k} = \mathbf{ij} = -\mathbf{ji} . \quad (65)$$

The norm of the quaternion algebra is

$$N(\gamma) = x_1^2 + x_2^2 + x_3^2 + x_4^2 .$$

- Non-associative normed division algebras are isomorphed to the octonion algebra

$$\gamma = \sum_{k=1}^8 x_k \mathbf{i}_k$$

with a complicated multiplication table and the norm given by the sum of 8 squares

$$N(\gamma) = \sum_{k=1}^8 x_k^2 .$$

Similarly, for an algebraic field \mathbb{F} of finite degree there exist quaternion normed algebras defined similarly to Eqs. (64) and (65) [56]. These algebras are labeled by two elements $a, b \in \mathbb{F}$ and it is a four-dimensional non-commutative algebra with basis $(\mathbf{1}, \mathbf{i}, \mathbf{j}, \mathbf{k})$ as in (64) with the following multiplication table

$$\mathbf{i}^2 = a , \quad \mathbf{j}^2 = b , \quad \mathbf{k} = \mathbf{ij} = -\mathbf{ji} . \quad (66)$$

Such algebra is denoted by $\left(\frac{a,b}{\mathbb{F}}\right)$ and its norm is

$$N(\gamma) = x_1^2 - ax_2^2 - bx_3^2 + abx_4^2 . \quad (67)$$

The matrix representation of the quaternion algebra (66) is obtained by the isomorphism

$$\mathbf{i} \rightarrow \begin{pmatrix} \sqrt{a} & 0 \\ 0 & -\sqrt{a} \end{pmatrix}, \mathbf{j} \rightarrow \begin{pmatrix} 0 & 1 \\ b & 0 \end{pmatrix}, \mathbf{k} = \mathbf{ij} \rightarrow \begin{pmatrix} 0 & \sqrt{a} \\ -b\sqrt{a} & 0 \end{pmatrix}.$$

Explicitly

$$\gamma = \begin{pmatrix} x_1 + x_2\sqrt{a} & x_3 + x_4\sqrt{a} \\ b(x_3 - x_4\sqrt{a}) & x_1 - x_2\sqrt{a} \end{pmatrix} \tag{68}$$

with $x_1, x_2, x_3, x_4 \in \mathbb{F}$. As it is a representation of the quaternion algebra the product of these matrices also has the same form. In this representation the norm of the algebra (67) equals the determinant of the matrix (68)

$$N(\gamma) = \det \gamma.$$

From an algebraic field \mathbb{F} of finite degree one can build also another simple set of matrices called $M(2, \mathbb{F})$ given by 2×2 matrices with entries from \mathbb{F}

$$\begin{pmatrix} x_1 & x_2 \\ x_3 & x_4 \end{pmatrix}. \tag{69}$$

Are the two sets (68) and (69) different or are they isomorphic? For example, if $a = u^2$ and $u \in \mathbb{F}$ the set (68) is, evidently, within $M(2, \mathbb{F})$.

Let us show that if $\sqrt{a} \notin \mathbb{F}$ but there exist certain elements $q_1, q_2, q_3, q_4 \in \mathbb{F}$ such that the determinant of the matrix (68) is zero

$$\det(\gamma) = q_1^2 - q_2^2 a - b(q_3^2 - q_4^2 a) = 0 \tag{70}$$

then matrices (68) are isomorphic to $M(2, \mathbb{F})$ [42].

Indeed, from the above expression it follows that in this case b has the form

$$b = (q_1^2 - q_2^2 a)(q_3^2 - q_4^2 a)^{-1} = (u_1 + u_2\sqrt{a})(u_1 - u_2\sqrt{a})$$

where

$$u_1 + u_2\sqrt{a} = (q_1 + q_2\sqrt{a})(q_3 + q_4\sqrt{a})^{-1}.$$

As all fractions of field elements belong to the field \mathbb{F} , u_1 and u_2 are also elements of \mathbb{F} . Now one can check that

$$\begin{pmatrix} x_1 + x_2\sqrt{a} & x_3 + x_4\sqrt{a} \\ b(x_3 - x_4\sqrt{a}) & x_1 - x_2\sqrt{a} \end{pmatrix} = S^{-1}MS$$

where

$$M = \begin{pmatrix} x_1 + x_3u_1 + x_4u_2a & x_2 - x_3u_2 - x_4u_1 \\ a(x_2 + x_3u_2 + x_4u_1) & x_1 - x_3u_1 - x_4u_2a \end{pmatrix}$$

and S is a fixed (independent of x_i) matrix

$$S = \begin{pmatrix} u_1 + u_2\sqrt{a} & 1 \\ \sqrt{a}(u_1 + u_2\sqrt{a}) & -\sqrt{a} \end{pmatrix}.$$

The importance of such representation lies in the fact that the matrix M contains only elements of our basis field \mathbb{F} and does not contain \sqrt{a} . Therefore,

it belongs to $M(2, \mathbb{F})$ and the set of matrices γ (68) is the conjugation of matrices from $M(2, \mathbb{F})$ by a fixed matrix S . In other words, when the equation $\det(\gamma) = 0$ has a solution $\gamma \in \mathbb{F}$ expression (68) is just a complicated way of writing matrices from $M(2, \mathbb{F})$. These considerations demonstrate that in order to construct a group different from $M(2, \mathbb{F})$ it is necessary to require that there exist *no* elements from the basis field such that the determinant (70) equals zero. Or, equivalently, any matrix (68) should have an inverse element. In the language of quaternion algebra this property corresponds to the division algebra for which any element has an inverse.

As for real fields this condition is quite restrictive and an explicit answer can be obtained only in simple cases. Let us consider for example the case when \mathbb{F} is the field of usual rational numbers $\mathbb{F} = \mathbb{Q}$. The following theorem [42] gives a series of division algebras over \mathbb{Q} .

Let b be a prime number and a be an integer such that the equation

$$x^2 \equiv a \pmod{b}$$

has no integer solution. Then the pair (a, b) defines a division algebra over \mathbb{Q} or equivalently the equation (70)

$$x_1^2 - x_2^2 a - b(x_3^2 - x_4^2 a) = 0 \quad (71)$$

has only zero rational solutions.

To prove the theorem note that due to homogeneity of this equation it is sufficient to consider integer solutions x_1, x_2, x_3, x_4 without common factors. From (71) it follows that

$$x_1^2 \equiv x_2^2 a \pmod{b}.$$

Consider first the case when b does not divide x_2 , $b \nmid x_2$. As b is assumed to be a prime, $x_2^{-1} \pmod{b}$ exists and $(x_1/x_2)^2 \equiv a \pmod{b}$ which contradicts our assumption. Hence $b \mid x_2$ but then $b \mid x_1$ and

$$x_3^2 \equiv x_4^2 a \pmod{b}.$$

The same arguments give that $b \mid x_4$ and $b \mid x_3$ which contradicts the assumption about the absence of common factors of x_i . Therefore there is no rational solution of (71) and the quaternion algebra defined by a and b is a division algebra.

Quaternion algebras with division are analogs of algebraic fields. How one can define integers of a quaternion algebra?

We have seen above that algebraic integers of a field of degree n form a free \mathbb{Z} -module of rank n i.e. they can be represented as a sum of n elements of the field with integer coefficients (see (61)). Similarly one can define 'integers' of a quaternion algebra over such field as a free \mathbb{Z} -module of rank $4n$ (which generates the whole algebra). For technical reasons they are called not integers but 'the order' in the algebra. The word 'integers' in algebras is reserved for

elements for which the trace and the determinant of matrix (68) are integers of the basis field. Different orders exist and the one which is not contained in any other order is called the maximal order.

The simplest case appears when a and b are integers of the basis field \mathbb{F} . Then matrices of the form

$$\begin{pmatrix} x_1 + x_2\sqrt{a} & x_3 + x_4\sqrt{a} \\ b(x_3 - x_4\sqrt{a}) & x_1 - x_2\sqrt{a} \end{pmatrix}$$

where all x_k are integers of \mathbb{F} form an order of the algebra $\left(\frac{a,b}{\mathbb{F}}\right)$ (but not necessarily the maximal order).

Matrices of the order in a division quaternion algebra with unit determinant form a group. Each matrix of this group belongs to the order and, therefore, is defined by $4n$ integers. The product of two group matrices have the same form and corresponds to a certain transformation of integers defining both matrices. It means that these groups can be represented by $4n \times 4n$ matrices with integer elements.

All such groups, all their subgroups, and all groups commensurable with them are *discrete arithmetic groups* with finite fundamental domain [42].

Example

As $x^2 \pmod{5}$ takes only values 0, 1, 4 the equation

$$x^2 \equiv 3 \pmod{5}$$

has no integer solution. Hence the pair (3,5) defines a division algebra over \mathbb{Q} .

A simplest order of this algebra has the form

$$\begin{pmatrix} m + n\sqrt{3} & k + l\sqrt{3} \\ 5(k - l\sqrt{3}) & m - n\sqrt{3} \end{pmatrix} \tag{72}$$

with integer m, n, k, l . When one considers these matrices with the unit determinant

$$m^2 - 3n^2 - 5k^2 + 15l^2 = 1$$

they form a discrete arithmetic group Γ_1 with a finite fundamental area.

The order (72) is not the maximal order. The latter can be chosen e.g. as follows

$$\begin{pmatrix} \frac{1}{2}(m + n\sqrt{3}) & \frac{1}{2}(k + l\sqrt{3}) \\ \frac{5}{2}(k - l\sqrt{3}) & \frac{1}{2}(m - n\sqrt{3}) \end{pmatrix} \tag{73}$$

with integer m, n, k, l such that $m \equiv k \pmod{2}$ and $n \equiv l \pmod{2}$. Matrices (73) with the unit determinant

$$m^2 - 3n^2 - 5k^2 + 15l^2 = 4$$

constitute another discrete arithmetic group Γ_2 whose fundamental domain is smaller than for the group (72) as Γ_1 is a subgroup of Γ_2 .

Using the representation (63) one concludes that to each 2×2 matrix of the group Γ_1 one can associate the 4×4 matrix with integer elements

$$M(\gamma) = \begin{pmatrix} m & n & k & l \\ 2n & m & 2l & k \\ 5k & -5l & m & -n \\ -10k & 5k & -2n & m \end{pmatrix}.$$

It is straightforward to check that (i) $M(\gamma_1\gamma_2) = M(\gamma_1)M(\gamma_2)$ for all $\gamma_1, \gamma_2 \in \Gamma_1$, (ii) $M(\mathbf{1}) = \mathbf{1}$, and (iii) $\det(M) = (\det(\gamma))^2 = 1$. Together these expressions mean that this group is an arithmetic group.

2.3 Criterion of arithmeticity

For general fields the situation is more complicated. To explain the general criterion of arithmetic groups let us first stress a difference between usual integers and algebraic integers.

The usual integers correspond to a discrete set of points. But for general algebraic integers this is not the case. For example, in the field $\mathbb{Q}(\sqrt{2})$ integers have the form $n + m\sqrt{2}$ with integer n and m . But it is evident that one can construct sequences of these algebraic integers converging to zero, e.g. $(\sqrt{2} - 1)^k = M_k - N_k\sqrt{2} \rightarrow 0$ when $k \rightarrow \infty$. Therefore the set of algebraic integers is not discrete as it has finite accumulation points.

How can one deal with such problem? The main point is that these small terms become large under the transformation

$$\sqrt{2} \rightarrow -\sqrt{2} \quad (74)$$

Let consider in the above example not all algebraic integers $n + m\sqrt{2}$ but only those which after transformation (74) remain bounded

$$|n - m\sqrt{2}| < \text{constant}.$$

It is clear that now arbitrary small integers are excluded and one gets a discrete set of points.

For more general fields the transformation (74) is generalized to all non-trivial isomorphisms of the field. To remove arbitrary small elements one has to require that for all isomorphisms of the field (except the identity), ϕ_i , transformed values of integers (61) are restricted

$$\left| \sum_{k=0}^{n-1} m_k \phi_i(\beta_k) \right| < \text{constant}. \quad (75)$$

In order to be sure that all small numbers are removed it is necessary that all roots of defining equation (60) are real. Otherwise, changing a root to its complex conjugate may not change modulus of integers.

These considerations make reasonable that in order to construct a *discrete* subset of algebraic integers (without finite accumulation points) it is necessary that (i) the field be a totally real field (i.e. all roots of defining equation (60) are real) and (ii) for all non-trivial isomorphisms of the field transformed integers remain bounded as in (75).

A precise criterion of arithmeticity obtained by Takeuchi [53] is quite similar (see also [6] and [24] for particular examples).

Takeuchi proved that a group Γ is an arithmetic group if and only if the traces of group matrices have the following properties

- All $\text{Tr}(\gamma)$ are integers of a totally real algebraic field of finite degree.
- For any non-trivial isomorphism ϕ of this field which changes some $|\text{Tr}(\gamma)|$ for certain γ , the value of the transformed trace satisfies $|\phi(\text{Tr}(\gamma))| \leq 2$.

There are two types of arithmetic groups. Non-compact groups, built from $SL(2, \mathbb{Z})$, and compact ones built from quaternion algebra different from $M(2, \mathbb{Q})$.

The above criterion is quite effective, in particular, it permits to find all possible arithmetic groups with triangular fundamental domains [53]. There are 85 triangular hyperbolic surfaces generated by discrete arithmetic groups. All of them are given in Table 1.

2.4 Multiplicities of Periodic Orbits for General Arithmetic Groups

The geometrical length of the periodic orbit, l , is connected with the trace of class of conjugate matrices by (28). When $l \rightarrow \infty$

$$\exp \frac{l}{2} = |\text{Tr}(\gamma)| .$$

Let us prove that for an arithmetic group the number of possible values of group matrix traces obeys

$$N(|\text{Tr}(\gamma)| \leq R) \xrightarrow{R \rightarrow \infty} C \cdot R$$

with a constant C depending on the group. The traces of matrices of arithmetic groups are dispatched as usual integers among real numbers.

Let Γ be an arithmetic group. The set of traces $\{\text{Tr}(\gamma), \gamma \in \Gamma\}$ are integers of an algebraic field \mathbb{F}

$$t_0 = \sum_{i=0}^{n-1} m_i \beta_i$$

where m_i are integers and β_i are linearly independent elements of the field. Consider the simplest case $\beta_i = \alpha^i$ then

$$\text{Tr}(\gamma) \equiv t_0 = \sum_{i=0}^{n-1} m_i \alpha^i .$$

Table 1. The list of arithmetic triangles from [53]. (n, m, p) in the first column corresponds to the three angles $(\pi/n, \pi/m, \pi/p)$. The second column indicates the algebraic field from which is built the corresponding arithmetic group.

(m,n,p)	IF
(2,3, ∞) (2,4, ∞) (2,6, ∞) (2, ∞ , ∞) (3,3, ∞)	\mathbb{Q}
(3, ∞ , ∞) (4,4, ∞) (6,6, ∞) (∞ , ∞ , ∞)	
(2,4,6) (2,6,6) (3,4,4) (3,6,6)	\mathbb{Q}
(2,3,8) (2,4,8) (2,6,8) (2,8,8) (3,3,4)	$\mathbb{Q}(\sqrt{2})$
(3,8,8) (4,4,4) (4,6,6) (4,8,8)	
(2,3,12) (2,6,12) (3,3,6) (3,4,12) (3,12,12)	$\mathbb{Q}(\sqrt{3})$
(6,6,6)	
(2,4,12) (2,12,12) (4,4,6) (6,12,12)	$\mathbb{Q}(\sqrt{3})$
(2,4,5) (2,4,10) (2,5,5) (2,10,10) (4,4,5)	$\mathbb{Q}(\sqrt{5})$
(5,10,10)	
(2,5,6) (3,5,5)	$\mathbb{Q}(\sqrt{5})$
(2,3,10) (2,5,10) (3,3,5) (5,5,5)	$\mathbb{Q}(\sqrt{5})$
(3,4,6)	$\mathbb{Q}(\sqrt{6})$
(2,3,7) (2,3,14) (2,4,7) (2,7,7) (2,7,14)	$\mathbb{Q}(\cos \pi/7)$
(3,3,7) (7,7,7)	
(2,3,9) (2,3,18) (2,9,18) (3,3,9) (3,6,18)	$\mathbb{Q}(\cos \pi/9)$
(9,9,9)	
(2,4,18) (2,18,18) (4,4,9) (9,18,18)	$\mathbb{Q}(\cos \pi/9)$
(2,3,16) (2,8,16) (3,3,8) (4,16,16) (8,8,8)	$\mathbb{Q}(\cos \pi/8)$
(2,5,20) (5,5,10)	$\mathbb{Q}(\cos \pi/10)$
(2,3,24) (2,12,24) (3,3,12) (3,8,24) (6,24,24)	$\mathbb{Q}(\cos \pi/12)$
(12,12,12)	
(2,5,30) (5,5,15)	$\mathbb{Q}(\cos \pi/15)$
(2,3,30) (2,15,30) (3,3,15) (3,10,30) (15,15,15)	$\mathbb{Q}(\cos \pi/15)$
(2,5,8) (4,5,5)	$\mathbb{Q}(\sqrt{2}, \sqrt{5})$
(2,3,11)	$\mathbb{Q}(\cos \pi/11)$

For a field of degree n there exist $n - 1$ non-trivial isomorphisms $\phi_k : \alpha \rightarrow \alpha_k$ where α_k is a root of the defining polynomial different from α .

Suppose that all such transformations change $|\text{Tr}(\gamma)|$. According to the criterion of Takeuchi all transformed traces satisfy

$$|t_k| \leq 2$$

where

$$t_k \equiv \phi_k(\text{Tr}(\gamma)) = \sum_{i=0}^{n-1} m_i \alpha_k^i.$$

Consider these equations as transformations from variables t_i to new variables m_i [27]. The volume elements in these two representations are related as

$$dt_0 dt_1 \dots dt_{n-1} = |\mathcal{J}| dm_0 dm_1 \dots dm_{n-1}$$

where

$$\mathcal{J} = \det \left(\frac{\partial t_j}{\partial m_k} \right)$$

is called the discriminant of the field and in our case (when $\beta_i = \alpha^i$)

$$\mathcal{J} = \det(\alpha_k^j) |_{k,j=0,\dots,n-1} = \prod_{i \neq j} (\alpha_i - \alpha_j).$$

As m_i are integers the volume of the smallest cell is one, and the total number of possible integer solutions is asymptotically

$$N(|\text{Tr}(\gamma)| \leq R) = N(|t_0| \leq R, |t_j| \leq 2) \simeq C \cdot R$$

where $C = 2^n / \mathcal{J}$.

For any surface of finite area generated by a discrete group the total number of periodic orbits with length less than a given value is asymptotically the following

$$N_{tot}(l < L) \xrightarrow{L \rightarrow \infty} \frac{e^L}{L}.$$

The number of periodic orbits with *different* lengths is the same as the number of group matrix traces

$$N_{\text{diff. lengths}}(l < L) \sim C \cdot e^{L/2}$$

Let $g(l)$ be the multiplicity of periodic orbits with length l . Then

$$\sum_{l < L} g(l) = \frac{e^L}{L} \text{ and } \sum_{l < L} 1 = C e^{L/2}$$

where the summation is done over different lengths.

Finally the *mean* multiplicity of arithmetic systems defined as in (59) has the following asymptotics

$$\langle g \rangle = \frac{(\sum_{l < L} g(l))'}{(\sum_{l < L} 1)'} \sim \frac{2e^{L/2}}{CL}. \quad (76)$$

Thus we demonstrate that the arithmetic nature of arithmetic groups leads to exponential multiplicities of periodic orbit lengths.

For generic systems one usually does not expect any degeneracy of periodic orbit lengths except the ones which follow from exact symmetries of the model. For example, systems with time-reversal invariance, in general, should have the mean multiplicity equal to 2, which corresponds to the same geometrical periodic orbits spanned in two directions. Therefore, arithmetic systems are *very exceptional* in this respect as they display exponentially large multiplicities of periodic orbit lengths. Notice, nevertheless, that according to Horowitz-Randol theorem [41], [51] this degeneracy is unbounded for any surface generated by a discrete group. However degeneracies of this theorem are much smaller than exponential.

3 Diagonal Approximation for Arithmetic Systems

The large multiplicities of periodic orbit lengths in arithmetical systems seem to have no importance in classical mechanics. These systems are as chaotic as any other models of free motion on constant negative curvature surfaces with finite area. Nevertheless, the quantum spectra of these systems are anomalous: is it connected to these degeneracies? In this Section we estimate the quantum two-point correlation form factor for arithmetic systems in the diagonal approximation as was done in Sect. 1.1 for generic chaotic systems.

Assume that there exist $g(l)$ periodic orbits with the same length l . Exactly as it was done in Sect. 1.1 one gets the following expression for the diagonal approximation of the two-point correlation function

$$R_2^{(diag)}(\epsilon) = \sum_{p,n} |A_{p,n}(l_p)|^2 g(l_p) e^{inT_p \epsilon} + \text{c.c.} \quad (77)$$

where the summation is done over all periodic orbits. The only difference with (48) is that in Sect. 1.1 it was assumed that g is a constant but here the multiplicity $g = g(l)$.

Define the two-point correlation form factor as the Fourier transform of $R_2(\epsilon)$

$$K(t) = \int_{-\infty}^{+\infty} R_2(\epsilon) e^{it\epsilon} .$$

This definition differs from the previous one by the absence of the factor 2π in the exponent. For later purposes it is more convenient.

Equation (77) leads to the following expression for the two-point correlation form factor in the diagonal approximation

$$K^{(diag)}(t) = 2\pi \sum_{p,n} |A_{p,n}(l_p)|^2 g(l_p) \delta\left(t - \frac{nl_p}{2k}\right) . \quad (78)$$

From (76) it follows that the mean multiplicity of periodic orbit lengths for arithmetic systems is asymptotically

$$\langle g(l_p) \rangle = \frac{2e^{l_p/2}}{Cl_p}$$

with a model dependent constant C (for the modular group $C = 1$).

For any models generated by discrete groups the summation over all periodic orbits is asymptotically equals the integration with the following measure

$$\sum_{l_p} \rightarrow \int \frac{dl}{l} e^l .$$

Taking into account that when $l \rightarrow \infty$ the term with $n = 1$ dominates and (see (47))

$$A_{p,1}(l) \xrightarrow{l \rightarrow \infty} \frac{le^{-l/2}}{4\pi k}$$

one obtains that in the diagonal approximation

$$K^{(diag)}(t) \sim \frac{e^{kt}}{2\pi kC}. \tag{79}$$

It means that the correlation form factor $K(t)$ for arithmetic systems grows much faster than was usually assumed and that for time of order of the Ehrenfest time it becomes of the order of 1.

The simplest approximation to the full form factor is the following

$$K(t) = \begin{cases} K^{diag}(t) & \text{for } t < t^* \\ \bar{d} & \text{for } t > t^* \end{cases}$$

where t^* is defined by the requirement that $K^{diag}(t^*) = \bar{d}$

$$t^* \sim \frac{1}{k} \ln(2\pi kC\bar{d}).$$

For the true Poisson statistics $K(t)$ always equals \bar{d} . For usual integrable systems $K(t)$ increases to this value during the time of the order of shortest periodic orbit periods, $t^* \sim 1/k$. For arithmetic systems $K(t)$ jumps to the universal saturation value in a time of order of the Ehrenfest time which has an additional logarithm of the momentum.

Therefore, spectral statistics of arithmetic systems is much closer to the Poisson prediction typical for integrable systems than to any of standard random matrix ensembles conjectured for generic ergodic systems.

4 Exact Two-Point Correlation Function for the Modular Group

The diagonal approximation gives quite crude estimate of the form factor. For the modular group it is possible to compute explicitly the two-point correlation function [23]. The calculations are based on a generalization of the Hardy–Littlewood method and depend strongly on the number-theoretical properties of the multiplicities of the periodic orbits of the modular group. In Sect. 4.1 using the Selberg trace formula the two-point correlation form factor is expressed through the two-point correlation function of multiplicities of periodic orbit lengths for the modular group. In Sect. 4.2 the latter is calculated by a certain generalization of the Hardy–Littlewood method. Quite tedious explicit formulas are given in Sect. 4.3 and the final expression for the two-point correlation form factor is presented in Sect. 4.4.

4.1 Basic Identities

The modular group has been considered in Sect. 1. It is the group of all 2×2 matrices with integer elements and unit determinant. The periodic orbits of the modular group correspond in a unique way to the conjugacy classes of hyperbolic elements of the modular group. The length of periodic orbit l_p is related with the trace of a representative matrix of the conjugacy class as follows

$$|\mathrm{Tr}M| = 2 \cosh l_p/2 .$$

As all elements of modular group matrices are integers the trace is also an integer

$$|\mathrm{Tr}M| = n .$$

In Sect. 1 it was demonstrated that the *mean* multiplicity of periodic orbit length for the modular group is

$$\langle g(l) \rangle = 2 \frac{e^{l/2}}{l} .$$

Denote by n the trace of a given conjugacy class and by $g(n)$ the number of distinct conjugacy classes corresponding to trace n . As n goes as $e^{L/2}$ when $n \rightarrow \infty$ one concludes that

$$\langle g(n) \rangle \xrightarrow{n \rightarrow \infty} \frac{n}{\ln n} . \quad (80)$$

According to the Selberg trace formula the density of eigenvalues for the modular surface $d(E) = \bar{d}(E) + d^{(osc)}(E)$ where the oscillating part of the density is represented by the following formal sum

$$d^{(osc)}(E) = \frac{2}{\pi k} \sum_n g(n) \frac{\ln n}{n} \cos(2k \ln n) .$$

From (80) it follows that mean value of $g(n) \ln n/n$ is one. Therefore we define

$$\alpha(n) = g(n) \frac{\ln n}{n} ,$$

so

$$d^{(osc)}(E) = \frac{1}{\pi k} \sum_n \alpha(n) \cos(2k \ln n)$$

and $\langle \alpha(n) \rangle = 1$.

As it was done in Sect. 1 one gets

$$R_2(\epsilon_1, \epsilon_2) = \bar{d}^2 + R_2^c(\epsilon_1, \epsilon_2)$$

where

$$R_2^\epsilon(\epsilon_1, \epsilon_2) = \frac{1}{(2\pi k)^2} \sum_{n_1, n_2} \alpha(n_1)\alpha(n_2) \left\langle e^{2i(k_1 \ln n_1 + k_2 \ln n_2)} + e^{2i(k_1 \ln n_1 - k_2 \ln n_2)} \right\rangle + \text{c.c.}$$

and

$$k_i \approx \sqrt{E + \epsilon_i} \xrightarrow{E \rightarrow \infty} k + \epsilon_i/2k.$$

As was discussed in Sect. 1 due to the energy average the first term will be washed out and the second one gives contributions only when

$$n_2 = n_1 + r \quad \text{with } r \ll n_1 .$$

Finally $R_2^\epsilon(\epsilon_1, \epsilon_2) = \bar{R}_2(\epsilon)$ where $\epsilon = \epsilon_1 - \epsilon_2$ and

$$\bar{R}_2(\epsilon) = \frac{1}{4\pi^2 k^2} \sum_n \sum_r \alpha(n)\alpha(n+r) \exp\left(-2i\frac{kr}{n} + i\epsilon\frac{\ln n}{k}\right) + \text{c.c.} .$$

Let assume that the following mean value exists

$$\gamma(r) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \alpha(n)\alpha(n+r) .$$

The dominant contribution to the two-point correlation function corresponds to replace the product $\alpha(n)\alpha(n+r)$ by its mean value $\gamma(r)$

$$\bar{R}_2(\epsilon) \approx \frac{1}{4\pi^2 k^2} \int_{n_0}^{\infty} dn \sum_{r=-\infty}^{\infty} \gamma(r) e^{-2ikr/n} \exp\left(i\epsilon\frac{\ln n}{k}\right) + \text{c.c.} \quad (81)$$

where we have used a continuum approximation for n starting formally from a certain fixed $n_0 \gg 1$, since only large values of n make a significant contribution.

Define a (real) function $f(x)$ as follows

$$f(x) = \sum_{r=-\infty}^{\infty} \gamma(r) e^{-irx} . \quad (82)$$

This function has the meaning of the Fourier transform of the two-point correlation function for multiplicities of the modular group.

After the changing variable $n \rightarrow e^{uk}$ in (81) one gets that the two-point correlation function for the modular group is expressed through $f(x)$ as follows

$$\bar{R}_2(\epsilon) \approx \frac{1}{2\pi^2 k} \int_0^{\infty} e^{ku} f(2ke^{-ku}) \cos \epsilon u \, du \quad (83)$$

and the two-point correlation form factor is

$$K(t) = \int_{-\infty}^{\infty} \overline{R}_2(\epsilon) e^{i\epsilon t} d\epsilon = \frac{1}{2\pi k} e^{kt} f(2ke^{-kt}) . \quad (84)$$

Therefore all non-trivial information is contained in functions $\gamma(r)$ or $f(x)$.

The simplest diagonal approximation is to assume that the $\alpha(n)$ are essentially uncorrelated, that is, $\gamma(r)$ is zero for $r \neq 0$. This gives for $f(x)$ a constant value which leads to an exponential growth of $K(t)$ as in (79). But from general considerations $K(t)$ obtained from a discrete spectrum has to saturate to a constant value for $t \rightarrow \infty$, consequently, the diagonal approximation cannot be correct for large t .

4.2 Two-Point Correlation Function of Multiplicities

The purpose of this Section is to calculate the two-point correlation function of modular group multiplicities, $\gamma(r)$, whose Fourier harmonics according to (84) determines the two-point correlation form factor.

The calculation will be done by a generalization of the Hardy-Littlewood method for prime pairs discussed in Sect. 2.1. As for primes one has to perform the three following steps.

The first step

Define the mean value of $\alpha(n)$ when n runs over integers of the form $mq + r$ for fixed q and $r < q$ in the following way

$$\alpha(q; r) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{m=0}^{N-1} \alpha(mq + r) .$$

Since $\langle \alpha(n) \rangle = 1$

$$\sum_{r=0}^{q-1} \alpha(q; r) = q .$$

Let M_q be the set of 2×2 matrices with entries being integers modulo q and having determinant equals one modulo q . These matrices form a group under multiplication modulo q which is sometimes called the modular group.

Define also $M_{q,r}$ to be the set of elements of M_q with trace equal to r modulo q . One can prove [23] that

$$\alpha(q; r) = \frac{q|M_{q,r}|}{|M_q|}$$

where $|M|$ is the number of elements of a set M .

The intuitive meaning of this result is the following: $g(n)$ is the number of conjugacy classes of modular matrices of trace n . To each modular matrix, one can associate an element of M_q in a unique way simply by taking the entries of the matrix modulo q . If n is equal to r modulo q , then all these matrices will belong to $M_{q,r}$. If we therefore assume that the matrices of the modular group cover the set M_q in some sense uniformly, the result follows. More careful treatment has been performed in [23].

Example

Let us consider $q = 2$. Integers modulo 2 are 0 and 1. The group M_2 consists of the following matrices

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

The dimension of the group M_2 , i.e. the total number of matrices, $|M_2| = 6$. Besides these matrices there are four matrices with zero trace (mod 2), i.e. $|M_{2,0}| = 4$, and two matrices with trace equals 1 (mod 2), $|M_{2,1}| = 2$. Therefore

$$\alpha(2; 0) = \frac{2 \cdot 4}{6} = \frac{4}{3}, \quad \alpha(2; 1) = \frac{2 \cdot 2}{6} = \frac{2}{3}.$$

The second step

Define as in the Hardy-Littlewood method the following function

$$\Phi(z) = \sum_{n=0}^{\infty} \alpha(n)z^n.$$

Since $\langle \alpha(n) \rangle = 1$ the convergence radius of this series is equal to one.

The importance of this function follows from the integral

$$J_r(e^{-u}) = e^{ru} \int_0^{2\pi} \frac{d\phi}{2\pi} \Phi^*(e^{-u+i\phi}) \Phi(e^{-u-i\phi}) e^{-ir\phi} = \sum_{n=1}^{\infty} \alpha(n)\alpha(n+r)e^{-2nu}$$

whose right-hand side by a Tauberian theorem is connected with the two-point correlation function of multiplicities, $\gamma(r)$.

The essence of the Hardy-Littlewood approach is the investigation of the function $\Phi(z)$ when $z = \exp(-u + i\epsilon + 2\pi ip/q)$ with $u \rightarrow 0$ and $\epsilon \rightarrow 0$, where p and q are co-prime integers. The main step is then to write n in the form $mq + r$ with r lying between 0 and $q - 1$ and prove that in the expression for $\Phi(z)$ the dominant contribution as u and ϵ go to zero will be given by the mean value of $\alpha(mq + r)$, that is, one may substitute it by $\alpha(q; r)$.

Accepting this, one has that as $u \rightarrow 0$ and $\epsilon \rightarrow 0$

$$\begin{aligned} \Phi(\exp(-u + 2\pi ip/q + i\epsilon)) &= \sum_{r=0}^{q-1} \sum_{m=0}^{\infty} \alpha(mq + r) e^{-(u-i\epsilon)(mq+r)} e^{2\pi irp/q} = \\ &= \sum_{r=0}^{q-1} \alpha(q; r) e^{2\pi ipr/q} \frac{1}{q} \int_0^{\infty} dn e^{-(u-i\epsilon)n} = \\ &= \frac{\beta(p, q)}{u - i\epsilon} \end{aligned}$$

where

$$\beta(p, q) = q^{-1} \sum_{r=0}^{q-1} \alpha(q; r) \exp\left(2\pi i \frac{p}{q} r\right).$$

Hence $\Phi(z)$ has a pole singularity at all rational points on the unit circle.

The third step

Divide the unit circle in intervals $I_{p,q}$ centered around $\exp(2\pi i p/q)$, where p and q are co-prime integers with $p < q$. If one neglects all terms in each interval except the pole term and extends the integration over ϵ to the whole line, one gets

$$\begin{aligned} J_r(e^{-u}) &= e^{ru} \sum_{(p,q)=1} \int_{-\infty}^{\infty} \frac{d\epsilon}{2\pi} \frac{|\beta(p, q)|^2}{u^2 + \epsilon^2} e^{ir(2\pi p/q + \epsilon)} = \\ &= \frac{1}{2u} \sum_{(p,q)=1} |\beta(p, q)|^2 \exp\left(2\pi i \frac{p}{q} r\right). \end{aligned}$$

Finally one obtains that

$$\gamma(r) = \sum_{(p,q)=1} |\beta(p, q)|^2 \exp\left(2\pi i \frac{p}{q} r\right).$$

The sum is performed over all q , and p co-prime to q with $0 < p < q$.

This is the two-point correlation function of *multiplicities* of the periodic orbits for the modular group. All other quantities of interest can be obtained from it. In particular, the function $f(x)$ (82) is given by

$$f(x) = 2\pi \sum_{(p,q)=1} |\beta(p, q)|^2 \delta\left(x - 2\pi \frac{p}{q}\right)$$

where the summation is done over all p and q co-prime, without the restriction $p < q$.

According to (83) and (84) the knowledge of $f(x)$ determines immediately the two-point correction function the form factor of modular domain eigenvalues.

4.3 Explicit Formulas

Let us define the so-called Kloosterman sums

$$S(n, m; c) = \sum_{(d,c)=1} \exp\left(\frac{2\pi i}{c}(nd + md^{-1})\right)$$

where the summation is taken over all $d < c$ co-prime with c and d^{-1} is an integer modulo c which obeys $d^{-1}d = 1 \pmod{c}$.

One can show (see [23]) that $\beta(p, q)$ can be expressed through these sums in the following way

$$\beta(p, q) = \frac{1}{q^2 \prod_{\omega|q} (1 - \omega^{-2})} S(p, p; q)$$

where ω are the prime divisors of q .

The function $\gamma(r)$ can be written as

$$\gamma(r) = \sum_{n=1}^{\infty} A_r(n)$$

where $A_r(q)$ is given by

$$A_r(q) = \sum_{p:(p,q)=1} |\beta(p, q)|^2 \exp\left(2\pi i r \frac{p}{q}\right).$$

One can prove that $A_r(q)$ is multiplicative function of q , i.e. $A_r(n_1 n_2) = A_r(n_1) A_r(n_2)$ provided $(n_1, n_2) = 1$, therefore one needs to know only its values on powers of primes and $\gamma(r)$ can be rewritten as the infinite product over all prime numbers

$$\gamma(r) = \prod_p \left(1 + \sum_{k=1}^{\infty} A_r(p^k)\right)$$

To present a closed expression for $A_r(q)$ let us introduce the standard definition of the Legendre symbol

$$\left(\frac{a}{q}\right) = \begin{cases} 1, & \text{if } a \equiv x^2 \pmod{q} \text{ has a solution } a \not\equiv 0 \pmod{q} \\ 0, & \text{if } a \equiv 0 \pmod{q} \\ -1, & \text{otherwise} \end{cases}.$$

The meaning of this symbol is perhaps best understood by saying that the number of *distinct* solutions of the equation $x^2 \equiv a \pmod{q}$ is $1 + (a/q)$.

A fairly tedious evaluation of $A_r(q)$ (see [23] for details) gives the following formulas.

Let $q = p^n$ where p is an odd prime. Then for $n = 1$ one has

$$A_r(p) = \frac{1}{(p^2 - 1)^2} \left[p \sum_{x=0}^{p-1} \left(\frac{(x^2 - 4)((x + r)^2 - 4)}{p} \right) - 1 \right].$$

For $n \geq 2$ we have, letting t be an arbitrary non-zero number modulo p ,

$$A_r(p^n) = \frac{1}{p^{2n}(1 - p^{-2})} \begin{cases} 2(1 - 1/p), & r \equiv 0 & \pmod{p^n} \\ -2/p, & r \equiv tp^{n-1} & \pmod{p^n} \\ \epsilon(n, p)(1 - 1/p), & r \equiv \pm 4 & \pmod{p^n} \\ -\epsilon(n, p)/p, & r \equiv \pm 4 + tp^{n-1} & \pmod{p^n} \end{cases}$$

where $\epsilon(n, p)$ takes the value -1 if n is odd and p is of the form $4k + 3$ and is equal to 1 in all other cases. For $p = 2$, we list down individual cases for low powers and eventually state a general rule

$$\begin{aligned} A_r(2) &= \frac{1}{9} \begin{cases} 1, & r \equiv 0 \pmod{2} \\ -1, & r \equiv 1 \pmod{2} \end{cases}, \\ A_r(4) &= \frac{1}{18} \begin{cases} 1, & r \equiv 0 \pmod{4} \\ -1, & r \equiv 2 \pmod{4} \end{cases}, \\ A_r(8) &= 0, \\ A_r(16) &= \frac{1}{9 \cdot 16} \begin{cases} 1, & r \equiv 0 \pmod{16} \\ -1, & r \equiv 8 \pmod{16} \end{cases}, \\ A_r(32) &= 0, \end{aligned}$$

and finally, for the general case $n \geq 6$

$$A_r(2^n) = \frac{1}{9 \cdot 2^{2n-4}} \begin{cases} 2, & r \equiv 0 \pmod{2^n} \\ -2, & r \equiv 2^{n-1} \pmod{2^n} \\ 1, & r \equiv \pm(4 + 2^{n-2}) \pmod{2^n} \\ -1, & r \equiv \pm(4 + 2^{n-2} + 2^{n-1}) \pmod{2^n} \end{cases}.$$

All terms not explicitly shown equal zero. In [50] these formulas were proved by a different method.

4.4 Two-Point Form Factor

These formulas give the explicit expression for the two-point correlation form factor

$$K(t) = \frac{1}{2\pi^2 k} \sum_{(p,q)=1} \left| \frac{q}{p} \beta(p, q) \right|^2 \delta(t - t_{p,q}).$$

where

$$t_{p,q} = \frac{1}{k} \ln \frac{kq}{\pi p}.$$

In the limit $k \rightarrow \infty$ and t fixed, the dominant contribution comes from terms with $p/q \ll 1$. Smoothing over such values one can show (see [23]) that in this limit $K(t)$ tends to the constant Poisson value

$$K(t) = \frac{A}{4\pi}$$

where $A = \pi/3$ is the area of the fundamental region of the modular group. For small t (of the order of the Ehrenfest time $\ln k/k$) $K(t)$ has number-theoretical oscillations due to cumulative contributions of degenerate periodic orbits. For very small values of t (of the order of $1/k$) the two-point form factor has δ peaks connected with short periodic orbits.

Though the modular group is by no means a generic system, it is the first ergodic dynamical system for which it was possible to compute explicitly the distribution of the energy levels.

5 Hecke operators

Arithmetic groups have many interesting properties. In particular, for all arithmetic groups it is possible to construct an infinite number of mutually commuting operators which commute also with the Laplace–Beltrami operator. These operators are of pure arithmetic origin and are called the *Hecke operators* [37], [54].

In a certain sense these operators permit to ‘understand’ why arithmetic systems have the Poisson statistics typical only for integrable systems. The point is that integrable systems are systems with sufficiently large number of independent commuting operators and Hecke operators may be viewed as a manifestation of a kind of arithmetic integrability of arithmetic systems which does the Poisson statistics for these models natural [27]. Unfortunately, precise relations along this line seem to be impossible.

Let us consider informally the construction of Hecke operators for the modular group. Choose two matrices A and B from the modular group with the same trace. As they have the same trace and determinant, they have the same eigenvalues and there exists a matrix γ such that

$$\gamma A \gamma^{-1} = B \text{ or } \gamma A = B \gamma \text{ and } \det(\gamma) \neq 0. \quad (85)$$

If A and B are not conjugated in the modular group, $\gamma \notin \text{PSL}(2, \mathbb{Z})$. But matrix γ can be chosen as a matrix with integer elements but with the determinant $\neq 1$.

Example.

Consider the following simple matrices

$$A = \begin{pmatrix} 3 & 1 \\ 2 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 3 & 2 \\ 1 & 1 \end{pmatrix}.$$

General form of matrices γ which obey (85) is

$$\gamma = \begin{pmatrix} 2\alpha + 2\beta & \alpha \\ \alpha & \beta \end{pmatrix}$$

with arbitrary α and β .

It is clear that there exists no $\gamma \in \text{PSL}(2, \mathbb{Z})$ but choosing different integer values of α and β one can construct an infinite number of integer matrices with determinant $\neq 1$ which obeys (85). For example,

$$\gamma = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} (\det = -1), \quad \gamma = \begin{pmatrix} 4 & 1 \\ 1 & 1 \end{pmatrix} (\det = 3) \dots$$

These considerations demonstrate that when dealing with the modular group it is quite natural to consider matrices with integer elements but with the determinant different from one

$$M_p = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mid a, b, c, d \text{ integers, } ad - bc = p \right\} .$$

Matrices M_p with $p \neq 1$ do not form a group because their product has not the same form.

A matrix $m_p \in M_p$ can uniquely be represented in the form

$$m_p = \mu \alpha_p \tag{86}$$

where $\mu \in \text{PSL}(2, \mathbb{Z})$ and α_p is one of matrices from the following finite set

$$\alpha_p = \left\{ \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \mid a, b, d \text{ integers, } ad = p, d > 0, 0 \leq b \leq d - 1 \right\} . \tag{87}$$

Instead of proving this fact let us transform a simple matrix

$$m_3 = \begin{pmatrix} 4 & 1 \\ 1 & 1 \end{pmatrix}$$

to the form (86). General proof (see e.g. [54]) follows the same steps. First, it is necessary to find a matrix

$$\mu' = \begin{pmatrix} \mu_1 & \mu_2 \\ \mu_3 & \mu_4 \end{pmatrix}$$

such that (i) $\det \mu' = 1$ and (ii)

$$\begin{pmatrix} \mu_1 & \mu_2 \\ \mu_3 & \mu_4 \end{pmatrix} \begin{pmatrix} 4 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} .$$

The condition of the zero of low-left element gives the equation $4\mu_3 + \mu_4 = 0$ and because μ_4 and μ_3 are coprime they can be chosen as follows: $\mu_4 = 4$ and $\mu_3 = -1$. Unit determinant condition gives $\mu_1 = k$ and $\mu_2 = 1 - 4k$ with an arbitrary integer k . Finally, $b = 1 - 3k$ and the smallest positive b modulo 3 corresponds to $k = 0$. Hence, our matrix m_3 has the following representation

$$\begin{pmatrix} 4 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 4 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 3 \end{pmatrix} .$$

An important property of the set M_p is that when one multiplies a matrix from this set by a matrix from the modular group the resulting matrix also belongs to M_p

$$M_p g = M_p \text{ for all } g \in \text{PSL}(2, \mathbb{Z}) .$$

Let $\Psi(z)$ be an *automorphic function* of the modular group, i.e.

$$\Psi(gz) = \Psi(z) , \text{ for all } g \in \text{PSL}(2, \mathbb{Z}) .$$

Then it is easy to see that the function

$$\Psi'(z) \equiv (T_p\Psi)(z) = \frac{1}{\sqrt{p}} \sum_{a,b,d} \Psi\left(\frac{az + g}{d}\right)$$

where the summation is performed over all $ad = p, d > 0, 0 \leq b \leq d - 1$ will also be an automorphic function for the modular group. This is a consequence of the fact that in the right-hand side of this expression there is effectively the summation over all matrices from M_p . As M_p goes not change after multiplication by a modular group matrix $\Psi'(z)$ is an automorphic function for the modular group. $(T_p\Psi)(z)$ is a kind of symmetrization of $\Psi(z)$ over images of z by all elements of M_p and the operators T_p are called Hecke operators.

These operators form a *commutative algebra* with the following product (see e.g. [54])

$$T_n T_m = \sum_{d|(n,m)} T_{nm/d^2} \tag{88}$$

where the summation is done over all divisors of the greatest common divisor of m and n . The most important case corresponds to Hecke operators with prime indices because all the others can be obtained from (88).

When p is a prime number

$$(T_p\Psi)(z) = \frac{1}{\sqrt{p}} \left[\Psi(pz) + \sum_{0 \leq j < p} \Psi\left(\frac{z + j}{p}\right) \right].$$

Since Hecke operators involve only fractional transformations all them commute with the Laplace–Beltrami operator. Consequently, if $\Psi(x, y)$ is an eigenfunction of the Laplace–Beltrami operator, then $(T_p\Psi)(x, y)$ will also be an eigenfunction with the same eigenvalue. If there is no spectral degeneracy (which strongly suggested by numerics) every eigenfunction of the Laplace–Beltrami operator is in the same time an eigenfunction of all Hecke operators

$$(T_p\Psi_n)(x, y) = \lambda_p(n)\Psi_n(x, y). \tag{89}$$

It is known (see e.g. [54]) that eigenfunctions of the Laplace–Beltrami operator for the modular group have the following Fourier expansion

$$\Psi_n(x, y) = y^{1/2} \sum_{p=-\infty}^{\infty} c_p(n) K_{s_n-1/2}(2\pi p y) e^{2\pi i p x}$$

where the eigenvalue of the Laplace–Beltrami operator $E_n = s_n(s_n - 1)$ and $K_\nu(x)$ is the Hankel function.

One has $z = x + iy$ and $(az + b)/d = (ax + b)/d + iay/d$, therefore

$$(T_m\Psi_n)(x, y) = \frac{1}{\sqrt{m}} \sum_{a,b,d} \left(\frac{ay}{d}\right)^{1/2} \sum_p c_p(n) K_{s_n-1/2}\left(2\pi p \frac{ay}{d}\right) e^{2\pi i p (ax+b)/d}$$

where the first summation is performed over all a, b, d as in (87).

The summation over b gives zero if d is not divide p . Otherwise

$$(T_m \Psi_n)(x, y) = y^{1/2} \sum_{d|p, d|m} c_p(n) K_{s_n-1/2}(2\pi y p m / d^2) e^{2\pi i p m x / d^2} .$$

Let $k = m/d$ and $u = p m / d^2$. Then $p = m u / k^2$ and

$$(T_m \Psi_n)(x, y) = y^{1/2} \sum_u \sum_{k|(m, u)} c_{m u / k^2}(n) K_{s_n-1/2}(2u y) e^{2\pi i u x} .$$

If $T_m \Psi_n = \lambda_m(n) \Psi_n$ then by comparing the first Fourier coefficient one gets

$$c_m(n) = \lambda_m(n) c_1 .$$

Assuming $c_1 \neq 0$ and using a convenient normalization $c_1 = 1$ one concludes that *eigenvalues of the Hecke operators coincide with the Fourier coefficients*.

We note also that similarly to the construction of the Selberg trace formula one can build the trace formulas for Hecke operators (see e.g. [24] and references therein). Such trace formula schematically has the form (cf. (29))

$$\sum_n \lambda_p(n) h(k_n) = \frac{1}{\sqrt{p}} \sum_{\text{hyperbolic}} \frac{l_p}{2 \sinh(L_p/2)} g(L_p) + \text{smooth, parabolic, and elliptic terms} .$$

Here $h(k)$ is a test function like in Sect. 2.8 and $g(l)$ is its Fourier transform. In the left-hand side the summation is performed over all eigenvalues $E_n = k_n^2 + 1/4$ of the Laplace–Beltrami operator and $\lambda_p(n)$ is the eigenvalue of the Hecke operator T_p (89) applied to the eigenfunction of the Laplace–Beltrami operator with eigenvalue E_n . In the right-hand side the summation is done over all ‘hyperbolic’ matrices from M_p with $\text{Tr } m_p \neq p + 1$. L_p is the ‘length’ associated with matrix m_p

$$2 \cosh(L_p/2) = |\text{Tr } m_p| / \sqrt{p}$$

and l_p is the minimal length of modular group matrices commuting with m_p .

6 The Jacquet–Langlands Correspondence

Another curious fact about arithmetic groups is the Jacquet–Langlands correspondence (see [40]) which claims that for a arithmetic group derived from a quaternion group over \mathbb{Q} (with a finite fundamental domain) one can find a subgroup of the modular group (with infinite fundamental domain) in such a way that amongst all automorphic eigenvalues of the Laplace–Beltrami operator for this modular group subgroup one can find all eigenvalues of a compact arithmetic group.

The simplest arithmetic group Γ derived from quaternion algebra over \mathbb{Q} with division is (see Sect. 2.2)

$$\Gamma = \begin{pmatrix} k_1 + k_2\sqrt{a} & k_3 + k_4\sqrt{a} \\ b(k_3 - k_4\sqrt{a}) & k_1 - k_2\sqrt{a} \end{pmatrix}$$

where b is a prime number, a is an integer such that the equation $x^2 \equiv a \pmod{b}$ has no integer solution (e.g. $a = 3, b = 5$), and integers k_i are such that

$$\det(\gamma) = k_1^2 - ak_2^2 - bk_3^2 + abk_4^2 = 1 .$$

Denote $z = x + iy, \tau = u + iv$ ($y, v > 0$) and define for all n_j

$$\alpha = n_1 + n_2\sqrt{a}, \beta = n_3 + n_4\sqrt{a},$$

$$\gamma = b(n_3 - n_4\sqrt{a}), \delta = n_1 - n_2\sqrt{a} .$$

Fix an arbitrary z_0 and compute the following kernel

$$\Phi(\tau, z) = \sum_{n_j=-\infty}^{+\infty} \exp K(\tau, z)$$

where

$$K(\tau, z) = -\pi \operatorname{Im} \tau \frac{|\alpha z_0 + \beta - z(\gamma \bar{z}_0 + \delta)|^2}{\operatorname{Im} z \operatorname{Im} z_0} + 2\pi i \bar{\tau}(\alpha \delta - \beta \gamma) .$$

Here \bar{z} is the complex conjugate of z .

Let $\psi_n(z)$ be an eigenfunction of the Laplace–Beltrami operator automorphic with respect to the quaternion group Γ . It means that

- $(\Delta_{L-B} + E_n) \psi_n(z) = 0,$
- $\psi_n(Mz) = \psi_n(z)$ for all $M \in \Gamma$.

Then the function

$$\Psi(\tau) = \operatorname{Im} \tau \int_{\mathcal{D}} \Phi(\tau, z) \psi_n(z) \frac{dx dy}{y^2}$$

where the integral is taken over the fundamental domain of the group Γ is an eigenfunction the Laplace–Beltrami operator with the same eigenvalue E_n but automorphic with respect to the congruence subgroup of the modular group $\Gamma_0(4ab)$ where

$$\Gamma_0(N) = \begin{pmatrix} m & n \\ k & l \end{pmatrix} \in \operatorname{SL}(2, \mathbb{Z})$$

with an additional condition that

$$k \equiv 0 \pmod{N} .$$

Direct (but tedious) proof of this statement can be found in [40].

7 Non-arithmetic Triangles

In the precedent Section we have seen that arithmetic systems have the Poisson spectral statistics. But what is about non-arithmetic models?

Let us consider, as example, the so-called *Hecke* triangles which are hyperbolic triangles with angles $(0, \pi/2, \pi/n)$. All of them tessellate the upper half-plane and are fundamental domains of the discrete groups generated by reflections across their sides. The modular billiard is a part of them corresponding to $n = 3$. Similar to it they all have an infinite cusp.

According to Table 1 the Hecke triangles are arithmetic only for $n = 3, 4, 6, \infty$. All these arithmetic triangles have an exponential degeneracies of periodic orbit lengths which leads to the Poisson-like statistics of energy levels.

The simplest non-arithmetic Hecke triangle is the one with $n = 5$. At Fig. 19 we present the results of numerical calculations of the nearest-neighbor distribution for 6000 first energy levels for this triangle with the Dirichlet boundary conditions. For others Hecke triangles one gets similar pictures. It is clearly seen that numerics agrees very well with the predictions of the Gaussian Orthogonal ensembles of random matrices as it should be for generic chaotic models.

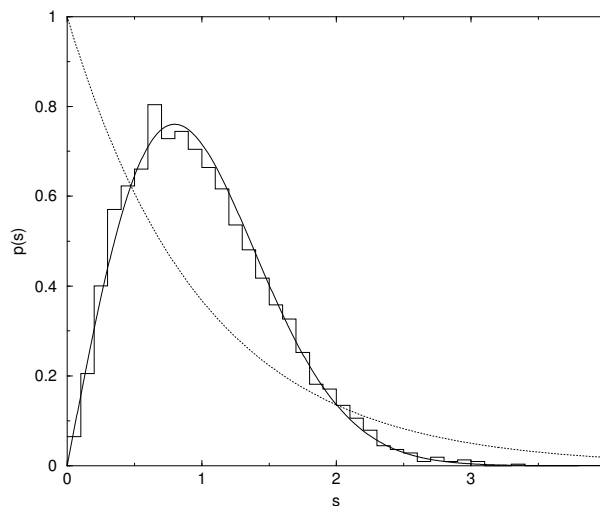


Fig. 19. The nearest-neighbor distribution of 6000 energy levels for the non-arithmetic Hecke triangular billiard with $n = 5$. The solid line – the GOE prediction. Dotted line – the Poisson result.

But what are the multiplicities of periodic orbit lengths for non-arithmetic Hecke triangles? As these model are not-arithmetic, one would expect that their length multiplicities should be equal to two as for generic time-reversal invariant systems. Nevertheless numerical calculations (see [24] for details)

demonstrated that this is not always the case. At Fig. 20 we present the

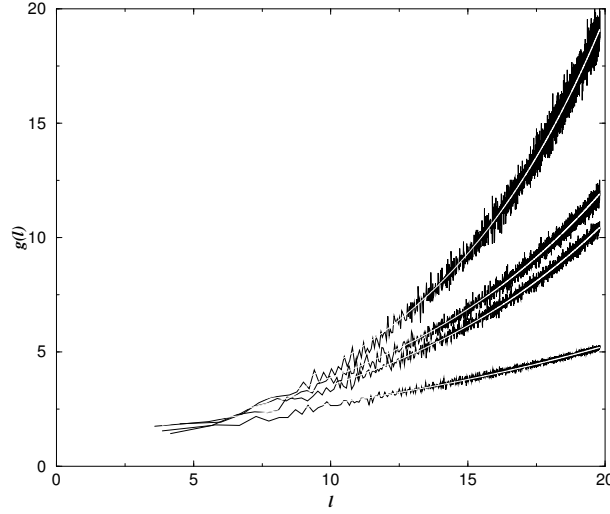


Fig. 20. Mean length multiplicities of periodic orbits for the Hecke triangles with (from top to bottom) $n = 12$, $n = 5$, $n = 8$, and $n = 10$. White lines are numerical fits (90).

numerically computed mean length multiplicities for the Hecke triangles with $n = 5, 8, 10, 12$ for lengths $l < 20$. White lines indicate a two-parameter fit to these data in the form $\bar{g}(l) = a_n e^{b_n l}$

$$\begin{aligned}
 n = 5 & : \bar{g}(l) \approx 1.235e^{.114l} , \\
 n = 8 & : \bar{g}(l) \approx 1.095e^{.114l} , \\
 n = 10 & : \bar{g}(l) \approx 1.143e^{.065l} , \\
 n = 12 & : \bar{g}(l) \approx .986e^{.150l} .
 \end{aligned} \tag{90}$$

These expressions fit numerical data in the given interval of lengths quite well and indicate that for, at least, certain Hecke triangles mean length multiplicity increases exponentially. We stress that (90) are only the best least-square numerical fits and no attempts were made to determine the accuracy of coefficients.

The discussion of the origin of such unexpected multiplicities for non-arithmetic triangles is beyond the scope of these lectures (on this subject see [26]). However it is of interest to understand why exponentially large multiplicities of periodic orbit lengths do not contradict the observed GOE behaviour of spectral statistics (cf. Fig. 19).

Assume that a system has an exponentially large number of periodic orbits with the same length l increasing as

$$g(l) \sim \frac{e^{\lambda l}}{l}$$

with $\lambda \leq 1/2$.

Let us repeat the arguments of Sect. 1.2 for this case with exact degeneracies. In Sect. 1.2 it was demonstrated that periodic orbits with different lengths can be treated in the diagonal approximation if

$$l_{p_1} - l_{p_2} \gg \frac{k}{\Delta E} \quad (91)$$

where k is the momentum and ΔE is the width of the energy average inherent in the definition of correlation functions of dynamical systems.

As the density of orbits with *different* lengths is

$$\rho_{\text{diff. lengths}} \approx \frac{e^l}{g(l)l} \sim e^{(1-\lambda)l}$$

it follows that the inequality (91) is valid till maximal length

$$l_m \sim \frac{1}{1-\lambda} \ln \frac{\Delta E}{k} \sim \frac{1}{1-\lambda} \ln k. \quad (92)$$

Notice that due to assumed large multiplicity l_m is different from (51).

From (78) it follows that the two-point correlation form factor in the diagonal approximation up to numerical factor is

$$K(t) \sim \frac{k}{l} |A(l)|^2 g(l) e^l$$

where $t = l/2k$ and $A(l) \sim le^{l/2}/k$. Combining all terms together one obtains that during the maximal time of applicability of the diagonal approximation $t_m = l_m/2k$ with l_m from (92) the form factor increases till

$$K(t_m) \sim \frac{e^{\lambda l_m}}{k} \sim k^{(2\lambda-1)/(1-\lambda)}.$$

Hence, if $\lambda = 1/2$ as for arithmetic systems the two-point correlation form factor during the time of validity of the diagonal approximation increases till a constant value of the order of 1. But if $\lambda < 1/2$ the form factor for the time of validity can reach only a value of the order of $k^{-\nu}$ with $\nu = (1-2\lambda)/(1-\lambda) > 0$. As $k \rightarrow \infty$ this value tends to zero and no apparent contradiction with standard random matrix ensembles can be derived within the diagonal approximation.

8 Summary

Arithmetic groups are a special sub-class of discrete groups characterized by the existence of a representation by matrices with integer elements. A readable

mathematical review of such groups is given in [42]. There are two types of arithmetic groups. The first includes groups commensurable with the modular group and having non-compact fundamental domains with infinite cusps. The second type of compact arithmetic groups combines groups commensurable with groups derived from quaternion algebras with division. These groups have finite fundamental domains.

From classical viewpoint the free motion on surfaces generated by arithmetic groups is as chaotic as for any hyperbolic surfaces. But quantum mechanics on these arithmetic surfaces is very special. In particular, spectral statistics of the Laplace–Beltrami operator automorphic with respect to arithmetic group is described by the Poisson statistics typical for integrable systems and not by the random matrix statistics typical for chaotic models.

The origin of this peculiarity can be traced to the existence in arithmetic systems of a very large number of periodic orbits with exactly the same length. For all arithmetic groups the mean multiplicity of periodic orbits with length l behaves like $e^{l/2}/l$. This has to be compared with the total density of periodic orbits which for all discrete groups is e^l/l . It is the cumulative effect of the interference of many periodic orbits with the same length which changes drastically the spectral statistics.

In the diagonal approximation the two-point correlation form factor $K(t)$ for arithmetic systems at small t increases exponentially like e^{kt}/k and during the Ehrenfest time (which is the limit of applicability of the diagonal approximation) reaches a constant value.

More detailed information can be obtained for the modular group where it is possible to compute the two-point correlation form factor analytically. The final answer is

$$K(t) = \frac{1}{2\pi^2 k} \sum_{(p,q)=1} \left| \frac{q}{p} \beta(p, q) \right|^2 \delta(t - t_{p,q})$$

where

$$t_{p,q} = \frac{1}{k} \ln \frac{kq}{\pi p}$$

and $\beta(p, q)$ is a number-theoretical function given in Sect. 4.3.

This formula means that the two-point form factor for the modular group is a sum over δ -functions at special points $t_{p,q}$ situated in a vicinity of the Ehrenfest time. The set of δ -functions is dense but the largest peaks correspond to the smallest ratios p/q . Nevertheless, small peaks with $p/q \ll 1$ are much more numerous and integrally they dominate. In the limit t fixed and $k \rightarrow \infty$ $K(t) \rightarrow \bar{d}$ thus confirming the Poisson nature of the spectral statistics of the modular group.

Arithmetic groups have many interesting properties. Hecke operators and the Jacquet–Langlands correspondence are the most remarkable.

Acknowledgement

It is a pleasure to thank Charles Schmit for his aid in the preparation of these lectures and A.M. Odlyzko for presenting numerical data of the two-point correlation function of Riemann zeros.

References

1. D. Alonso and P. Gaspard, \hbar -Expansion for the Periodic Orbit Quantization of Chaotic Systems, *Chaos* 3 (1993) 601-612; Erratum *ibid* 4 (1994) 105.
2. A.V. Andreev and B.L. Altshuler, Spectral Statistics beyond Random Matrix Theory, *Phys. Rev. Lett.* 75 (1995) 902-905.
3. R. Aurich, M. Sieber, and F. Steiner, Quantum Chaos on the Hadamard-Gutzwiller problem, *Phys. Rev. Lett.* 61 (1988) 483.
4. A. Aurich and F. Steiner, On the Periodic Orbits of a Strongly Chaotic System, *Physica D* 32 (1988) 451.
5. H.P. Baltes and E.R. Hilf, *Spectra of Finite Systems*, Wissenschaftsverlag, Mannheim, 1976.
6. R. Aurich, E. Bogomolny, and F. Steiner, Periodic Orbits on the Regular Octagon, *Physica D* 48 (1991) 91-101.
7. N.L. Balatz and A. Voros, Chaos on the Pseudosphere, *Phys. Rep.* 143 (1986) 109.
8. R. Balian and C. Bloch, Distribution of eigenfrequencies for the wave equation in a finite domain: I Three-dimensional problems with smooth boundary surface, *Ann. Phys.* 60 (1970) 401; II Electromagnetic Field. Riemannian spaces, *Ann. Phys.* 64 (1971) 271; III Eigenfrequency density fluctuations, *Ann. Phys.* 69 (1972) 76; *ibid* Asymptotic evaluation of the Green's function for large quantum numbers, *Ann. Phys.* 63 (1971) 592; *ibid* Solutions of the Schrödinger equation in terms of classical paths, *Ann. Phys.* 85 (1974) 514.
9. M.V. Berry and M. Tabor, Closed Orbits and the Regular Bound Spectrum, *Proc. R. Soc. Lond. A* 349 (1976) 101-123.
10. M.V. Berry and M. Tabor, Level Clustering in the Regular Spectrum, *Proc. R. Soc. London A* 356 (1977) 375-394.
11. M.V. Berry, Semiclassical Theory of Spectrum Rigidity, *Proc. R. Soc. London A* 400 (1985) 229-251.
12. M.V. Berry, Semiclassical formula for the number variance of the Riemann zeros, *Nonlinearity*, 1 (1988) 399-407.
13. M.V. Berry, Some Quantum-to-Classical Asymptotics, in [28] (1989) 251-303.
14. M.V. Berry and C.J. Howls, High Orders of the Weyl Expansion for Quantum Billiards, Resurgence of Periodic Orbits, and the Stokes Phenomenon, *Proc. R. Soc. Lond A* 447 (1994) 527-555.
15. M.V. Berry and J.P. Keating, $H = xp$ and the Riemann zeros, in 'Supersymmetry and trace formulas', eds. I.V. Lerner and J.P. Keating, Plenum, New York (1999) 355-367.
16. O. Bohigas, Random Matrix Theories and Chaotic Dynamics, in [28] (1989) 87-199.
17. O. Bohigas, M.-J. Giannoni, and C. Schmit, Characteristic of Chaotic Quantum Spectra and Universality of Level Fluctuations Law, *Phys. Rev. Lett.* 52 (1984)

- 1 ; Spectral Properties of the Laplacian and Random Matrix Theory, *J. Physique Lett.* 45 (1984) L-1015.
18. E. Bogomolny, B. Georgeot, M.J. Giannoni, and C. Schmit, Chaotic Billiards Generated by Arithmetic Groups, *Phys. Rev. Lett.* 69 (1992) 1477-1480.
 19. E. Bogomolny and C. Schmit, Semiclassical Computation of High-Excited Energy Levels, *Nonlinearity* 6 (1993) 523-547.
 20. E. Bogomolny, Introduction to models on constant negative curvature surfaces, in *Quantum Dynamics of Simple Systems, The Forty Fourth Scottish Universities Summer School in Physics, Stirling, 1994*, Eds. G-L Oppo, S. M. Barnett, E. Riis, and M. Wilkinson.
 21. E. Bogomolny and J.P. Keating, Gutzwiller's Trace Formula and Spectral Statistics: Beyond the Diagonal Approximation, *Phys. Rev. Lett.* 77 (1996) 1472-1475.
 22. E. Bogomolny and J.P. Keating, Random Matrix Theory and the Riemann Zeros I: , *Nonlinearity* 8 (1995) 1115-1131; Random Matrix Theory and the Riemann Zeros II: n-point correlations, *Nonlinearity* 9 (1996) 911-935.
 23. E. Bogomolny, F. Leyvraz, and C. Schmit, Distribution of Eigenvalues for the Modular Group, *Commun. Math. Phys.* 176 (1996) 577-617.
 24. E. Bogomolny, B. Georgeot, M.-J. Giannoni, and C. Schmit, Arithmetic Chaos, *Phys. Rep.* 291 (1997) 219-324.
 25. E. Bogomolny, Spectral Statistics and Periodic Orbits, in *Proceedings of the International School of Physics 'Enrico Fermi', Varenna (1999)*, New Directions in Quantum Chaos, Eds. G. Casati, I. Guarneri, and U. Smilansky, IOS Press, Amsterdam, Oxford, Tokyo, Washington, 2000, 333-369.
 26. E. Bogomolny and C. Schmit, Multiplicities of Periodic Orbits Lengths for Non-Arithmetic Models, (2003) nlin.CD/0312057.
 27. J. Bolte, G. Steil, and F. Steiner, Arithmetic Chaos and Violations of Universality in Energy Level Statistics, *Phys. Rev. Lett.* 69 (1992) 2188.
 28. Chaos and Quantum Physics, *Proceedings of the Les Houches Summer School (1989)* Eds. M.-J. Giannoni, A. Voros, J. Zinn-Justin. North Holland, Amsterdam, London, New York, Tokyo, 1991.
 29. Y. Colin de Verdière, Hyperbolic Geometry in Two-Dimensions and Trace Formulas, in [28] (1989) 305-330.
 30. A. Connes, Trace Formula in Non-Commutative Geometry and the Zeros of the Riemann Zeta Function, (1997) arXiv: math.NT/9811068.
 31. H. Davenport, *Multiplicative Number Theory*, revised by H. Montgomery, Springer-Verlag, New York, Heidelberg, Berlin, 1980.
 32. A. Erdélyi et al. *Higher Transcendental Functions Vol. 1* (Bateman Manuscript Project). McGraw-Hill, New York, 1953.
 33. M.C. Gutzwiller, *Chaos in Classical and Quantum Mechanics*, Springer, New York, 1990.
 34. J.H. Hannay and A.M. Ozorio de Almeida, *J. Phys. A* 17 (1984) 3429-3440.
 35. G.H. Hardy and J.E. Littlewood, Some Problems of 'Partitis Numerorum'; III: On the expression of a Number as a Sum of Primes, *Acta Mathematica*, 44 (1923) 1-70.
 36. G.H. Hardy and E.M. Wright, *An Introduction to the Theory of Numbers*, Clarendon Press, Oxford, 1979.
 37. E. Hecke, *Lectures on Dirichlet Series, Modular Functions and Quadratic Forms*, Vandenhoeck and Ruprecht, Göttingen, 1983.
 38. D. Hejhal, The Selberg Trace Formula and the Riemann Zeta Function, *Duke Math. J.* 43 (1976) 441-482.

39. D. Hejhal, The Selberg Trace Formula for $\mathrm{PSL}(2, \mathbb{R})$, Vol. 1, Lectures Notes in Mathematics 548 (1979); Vol. 2, *ibid* 1001 (1983).
40. D. Hejhal, A classical approach to a well known spectral correspondence on quaternion groups, in Number Theory, D.V. Chudnovsky, G.V. Chudnovsky, H. Cohn, M.B. Nathanson Eds., Lectures Notes in Mathematics 1135 (1985) 127.
41. R.D. Horowitz, Characters of Free Groups Represented in the Two-Dimensional Special Linear Group, *Comm. Pure Appl. Math.* 25 (1972) 635.
42. S. Katok, Fuchsian Groups, University of Chicago Press, Chicago and London, 1992.
43. J.P. Keating and N.C. Snaith, Random Matrix Theory and $\zeta(1/2+it)$, *Commun. Math. Phys.* 214 (2000) 57-89.
44. A.G. Kurosh, Lectures in General Algebra, Pergamon Press, Oxford, London, Edinburgh, New York, Paris, Frankfurt, 1965.
45. H.L. Montgomery, The pair correlation of zeros of zeta-function, *Proc. Symp. Pure Math.* (1973) 181-193.
46. M.L. Mehta, Random Matrices and the Statistical Theory of Energy levels, Academic Press, New York, 1967.
47. A.M. Odlyzko, On the Distribution of Spacing Between Zeros of Zeta Function, *Math. of Comp.* 48 (1987) 273.
48. The web site of A.M. Odlyzko: www.dtc.umn.edu/~odlyzko/.
49. A.M. Odlyzko, private communication (2003).
50. Manfred Peter, The Correlation Between Multiplicities of Closed Geodesics on the Modular Surface (2001) arXiv: math.NT/0104234.
51. B. Randol, The Length Spectrum of Riemann Surface is Always of Unbounded Multiplicity, *Proc. Amer. Math. Soc.* 78 (1980) 455.
52. C. Schmit, Quantum and Classical Properties of Some Billiards on the Hyperbolic Plane, in [28] (1989) 331-369.
53. K. Takeuchi, On some Discrete Subgroups of $\mathrm{SL}(2, \mathbb{R})$, *J. Fa. Sci. Univ. Tokyo Sect. 1A* 16 (1969) 97-100; A Characterization of Arithmetic Fuchsian Groups, *J. Math. Soc. Japan* 27 (1975) 600-612; Arithmetic Triangle Groups, *J. Math. Soc. Japan* 29 (1977) 91-106; Commensurability Classes of Arithmetic Triangles Groups, *J. Fac. Sci. Univ. Tokyo Sect. 1A* 24 (1977) 201-212.
54. A. Terras, Harmonic Analysis on Symmetric Spaces and Applications, Springer, Berlin, 1979.
55. E. C. Titchmarsh, The Theory of the Riemann Zeta-Function. Oxford, Clarendon Press, 1951.
56. M.F. Vignéras, Arithmetique des algèbres de quaternions, Lectures Note in Mathematics, 800 (1980).