Commun. Math. Phys. 176, 577–617 (1996)

Communications in
Mathematical
Physics

Distribution of Eigenvalues for the Modular Group

E. Bogomolny★, F. Leyvraz★★, C. Schmit

Division de Physique Théorique, Unité de Recherche des Universités Paris 11 et Paris 6, Associée au CNRS, Institut de Physique Nucléaire, F-91406 Orsay Cedex, France

Received: 25 November 1994

Abstract: The two-point correlation functions of energy levels for free motion on the modular domain, both with periodic and Dirichlet boundary conditions, are explicitly computed using a generalization of the Hardy–Littlewood method. It is shown that in the limit of small separations they show an uncorrelated behaviour and agree with the Poisson distribution but they have prominent number-theoretical oscillations at larger scale. The results agree well with numerical simulations.

1. Introduction

Free motion on constant negative curvature surfaces (CNCS) generated by discrete groups is the oldest and in some sense the best example of classically chaotic motion (see e.g. [1–3]). In recent years this subject has attracted wide attention also within the context of quantum chaos. There the main question is the way in which classical chaos manifests itself in the properties of the corresponding quantum systems (see e.g. [4, 5]). An important property of CNCS models is the existence of an *exact* relationship between the density of eigenvalues of the Laplace–Beltrami operator on the surface (= energy levels) and the geodesic on the surface, which correspond to classical periodic orbits. This is known as the Selberg trace formula (see e.g. [6, 7]). For arbitrary systems, only an approximate connection of this type is known, namely the so-called Gutzwiller trace formula [8, 5], which is asymptotically valid in the limit of highly excited states but does not have a good estimate on the error. It is therefore the coexistence of hard classical chaos and the exact Selberg trace formula which makes the study of CNCS models so important.

The simplest hallmark of classical chaos for a quantum system is the nature of the spectral fluctuations of the energy levels. It was conjectured that, for ergodic

★On leave of absence from L.D. Landau Institute of Theoretical Physics, 142432 Cherogolovka, Russia
★★Permanent address: Instituto de Física, University of Mexico, Apdo. postal 20-364, 01000 Mexico City, Mexico

systems with strong chaotic properties, the fluctuation properties of energy levels should be that of the classical random matrix ensembles [9]. This result has found considerable numerical confirmation, but numerical work on various models on CNCS [11–14] gave unexpected results. It was observed that the distribution of energy levels for these systems is quite close to a Poisson distribution, normally typical of integrable systems [15] and not to random matrix ensembles which are characterized by strong level repulsion.

In [16, 17], it was shown that this anomalous behaviour could be traced back to a non-generic feature of these systems, namely the fact that the underlying group belonged to a very specific subclass of the discrete subgroups of the motions of CNCS, namely the so-called arithmetic groups.

Arithmetic groups are groups which permit a representation by $n \times n$ matrices with integer entries [20]. The important consequence of the arithmetic nature of these groups is that in such cases the corresponding CNCS shows an exponential proliferation of geodesics having exactly degenerate lengths [16, 18]. It is the cumulative effect of the interference of these degenerate orbits which leads to the Poisson-like distribution of energy levels. In [16, 18] the two-point correlation function of energy levels was computed in the diagonal approximation [22], where one neglects all correlations between orbits that are not exactly degenerate. It was shown that this function definitely differs from the result predicted by random matrix theory [10]. Unfortunately, the diagonal assumption is quite crude and is only expected to give good results when the separation between two energy levels in the two-point function is large. In particular, it does not allow one to compute the correlations in the region where these are believed to be universal, namely for energy differences of the order of a level spacing.

The purpose of this paper is to compute explicitly the two-point correlation function for the energy levels of the modular domain and the corresponding billiard. The calculations are based on a generalization of the Hardy–Littlewood method [23] which was originally developed to compute the distribution of prime numbers, and depend strongly on the number-theoretical properties of the multiplicities of the periodic orbits of the modular group.

This paper is organized as follows: In Sect. 2, we give a quick review of hyperbolic geometry and the Selberg trace formula and derive various basic relations between the spectrum (in particular its two-point function, which is the basic object of interest here) and the properties of the classical periodic orbits of the system. In Sect. 3, we describe a method originally due to Hardy and Littlewood to describe the structure of singularities of certain peculiar power series. This leads to the consideration of certain quantities which are evaluated exactly. These results are presented in detail in Sect. 4. This leads to a closed form for the two-point form factor of the spectrum of the modular domain which has δ-function singularities at all rational points. In Sect. 5, we demonstrate that after a suitable smoothing it becomes constant as it should for the Poisson distribution. This result is valid when the separation is fixed and energy tends to infinity. On a scale of $\ln k/k$ (where k is the momentum) the two-point form factor has oscillations of number-theoretical origin. In Sect. 6 we generalize these formulas for the case of modular billiard with Dirichlet and Neumann boundary conditions. In the conclusion (Sect. 7) we briefly repeat the main steps necessary for our derivation. In Appendices A–F the details of calculations are presented. Much standard material in hyperbolic geometry and number theory has been presented to make the article self-contained.

2. Basic Identities

In what follows we shall use the standard realization of the surface of constant negative curvature as the so-called Poincaré plane, that is, the upper half-plane $z = x + iy$, where $y > 0$, endowed with the metric $ds^2 = y^{-2}(dx^2 + dy^2)$, (see e.g. [3, 24, 25]). The hyperbolic distance between two points $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$ is given by

$$\cosh d(z_1, z_2) = 1 + \frac{|z_1 - z_2|^2}{2y_1 y_2} . \tag{2.1}$$

The geodesics are then circles perpendicular to the real axis and the group of isometries is the group of linear fractional transformations, that is

$$z' = gz = \frac{az + b}{cz + d} , \tag{2.2}$$

where a, b, c and d are arbitrary real numbers, which can, without loss of generality, be chosen so as to satisfy the condition $ad - bc = 1$. It is then easily verified that the composition of two such transformations gives another such transformation according to the multiplication of the corresponding matrices

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} . \tag{2.3}$$

One can now introduce finite surfaces through a device which has an analogue in the Euclidean case: there, one can construct a torus by identifying all points on the plane which differ from each other by a translation belonging to a discrete subgroup of the Euclidean group (which is the isometry group of the Euclidean plane). In the case of CNCS one proceeds similarly: One takes a given discrete subgroup G of $SL(2, R)$ and identifies all points which are connected by a transformation g belonging to G.

Under certain assumptions, this procedure leads to a CNCS on which free motion is ergodic. This motion has very strong chaotic properties [1, 3, 2] and is therefore a natural object of study.

To visualize this construction, it is convenient to introduce the notion of the fundamental domain of a given discrete group G. This is defined as a region in the upper half-plane such that

1. No two points inside the fundamental domain are connected by a transformation g belonging to G.
2. For any point z' outside of the fundamental domain, there is a point z inside the fundamental domain such that there is a g in G with

$$z' = gz . \tag{2.4}$$

Equation (2.4) leads to the identification of certain points on the *boundary* of the fundamental domain. Gluing these together yields a compact surface of constant negative curvature. The trajectory on the whole upper half-plane is given by a half-circle perpendicular to the real axis. After the identification has been carried out, this geodesic can be reduced to a curve lying entirely within the fundamental domain and consisting of segments of geodesics.

Periodic orbits of this geodesic flow are in one-to-one correspondence with conjugacy classes of elements of G (see e.g. [3]). If M is any matrix belonging to G, then the (hyperbolic) length of the corresponding periodic orbit is given by the relation

$$2\cosh l/2 = |\mathrm{Tr}\, M| \,. \tag{2.5}$$

The natural "quantization" of such systems consists in the investigation of the spectrum of the invariant Laplace–Beltrami operator

$$-\frac{y^2}{2}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)\Psi_n(x,y) = E_n\Psi_n(x,y) \tag{2.6}$$

on the space of functions obeying the periodic boundary conditions

$$\Psi_n(gz) = \Psi_n(z) \tag{2.7}$$

for any element g of G. (Note that our definition of E_n differs by a factor of one half from the one commonly used in the literature). Due to their peculiar mathematical structure, there exists for these models an exact relation between "quantum" eigenvalues and the periodic orbits of the corresponding classical motion. This relation – the celebrated Selberg trace formula [6, 7] – can be stated in the following way: Let $h(r)$ be an arbitrary even function with appropriate smoothness properties and $g(u)$ its Fourier transform, given by

$$g(u) = \frac{1}{2\pi}\int_{-\infty}^{\infty} h(r)e^{-iur}dr \,. \tag{2.8}$$

Let the index p run over all classical periodic orbits, let l_p denote the length of the corresponding orbit and L_p be the length of the *primitive* periodic orbit. Let A be the area of the fundamental domain under consideration. One then has the identity:

$$\sum_{n=1}^{\infty} h(r_n) = \frac{A}{4\pi}\int_{-\infty}^{\infty} rh(\tau)\tanh \pi r\, dr + \sum_p \frac{L_p}{2\sinh l_p/2}g(l_p)$$
$$+\text{"corner and horn" terms,} \tag{2.9}$$

where r_n is equal to $\sqrt{2E_n - 1/4}$, with E_n being the eigenvalues of the Laplace–Beltrami operator with periodic boundary conditions.

The first terms correspond to the smooth part of the level density and the second gives the contribution from periodic orbits. In mathematical language, it equals the sum over all conjugacy classes of matrices with trace larger than two. The "corner and horn" terms refer to contributions from group matrices with trace less than and equal to two respectively, whenever such elements exist. Their explicit form can be found in [6, 7].

In the following, we shall apply the Selberg trace formula to

$$h(r) = \delta\left(E - \frac{1}{2}\left(r^2+\frac{1}{4}\right)\right) \,. \tag{2.10}$$

This formally causes some problems, as the delta functions are not smooth enough for the series to converge. We will here proceed in a largely formal manner and will discuss the regularization later.

In this way one obtains for the eigenvalue density $d(E) = \sum_{n=1}^{\infty}\delta(E-E_n)$,

$$d(E) = \langle d(E)\rangle + d_{osc}(E) + \tilde{d}(E) \,, \tag{2.11}$$

where the first term is given $A/(2\pi)$,

$$d_{osc}(E) = \sum_p \frac{l_p}{\pi k} \sum_{n=1}^{\infty} \frac{\cos k l_p n}{2 \sinh l_p n/2}, \tag{2.12}$$

and $\tilde{d}(E)$ includes all other terms which enter the exact Selberg trace formula. Here k is the momentum, defined by $E = k^2/2 + 1/8$.

Without the last term the Selberg trace formula agrees with the Gutzwiller trace formula. In principle, one could use a suitable regularization of the delta functions and take the appropriate limits at the end of the calculations.

So far, the above formulae are valid for an arbitrary group. Now let us consider the simplest example of an arithmetic group, namely the modular group. This group is defined as the group of all 2×2 matrices with integer entries and unit determinant modulo the subgroup $\{1, -1\}$. It is well-known (see e.g. [25]) that this group is generated by two of its elements: The translation T, which maps z to $z + 1$ and the inversion S, which maps z to $-1/z$, with their corresponding matrices:

$$t = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \qquad s = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{2.13}$$

The fundamental domain of the modular group has the form shown in Fig. 1 and the area $\pi/3$. The arrows indicate the lines identified under the action of the generators T and S and the periodic boundary conditions of Eq. (2.6) mean that the function $\Psi(x, y)$ takes the same values on corresponding lines. For this problem there is an evident reflection symmetry $x \longrightarrow -x$, which leads to a splitting of the eigenfunctions in two classes, namely odd and even. These two classes correspond to the problem of finding the eigenvalues of the Laplace–Beltrami operator with Neumann and Dirichlet conditions on the boundary of the fundamental domain shown in Fig. 1. This problem is called the modular billiard (or Artin billiard [27]) and we shall consider it in Sect. 6.

We have said that the modular group is the group of all 2×2 matrices with integer entries and unit determinant. The periodic orbits on the modular domain then correspond in a unique way to the conjugacy classes of those elements of the

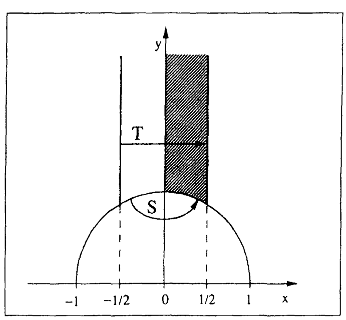

Fig. 1. The fundamental domain of the modular group. T denotes the translation $z \longrightarrow z + 1$, S is the inversion $z \longrightarrow -1/z$ and the arrows connect the boundaries identified under these transformations. The shaded region is the fundamental domain of the modular billiard

modular group which have trace larger than two (hyperbolic elements). Further one has the general relation

$$|\mathrm{Tr}\, M| = 2\cosh l_p/2\,, \tag{2.14}$$

which connects the length of the periodic orbit l_p with the trace of a representative matrix of the conjugacy class. But as all entries of the matrix are integers, the trace is also an integer. Here the arithmetical nature of the group comes into play.

Therefore one sees that for all periodic orbits with a length less than L, the number of possible different lengths is the number of different integers N less than $2\cosh L/2$, or asymptotically as $L \to \infty$, N goes as $e^{L/2}$. Now it is well known that for any group the number of periodic orbits of length less than L grows as [6, 7, 28]

$$N(l_p < L) = \frac{e^L}{L}\,. \tag{2.15}$$

These two estimates show that in the case of the modular group (as well as other arithmetic groups) periodic orbits are degenerate, i.e., there are many periodic orbits with exactly the same length.

Let $g(l)$ be the number of periodic orbits of length l. The above estimates mean that asymptotically

$$\sum_{l<L} g(l) = \frac{e^L}{L}\,,$$

$$\sum_{l<L} 1 = e^{L/2}\,, \tag{2.16}$$

where the summation extends over different lengths of periodic orbits, counted without taking multiplicity into account. If we now define the *mean* multiplicity $\langle g(l)\rangle$ as follows

$$\langle g(l)\rangle = \frac{\text{Number of periodic orbits with } l < l_p < l+\Delta l}{\text{Number of different lengths with } l < l_p < l+\Delta l}\,, \tag{2.17}$$

one concludes that

$$\langle g(l)\rangle = 2\frac{e^{l/2}}{l}\,. \tag{2.18}$$

For other arithmetic groups one obtains the same asymptotic behaviour but with a numerical prefactor which depends on the group [16, 18, 19]. This extraordinary degeneracy of the lengths of periodic orbits has been discussed before and is at the root of the remarkable structure to be found in these systems.

We now proceed to reexpress Eq. (2.12) in terms of a sum over all conjugacy classes of hyperbolic elements of the modular group. Denote by n the trace of a given conjugacy class and by $g(n)$ the number of distinct conjugacy classes corresponding to trace n. Taking into account the fact that n goes as $e^{L/2}$ as $n \to \infty$ one concludes that

$$\langle g(n)\rangle = \frac{2e^{L/2}}{L} = \frac{n}{\ln n}\,. \tag{2.19}$$

Therefore the mean multiplicity of periodic orbits having trace n grows asymptotically as the number of primes less than n. While such a fact is suggestive of a deeper connection, the authors are as yet unable to state anything further. The multiplicity $g(n)$ can also be identified with the proper class number of quadratic forms ([25, 26]), but we shall not require this representation in the following.

Applying the Selberg trace formula to $d(E)$, we split the contributions to $d_{osc}(E)$ into two parts. In the first we collect all periodic orbits whose matrices have traces less than a certain value $n_0 \gg 1$ and all others are put in the second part. The first one we simply add to $\tilde{d}(E)$ which from now will contain all "non-interesting" terms which are explicitly known and could be calculated without difficulties of principle. Neglecting the difference between 2 cosh L and e^L and further disregarding the existence of multiple traversals (since they are exponentially few in number compared to the primitive orbits) we concentrate on

$$d_{osc}(E) = \frac{2}{\pi k} \sum_{n=n_0}^{\infty} g(n) \frac{\ln n}{n} \cos(2k \ln n) . \tag{2.20}$$

(All other terms are put to $\tilde{d}(E)$.)

We have seen above that the coefficient entering in Eq. (2.20) is on the average of order one. Defining

$$\alpha(n) = g(n) \ln n/n , \tag{2.21}$$

we finally obtain that

$$d_{osc}(E) = \frac{2}{\pi k} \sum_{n=n_0}^{\infty} \alpha(n) \cos(2k \ln n) , \tag{2.22}$$

where $\langle \alpha(n) \rangle$ is one.

Equation (2.22) as it is written diverges when the sum is performed over all values of n. This is the well-known divergence of the summation over long periodic orbits, which is inherent in all semiclassical formulae. Mathematically, one treats such problems by using a suitable function $h(r)$ in the Selberg trace formula. We here use a different method of regularization based on the subtraction of the main term in the density of periodic orbits (see e.g., [22, 30, 13]). In Eq. (2.22), it corresponds to the substitution of $\alpha(n)$ by $\tilde{\alpha}(n) = \alpha(n) - 1$, i.e., the subtraction of the mean value of $\alpha(n)$[1]. Of course, one has to add to $\tilde{d}(E)$ the rest,

$$\sum_{n=n_0}^{\infty} \cos(2k \ln n) = \Re\, \zeta(2ik) - \sum_{n=1}^{n_0-1} \cos(2k \ln n) ,$$

where $\zeta(s)$ is the Riemann zeta function whose analytical continuation can be done easily by using the well-known functional relation [41].

Finally the level density of energy levels for the modular group can be written in the following form:

$$d(E) = \langle d(E) \rangle + d_{osc}(E) , \tag{2.23}$$

where

$$d_{osc}(E) = \frac{2}{\pi k} \sum_{n=n_0}^{\infty} \tilde{\alpha}(n) \cos(2k \ln n) ,$$

and for simplicity we redefine $\langle d(E) \rangle$:

$$\langle d(E) \rangle = \frac{1}{6} + \tilde{d}(E) .$$

[1] This subtraction is not necessary for billiard problems with Dirichlet boundary conditions

Usually the mean density of levels includes only the Weyl term (plus corrections). Sometimes it is convenient to put to it other terms as well. Here we shall consider all explicitly known terms and the convergent contributions from repetitions of primitive periodic orbits as part of $\langle d(E)\rangle$. They are unessential for our purposes but can be important for numerical calculations.

Having the expression for the level density, one can formally compute the n-point correlation functions:

$$R_n(\varepsilon_1,\ldots,\varepsilon_n) = \langle d(E+\varepsilon_1)d(E+\varepsilon_2)\cdots d(E+\varepsilon_n)\rangle\,. \tag{2.24}$$

The important point to note here is the energy smoothing denoted by the brackets:

$$\langle f(E)\rangle = \int f(E')\sigma(E-E')dE' \tag{2.25}$$

and $\int\sigma(x)dx = 1$. Here $\sigma(x)$ is a function which is peaked near $x=0$ and has a width ΔE. It is usually assumed that $\langle d\rangle^{-1} \ll \Delta E \ll E$, where $\langle d\rangle$ is the mean level density. The standard choice is the Gaussian:

$$\sigma(x) = \frac{1}{\sqrt{2\pi}\Delta E}\exp\left(-\frac{x^2}{2(\Delta E)^2}\right)\,,$$

but other smoothing functions are also possible. They should be chosen in such a way that if u is a constant $\langle\exp(iku)\rangle\to 0$ sufficiently fast as $k\to\infty$. This kind of averaging procedure is inevitable for the statistical analysis of a system in which there are no random parameters.

In this paper we concentrate on the two-point function

$$R_2(\varepsilon_1,\varepsilon_2) = \langle d(E+\varepsilon_1)d(E+\varepsilon_2)\rangle\,, \tag{2.26}$$

where for the modular group we use the following expressions:

$$d(E) = \langle d(E)\rangle + d_{osc}(E)\,, \tag{2.27}$$

where as $E\to\infty\langle d(E)\rangle\to 1/6$ and $d_{osc}(E)$ is given by Eq. (2.22). One then finds

$$R_2(\varepsilon_1,\varepsilon_2) = \langle d(E)\rangle^2 + \overline{R}_2(\varepsilon_1,\varepsilon_2)\,, \tag{2.28}$$

where

$$\overline{R}_2(\varepsilon_1,\varepsilon_2) = \frac{1}{(\pi k)^2}\sum_{n_1,n_2}\tilde{\alpha}(n_1)\tilde{\alpha}(n_2)\langle e^{2i(k_1\ln n_1+k_2\ln n_2)} + e^{2i(k_1\ln n_1-k_2\ln n_2)} + \text{c.c.}\rangle \tag{2.29}$$

and k_i is $\sqrt{2(E+\varepsilon_i)}$ which goes as $k+\varepsilon_i/k$ as $k\to\infty$.

Due to the energy average, the first term will be washed out and the second one will give contributions only when $n_2 = n_1 + r$ with $r \ll n_1, n_2$. Finally

$$\overline{R}_2(\varepsilon_1,\varepsilon_2) = \frac{1}{\pi^2k^2}\sum_{n=n_0}^{\infty}\sum_{r=-\infty}^{\infty}\tilde{\alpha}(n)\tilde{\alpha}(n+r)\left[\exp\left(2i\frac{kr}{n} - 2i\varepsilon\frac{\ln n}{k}\right) + \text{ c.c.}\right]\,, \tag{2.30}$$

where $\varepsilon = \varepsilon_1 - \varepsilon_2$. Note that the two-point function depends on the difference $\varepsilon = \varepsilon_2 - \varepsilon_1$ as it should be.

Let us now assume that the following mean value exists:

$$\gamma(r) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \tilde{\alpha}(n)\tilde{\alpha}(n+r)\,. \tag{2.31}$$

Then the dominant contribution to the two-point correlation function will be given by

$$\overline{R}_2(\varepsilon) = \frac{2}{\pi^2 k^2} \Re \int_{n_0}^{\infty} dn \sum_{r=-\infty}^{\infty} \gamma(r) e^{2ikr/n} \exp\left(-2i\varepsilon \frac{\ln n}{k}\right)\,, \tag{2.32}$$

where we have used a continuum approximation for n, since only large values of n make a significant contribution. If we now define $f(x)$ as follows:

$$f(x) = \sum_{r=-\infty}^{\infty} \gamma(r) e^{irx}\,, \tag{2.33}$$

we can finally express the two-point function as

$$\overline{R}_2(\varepsilon) = \frac{1}{\pi^2 k} \int_{\tau_0}^{\infty} e^{ku/2} f(2ke^{-ku/2}) \cos \varepsilon u \, du\,, \tag{2.34}$$

where $\tau_0 = 2 \ln n_0 / k$.

Another frequently used quantity is its Fourier transform $K(t)$, (the two-point form factor)

$$K(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \overline{R}_2(\varepsilon) e^{i\varepsilon t} dt = \frac{1}{2\pi^2 k} e^{kt/2} f(2ke^{-kt/2}) = \frac{1}{\pi^2 w} f(w)\,, \tag{2.35}$$

where w is $2ke^{-kt/2}$ and $K(t) = 0$ when $t < \tau_0$. Therefore all the non-trivial information is contained in the functions $\gamma(r)$ and $f(x)$. In the following section, we will outline a general method for evaluating them. The simplest approximation, known as the diagonal approximation [22], would be to assume that the $\alpha(n)$ are essentially uncorrelated, that is, that $\gamma(r)$ is zero for $r \neq 0$. This gives for $f(x)$ a constant value, which leads to an exponential growth of $K(t)$ as is clear from Eq. (2.35) [16, 18]. On the other hand, from a general consideration [22] (see Sect. 5) one sees that $K(t)$ must saturate to a constant value for $t \to \infty$ if it was originally obtained from a discrete spectrum. The exact expression for $f(x)$ found in the next section will give a resolution of this discrepancy, as will be seen in great detail in the final sections.

As a final remark, we note that Eq. (2.35) indicates that the modular domain in fact behaves much as an ordinary integrable system, in spite of its chaotic classical behaviour. Indeed, as we shall see, $K(t)$ reaches the constant value of the order of one at a time $t^\star \approx \ln k/k$ which goes to zero as $k \to \infty$. This means that in the region where the two-point correlation function is expected to be universal, namely for distances of the order of a mean level spacing, and hence for t not excessively small, the eigenvalues of the modular domain do not show correlation. On the other hand, there is structure present at small times, due to the cumulative effect of short periodic orbits.

3. Two-point Correlation Function of Multiplicities

In this section we shall give a way of evaluating the quantities $\gamma(r)$ and $f(x)$ defined in the previous section. The first remark concerning the correlations $\gamma(r)$ is that they have undamped oscillations related to their number-theoretical nature. This recalls to some extent the correlation between prime numbers which show similar oscillations, as shown by the Hardy–Littlewood conjecture [23]. We shall therefore follow a similar line of approach to evaluate $\gamma(r)$.

The key point is to find a suitable expression for the multiplicity of periodic orbits of the modular group. In the previous section we introduced the normalized multiplicities $\alpha(n)$ equal to $g(n)\ \ln/n$ and it was shown that $\langle\alpha(n)\rangle$ is equal to one. In Fig. 2 we present the correlations of the multiplicities $\alpha(n)$, that is, we show $\gamma(r)$. What is striking about that graph is that it shows correlations which do not decay as $r \to \infty$. This is in a sense similar to the situation prevailing for the correlations between primes, as shown by the Hardy–Littlewood conjecture. We shall therefore follow a somewhat similar path. We define

$$\begin{aligned}\alpha(q;r) &= \lim_{N\to\infty} \frac{1}{N}\sum_{m=0}^{N-1} \alpha(mq+r)\\ &= \lim_{u\to 0}(qu)\sum_{m=0}^{\infty} \alpha(mq+r)e^{-(mq+r)u}\,,\end{aligned} \tag{3.1}$$

where the equality between both is guaranteed by a Tauberian theorem [32, 33]. The intuitive meaning of this quantity is the average value of $\alpha(n)$ when n only runs over numbers of the form $mq+r$ for given q and r. Clearly, since $\langle\alpha(n)\rangle$ is one,

$$\sum_{r=0}^{q-1} \alpha(q;r) = q\,. \tag{3.2}$$

Now, if $\alpha(n)$ were a smooth function, we would expect all $\alpha(q;r)$ to be equal to each other and hence to one. This, as we shall see, is not at all the case. Rather, the dependence of $\alpha(q;r)$ on its arguments is exceedingly complex and highly irregular. This will in fact be one of the principal technical difficulties of this subject.

In Appendix A, we shall show the exact way to compute $\alpha(q;r)$. To state the result, we need a few definitions: We call M_q the set of 2×2 matrices with entries

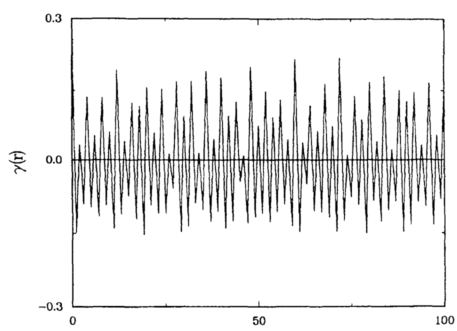

Fig. 2. Two-point correlation function of the multiplicities of the periodic orbits of the modular group computed from the knowledge of whose corresponding matrices have traces up to 8000

being integers modulo q and having determinant one modulo q. These matrices form a group under multiplication modulo q and is sometimes called the modulary group [42]. Additionally, we define $M_{q,r}$ to be the set of elements of M_q with trace equal to r modulo q. We generally denote the number of elements of a set M by $|M|$. The result of Appendix A can then be stated as follows:

$$\alpha(q;r) = \frac{q|M_{q,r}|}{|M_q|} . \tag{3.3}$$

In Table 1 we present the calculated values of $\alpha(q;r)$ for $q \leqq 11$. These results are in fact in very good agreement with direct numerical computations of $\alpha(q;r)$. The intuitive meaning of this apparently strange 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, Eq. (3.3) appears reasonable. The argument presented in Appendix A makes these ideas more rigorous.

Now, in order to compute the correlation $\gamma(r)$ between the $\alpha(n)$ we proceed as follows: Define as in [23],

$$\Phi(z) = \sum_{n=0}^{\infty} \alpha(n) z^n . \tag{3.4}$$

Since $\langle \alpha(n) \rangle$ is one, the convergence radius of this series is equal to one. The importance of this function comes from the fact that

$$J_r(e^{-u}) = e^{ru} \int_0^{2\pi} \frac{d\phi}{2\pi} \Phi^*(e^{-u+i\phi}) \Phi(e^{-u-i\phi}) e^{-ir\phi} = \sum_{n=1}^{\infty} \alpha(n)\alpha(n+r) e^{-2\pi u} , \tag{3.5}$$

and the right-hand side, again by a Tauberian theorem [32, 33], is connected to the quantity $\gamma(r)$ which we wish to obtain.

The essence of the Hardy–Littlewood approach [23] is the investigation of the function $\Phi(z)$ as $z = \exp(-u + i\varepsilon + 2\pi i p/q)$ as $u \to 0$ and $\varepsilon \to 0$, where p and q are coprime 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 $\Phi(z)$ in Eq. (3.4) 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 $\alpha(q;r)$. We present a more detailed

Table 1.

q	r modulo q										
	0	1	2	3	4	5	6	7	8	9	10
2	4/3	2/3									
3	3/4	9/8	9/8								
4	1	2/3	5/3	2/3							
5	5/4	5/6	25/24	25/24	5/6						
6	1	3/4	3/2	1/2	3/2	3/4					
7	7/8	7/6	49/48	7/8	7/8	49/48	7/6				
8	1	2/3	5/3	2/3	1	2/3	5/3	7/2			
9	3/4	1	11/8	3/4	1	1	3/4	11/8	1		
10	5/3	5/9	25/18	25/36	10/9	5/6	10/9	25/36	25/18	5/9	
11	11/12	11/12	121/120	11/10	11/10	11/12	11/12	11/10	11/10	121/120	11/12

discussion of the validity of this assumption in Appendix B. Here it may be sufficient to say that the basic assumption involved is the one that there are *no* ordinary short-range correlations involved. Rather, all correlations have the oscillatory long-range behaviour shown in Fig. 2. Accepting this, one has that as $u \to 0$ and $\varepsilon \to 0$,

$$\Phi\left(\exp\left(-u+\frac{2\pi ip}{q}+i\varepsilon\right)\right)=\sum_{r=0}^{q-1}\sum_{m=0}^{\infty}\alpha(mq+r)e^{-(u-i\varepsilon)}e^{2\pi ipr/q}$$

$$=\sum_{r=0}^{q-1}\alpha(q;r)e^{2\pi ipr/q}\frac{1}{q}\int_0^{\infty}dn\,e^{-(u-i\varepsilon)n}$$

$$=\frac{\beta(p,q)}{u-i\varepsilon}, \tag{3.6}$$

where

$$\beta(p,q)=q^{-1}\sum_{r=0}^{q-1}\alpha(q;r)\exp\left(\frac{2\pi ip}{q}r\right). \tag{3.7}$$

Therefore the function $\Phi(z)$ has a pole singularity at all points on the unit circle which have a rational multiple of 2π as phase. Its explicit form is given in Eq. (3.6). The reason for the appearance of such poles is the irregular behaviour of $\alpha(n)$ and, connected with it, the fact that the $\alpha(q;r)$ depend in a highly non-trivial way on q and r instead of being independent of them, as would be the case for more "reasonable" sequences.

The next step is to substitute Eq. (3.6) into Eq. (3.5). We therefore divide the unit circle in intervals $I_{p,q}$ centered around $\exp(2\pi ip/q)$, where p and q run over all the relatively prime numbers with $p<q$ and q less than some prescribed upper bound Q which later goes to infinity. If one now divides the integral in this way, neglects all terms in each interval except the pole terms of Eq. (3.6), and finally extends the integration over ε to the whole line, one obtains:

$$J_r(e^{-u})=e^{ru}\sum_{(p,q)=1}\int_{-\infty}^{\infty}\frac{d\varepsilon}{2\pi}\frac{|\beta(p,q)|^2}{u^2+\varepsilon^2}e^{ir(2\pi p/q+\varepsilon)}$$

$$=\frac{1}{2u}\sum_{(p,q)=1}|\beta(p,q)|^2\exp\left(\frac{2\pi ip}{q}r\right) \tag{3.8}$$

as $u\to 0$. Here and in the following, (p,q) will denote the greatest common divisor of p and q. Finally, from the definition of $J_r(e^{-u})$ in Eq. (3.5) one obtains for $\gamma(r)$

$$\gamma(r)=\sum_{(p,q)=1}|\beta(p,q)|^2\exp\left(\frac{2\pi ip}{q}r\right). \tag{3.9}$$

The sum is performed over all q, all $0<p<q$ and coprime to q, and the term $p=0$ and $q=1$ is omitted as we defined $\gamma(r)$ through $\tilde{\alpha}(n)$ (see 2.30) whose mean value is zero.

This is the two-point correlation function of the multiplicities of the periodic orbits for the modular group. All other quantities of interest can be obtained from it. In particular the function $f(x)$ introduced in the previous section is given by

$$f(x)=2\pi\sum_{(p,q)=1}|\beta(p,q)|^2\delta(x-2\pi p/q), \tag{3.10}$$

where the summation is over all p and q coprime, without the restriction $p<q$.

The subtraction needed to make the trace formula converge (see Sect. 2) is equivalent to setting $f(0)$ equal to zero which is equivalent to removing from the sum in Eq. (3.10) those terms for which $p = mq$, where m is an integer. From now on we assume implicitly that these terms are removed from the sum and the "renormalized" $f(x)$ satisfies $f(0) = 0$. As we saw above, the knowledge of $f(x)$ determines immediately the function $K(t)$ and consequently its Fourier transform, which is the two-point function.

This is in essence the main result of our paper. Combined with a fairly technical evaluation of $\beta(p,q)$ which is carried out in Appendices C and D, this gives a closed form for the Fourier transform of the two-point function which is presumably exact. Note that in these evaluations there are no (or merely trivial) approximations. The substantial problems arise from the approximations involved in the use of the Hardy–Littlewood method as well as in the various simplifications of the Selberg trace formula.

4. Results

The fundamental equation worked out at the end of the preceding section gives an exact form for the two-point function as well as its Fourier transform, but they are still somewhat unwieldy. To simplify them and cast them in a useful form will be the purpose of this section.

To this end we need to point out a basic number-theoretical property of the functions we have been discussing: they are all so-called *multiplicative* functions. One says that the function $g(n)$ is multiplicative when it has the following property:

$$g(mn) = g(m)g(n) \quad \text{whenever } (m,n) = 1 . \tag{4.1}$$

Multiplicative functions are therefore uniquely determined by their values on numbers of the form p^a, where p is a prime number. It turns out that $\alpha(q;r)$ is multiplicative in the argument q and the functions $\beta(p,q)$ as well. The first follows from the fact that $M_{q,r}$ and M_q can be expressed as the number of solutions of a given set of congruences modulo q. These are therefore multiplicative in q as a consequence of the following well-known fact known as the Chinese Remainder Theorem (see e.g. [35]): If q_1 and q_2 are relatively prime, then to every solution of a congruence or set of congruences modulo q_1q_2 there corresponds uniquely a pair of solutions to the same congruences mod q_1 and q_2 respectively and vice versa. The fact that more complicated expressions such as $\beta(p,q)$ maintain this multiplicative property is a tedious but straightforward exercise.

To simplify the expression for $f(x)$, we will use the following identity valid for an arbitrary multiplicative function $g(n)$:

$$\sum_{n=1}^{\infty} g(n) = \prod_{p \text{ prime}} \left(1 + \sum_{k=1}^{\infty} g(p^k)\right) , \tag{4.2}$$

which is known as Euler's identity. This leads to

$$\gamma(r) = \sum_{n=1}^{\infty} A_r(n) - 1 = \prod_{p} \left(1 + \sum_{k=0}^{\infty} A_r(p^k)\right) - 1 , \tag{4.3}$$

where $A_r(q)$ is given by

$$A_r(q) = \sum_{p:(p,q)=1} |\beta(p,q)|^2 \exp\left(\frac{2\pi i p}{q} r\right) . \tag{4.4}$$

To give a closed expression for $A_r(q)$ we still need one standard definition from number theory: we define the Legendre symbol of a and q, where q is an odd prime as follows:

$$\begin{aligned} \left(\frac{a}{q}\right) &= 1 \quad \text{if there is an } x \not\equiv 0 \pmod q \text{ such that } a \equiv x^2 \pmod q, \\ &= 0 \quad \text{if } a \equiv 0 \pmod q \\ &= -1 \quad \text{otherwise}. \end{aligned} \tag{4.5}$$

The meaning of this number is perhaps best understood by saying that the number of *distinct* solutions of the equation $x^2 \equiv a \pmod q$ is $1 + (a/q)$. The properties of the Legendre symbol are stated in any standard reference on number theory (see e.g. [35, 34]). A fairly tedious evaluation of $A_r(q)$ (performed in Appendix D) gives the following for q a prime power (which is all that is necessary, since $A_r(q)$ is also multiplicative in q): First, let q be equal to p^n, where p is an odd prime. Then we have for $n = 1$:

$$A_r(p) = \frac{1}{(p^2-1)^2}\left[p\sum_{x=0}^{p-1}\left(\frac{(x^2-4)((x+r)^2-4)}{p}\right) - 1\right] . \tag{4.6}$$

For $n \geqq 2$, we have, letting t be an arbitrary non-zero number modulo p

$$A_r(p^n) = \frac{1}{p^{2n}(1-p^{-2})}\begin{cases} 2(1-1/p) & r \equiv 0 \pmod{p^n} \\ -2/p & r \equiv tp^{n-1} \pmod{p^n} \\ \varepsilon(n,p)(1-1/p) & r \equiv \pm 4 \pmod{p^n} \\ -\varepsilon(n,p)/p & r \equiv \pm 4 + tp^{n-1} \pmod{p^n} \end{cases} , \tag{4.7}$$

where $\varepsilon(n, p)$ 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 must list down individual cases for low powers and eventually state a general rule:

$$A_r(2) = \begin{cases} 1/9 & r \equiv 0 \pmod 2 \\ -1/9 & r \equiv 1 \pmod 2 \end{cases} , \tag{4.8}$$

$$A_r(4) = \begin{cases} 1/18 & r \equiv 0 \pmod 4 \\ -1/18 & r \equiv 2 \pmod 4 \end{cases} , \tag{4.9}$$

$$A_r(8) = 0 , \tag{4.10}$$

$$A_r(16) = 1/(9 \cdot 16)\begin{cases} 1 & r \equiv 0 \pmod{16} \\ -1 & r \equiv 8 \pmod{16} \end{cases} , \tag{4.11}$$

$$A_r(32) = 0 , \tag{4.12}$$

and finally, for the general case $n \geqq 6$

$$A_r(2^n) = 1/(9 \cdot 2^{2n-4}) \begin{cases} 2 & r \equiv 0 \,(\mathrm{mod}\, 2^n) \\ -2 & r \equiv 2^{n-1} \,(\mathrm{mod}\, 2^n) \\ 1 & r \equiv \pm(4 + 2^{n-2}) \,(\mathrm{mod}\, 2^n) \\ -1 & r \equiv \pm(4 + 2^{n-2} + 2^{n-1}) \,(\mathrm{mod}\, 2^n) \end{cases} \tag{4.13}$$

All the terms not explicitly shown above are of course equal to zero.

5. Two-point Form Factor of Energy Levels

We have shown in Sect. 2 that the Fourier transform of the two-point correlation function of the energy levels for the modular group can be written as follows:

$$K(t) = \frac{1}{\pi^2 w} f(w) , \tag{5.1}$$

where $w = 2ke^{-k|t|/2}$ and

$$f(x) = 2(\pi) \sum_{(p,q)=1} |\beta(p,q)|^2 \delta(x - 2\pi p/q) , \tag{5.2}$$

and $\beta(p,q)$ can be expressed in terms of Kloosterman sums as in Eq. (C.8).

The two-point correlation function of any system with a non-degenerate discrete spectrum should have the following asymptotic behaviour [22]:

$$R_2(\varepsilon_1, \varepsilon_2) \to \langle d(E) \rangle \delta(\varepsilon_2 - \varepsilon_1) \quad \text{as } \varepsilon_2 - \varepsilon_1 \to 0 . \tag{5.3}$$

This is a simple consequence of the fact that, in the absence of systematic level degeneracies in the sum

$$R_2(\varepsilon_1, \varepsilon_2) = \sum_{n_1, n_2} \delta(E - E_{n_1} + \varepsilon_1) \delta(E - E_{n_2} + \varepsilon_2) \tag{5.4}$$

only the terms with $E_{n_1} = E_{n_2}$ are important in the limit $\varepsilon_2 \to \varepsilon_1$. From this the asymptotic relation of Eq. (5.3) follows. Consequently the two-particle form factor $K(t)$ should have the following large t-asymptotics:

$$K(t) \to \frac{1}{2\pi} \langle d(E) \rangle \quad \text{as } t \to \infty . \tag{5.5}$$

(For the modular domain $\langle d(E) \rangle \to 1/6$ as $E \to \infty$.) For the Poisson distribution, the form factor $K(t)$ takes its asymptotic value throughout the range of t of order one (there is a non-universal region near $t = 0$). In the case of the standard matrix ensembles (GOE, GUE and GSE) these form factors reach their asymptotic value when $t \gg 1$.

Although the above limiting behaviour follows from quite general considerations, it is generally not possible to derive it from the semiclassical formulae. Indeed, this fact depends on interference between non-degenerate long periodic orbits and is therefore quite difficult to ascertain. In the case of the modular domain, our knowledge is by now sufficiently detailed that we can attempt an explicit verification of this behaviour. Of course, since the Selberg trace formula is exact and the spectrum of the modular domain is discrete, there is every reason to expect that it will work. Nevertheless, such a calculation provides an extremely powerful consistency check on the various approximations made in order to arrive at the final result. This is what we shall do in this section.

The "exact" form factor as given in Eq. (2.34) is really a sum of delta functions at the points

$$t_{p,q} = \frac{2}{k} \ln \frac{kq}{\pi p} . \tag{5.6}$$

Equation (2.34) can then be rewritten as

$$K(t) = \frac{1}{\pi^3 k} \sum_{(p,q)=1} \left| \frac{q}{p} \beta(p,q) \right|^2 \delta(t - t_{p,q}) . \tag{5.7}$$

The limit $t \to \infty$ then corresponds to the behaviour of $f(x)$ as $x \to 0$.

In Fig. 3a we present a plot of $f(x)$ computed from the periodic orbits of the modular domain using a direct Fourier transformation of the computed $\gamma(r)$ with an appropriate smoothing procedure

$$\tilde{f}(x) = \frac{1}{N} \sum_{k=0}^{N} \gamma(k) e^{-\lambda k^2} \cos kx , \tag{5.8}$$

where the smoothing parameter λ has been taken equal to $3/N^2$ and $N = 1000$. The big peaks corresponding to small values of p and q are well pronounced, but its behaviour between peaks is not clear. It can be visualized if instead of $f(x)$ one computes its integral

$$G(x) = \int_0^x f(y)\, dy . \tag{5.9}$$

The plot of $G(x)$ is given in Fig. 3b. These pictures suggest that it is convenient to separate the function $f(x)$ into two parts

$$f(x) = f_N^{osc}(x) + f_N(x) , \tag{5.10}$$

where the first function when $0 < x < 2\pi$ is given by Eq. (5.1) but with finite number of terms for which $q \leqq N$

$$f_N^{osc}(x) = 2\pi \sum_{\substack{(p,q)=1 \\ 0<p<q\leqq N}} |\beta(p,q)|^2 \delta(x - 2\pi p/q) , \tag{5.11}$$

and the second function $f_N(x)$ includes contributions from all terms with $q > N$,

$$f_N(x) = 2\pi \sum_{\substack{(p,q)=1 \\ q>N}} |\beta(p,q)|^2 \delta(x - 2\pi p/q) . \tag{5.12}$$

The calculation of the first function is straightforward and only the second one needs special attention.

Of course, one can simply calculate $f_N^{osc}(x)$ for sufficiently large N but we shall show that adding to it an approximate expression for $f_N(x)$ leads to much better approximation even for small N.

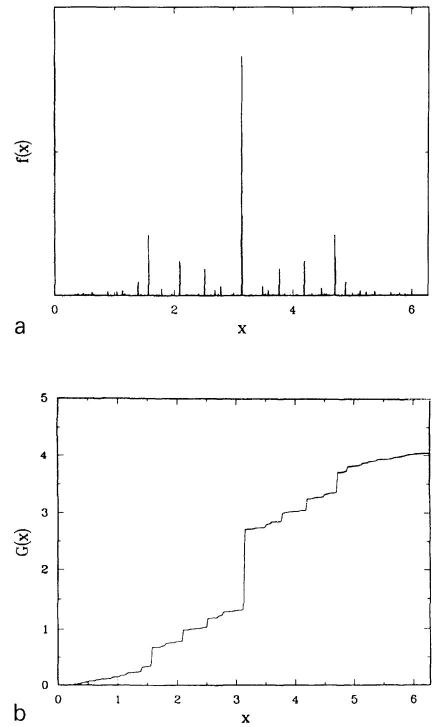

Fig. 3a. The Fourier transform of the two-point correlation function of the multiplicities of the periodic orbits for the modular domain **b.** Its integral

It is shown in Appendix C that $|\beta(p,q)| \to 0$ as $q \to \infty$. Therefore all peaks in $f_N(x)$ are small and it is reasonable to compute its mean value. We shall proceed in the following way (see e.g. [35]). Let us define the function

$$g_N(x) = \int_0^x f_N(y)\, dy\,. \tag{5.13}$$

If we now find a smooth function $G_N(x)$ such that

$$|G_N(x) - g_N(x)| \ll G_N(x)\,, \tag{5.14}$$

then we shall say that $g_N(x)$ and $G_N(x)$ are of the same order. We may then define

$$\langle f_N(x) \rangle = \frac{dG_N(x)}{dx}\,. \tag{5.15}$$

Note that this method is equivalent to smoothing the form factor over a small energy interval as in Eq. (2.25).

Formally:

$$g_N(x) = 2\pi \sum_{\substack{(p,q)=1 \\ p/q < x/(2\pi) \\ q > N}} |\beta(p,q)|^2\,, \tag{5.16}$$

where the summation is over all coprime integers such that

$$0 < p/q < \frac{x}{2\pi}, \quad q > N\,.$$

When q is fixed, $\beta(p,q)$ is a number-theoretical function of p. It is in fact quite erratic. Therefore we conjecture that it can to some degree of approximation be replaced by its mean value over all p, which we call $\beta(q)$:

$$\beta(q) = \langle|\beta(p,q)|^2\rangle = \frac{\sum_{(p,q)=1} |\beta(p,q)|^2}{\sum_{(p,q)=1} 1} , \tag{5.17}$$

where the summation is performed over all p coprime to and less than q. But it is easy to see that

$$\sum_{(p,q)=1} |\beta(p,q)|^2 = A_0(q) = \prod_{\iota} A_0(\omega_\iota^{n_\iota}) , \tag{5.18}$$

where the ω_ι are the prime factors of q and n_ι the power with which they occur. These values have already been computed in Appendix D. It is worth pointing out that in the case of an odd prime one can reduce the intractable expression of Eq. (4.6) to the following:

$$A_0(p) = \frac{p^2 - 2p - 1}{p^4(1 - p^{-2})^2} . \tag{5.19}$$

Further, since one has, by elementary number theory [35, 34], the following facts about the Euler function

$$\varphi(q) = \sum_{(p,q)=1} 1 = q \prod_{\omega|q} \left(1 - \frac{1}{\omega}\right) , \tag{5.20}$$

it follows that $\beta(q)$ is known explicitly (see Appendix F). Let

$$G_N(x) = 2\pi \sum_{\substack{(p,q)=1 \\ 0<p/q<x/(2\pi) \\ q>N}} \beta(q) . \tag{5.21}$$

The calculation of this function is performed in Appendix F. Here for clarity we consider a simple prototype of such function.

Let

$$j_N(x) = \sum_{\substack{0<p/q<x \\ q>N}} \frac{1}{q^3} . \tag{5.22}$$

It differs from the exact function (5.20) in two things. First, the sum includes all p and not only coprime to q. Second, the function $\beta(q)$ is substituted by its asymptotic behaviour. Now

$$j_N(x) = \sum_{q=N+1}^{\infty} \frac{[xq]}{q^3} , \tag{5.23}$$

where $[y]$ is an integer part of y.

When $Nx \ll 1$ and $N \to \infty$ one can replace the sum by an integral and one obtains

$$j_N(x) \approx x^2 \int_0^\infty \frac{dq[q]}{q^3} = \frac{x^2}{2} \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{12} x^2 . \tag{5.24}$$

When $Nx \gg 1$, one has $[xq] \approx xq$ and

$$j_N(x) \approx x \sum_{q=N+1}^{\infty} \frac{1}{q^2} = \frac{x}{N} . \tag{5.25}$$

Therefore at small $x \ll 1/N$ $j_N(x)$ behaves as $\pi^2 x^2/12$ but at $x \gg 1/N (x < 2\pi)$ it grows as x/N. The exact function (5.20) has similar behaviour but the computations are more complicated. The details are presented in Appendix F.

When $x \ll 1/N$,

$$G_N(x) = \frac{\pi}{24} x^2 . \tag{5.26}$$

Therefore one has

$$\langle f_N(x) \rangle = \frac{\pi x}{12} \quad (x \to 0) . \tag{5.27}$$

When $x \gg 1/N$ and $(2\pi - x) \gg 1/N$ the asymptotics of $G_N(x)$ changes. In this region

$$G_N(x) \to \frac{C}{N} x , \tag{5.28}$$

where

$$C = \prod_p \left(1 - \frac{1}{p(p+1)} \right)$$

and the product is taken over all primes including $p = 2$. Numerically $C \approx .704$.

Combining these values one concludes that the function $\langle f_N(x) \rangle$ as $0 < x < 2\pi$ approximately has the shape as in Fig. 4 and continues periodically beyond this interval.

In our approximation the function $f(x)$ can be written in the simple form:

$$f(x) = f_N^{osc}(x) + \langle f_N(x) \rangle .$$

Together with the explicit expression for $f_N^{osc}(x)$ in Eq. (5.10) it gives a quite accurate description of the two-point form factor for the modular domain.

In Fig. 5 we presented the difference between the "exact" function $G(x)$ computed by taking into account all terms with $q \leqq 1000$ and the sum of these two terms for a different value of N. Note the difference in scales with respect to Fig. 3. Quite good agreement is observed even for $N = 20$.

The knowledge of the function $\langle f_N(x) \rangle$ is particularly important in the region of small x because it gives the dominant contribution to the asymptotics of the two-point correlation formfactor. From Eqs. (5.26) and (5.27) it follows that as $x \to 0$,

$$f(x) \to \frac{\pi}{12} x \quad \text{and} \quad G(x) \to \frac{\pi}{24} x^2 . \tag{5.29}$$

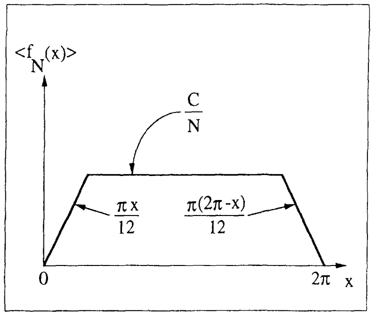

Fig. 4 The schematic picture of the function $\langle f_N(x) \rangle$ for $0 < x < 2\pi$

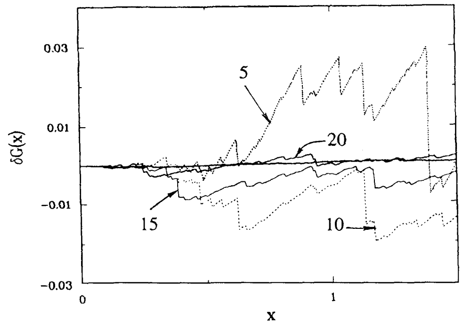

Fig. 5. The difference between the exact $G(x)$ and the approximate formula (5.28) for different values of N which are indicated near the curves. The middle line corresponds to $N = 50$

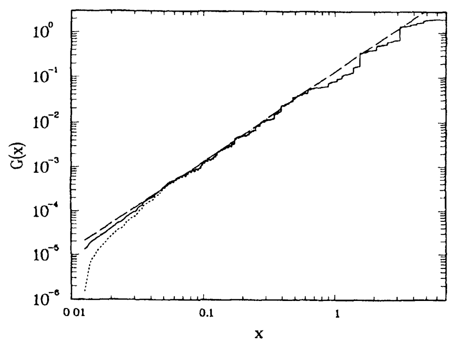

Fig. 6. Behaviour of the integral of the two-point correlation form factor at small x. The dotted line corresponds to the sum of all terms up to $N = 500$ and the solid one to $N = 1000$. The straight line indicates asymptotics (5.26)

In Fig. 6 we plot the function $G(x)$ computed as the sum of all terms up to $N = 500$ and $N = 1000$ in the double logarithmic scale which is the most sensitive to small x behaviour. The solid line is the asymptotics (5.28).

The most important consequence is that in the limit $k \to \infty$ and t fixed the two-point formfactor tends to the constant value

$$K(t) = \frac{1}{12\pi}, \tag{5.30}$$

as it should be for the Poisson distribution.

6. Billiard Problems

We have mentioned that eigenfunctions of the Laplace–Beltrami operator for the modular group can be classified by the parity with respect to the inversion $x \to -x$. The odd (even) functions are eigenfunctions of the billiard problem with the Dirichlet (Neumann) conditions on the boundary of half of the modular domain (see Fig. 1). In this section we compute the two-point correlation functions for these problems separately.

In the modular billiard problem group matrices are 2×2 matrices with integer entries but with the determinant equals both 1 and -1 (see e.g. [3]). Matrices with determinant -1 describe geometrically an inversion with respect to a circle and they correspond to the following transformation:

$$z' = \frac{az^* + b}{cz^* + d}, \tag{6.1}$$

where z^* is the complex conjugate of z.

The periodic orbits of the billiard problems can be identified with classes of conjugated matrices (both with determinants ± 1) but their length is given by Eq. (2.4) only if the matrix determinant equals 1. If it equals -1 the length of corresponding periodic orbits should be computed by

$$2\,\sinh(l/2) = |TrM|\,. \tag{6.2}$$

Physically matrices with determinant $+1$ (-1) correspond to periodic orbits with even (odd) number of reflections from the billiard boundary.

For billiard problems there exist an exact Selberg-type trace formula which expresses the density of states through periodic orbits [39, 3]. It looks like the usual trace formula (2.12) but periodic orbits with odd number of bounces have an additional minus sign for Dirichlet boundary conditions.

Performing the same steps as in Sect. 2 one obtains:

$$d_{bil}(E) = \langle d(E)\rangle + \tilde{d}(E) + d_{osc}(E)\,, \tag{6.3}$$

where $\langle d(E)\rangle = A/2\pi$ is a smooth part of the level density. A here is the area of the modular billiard equals a half of that of the modular domain: $A = \pi/6$,

$$d_{osc}(E) = \frac{1}{\pi k}\sum_{n=n_0}^{\infty}(a^{(+)}(n) + \varepsilon a^{(-)}(n))\cos(2k\ln n)\,. \tag{6.4}$$

$a^{(\pm)}(n)$ here are the normalized multiplicities of periodic orbits corresponding to classes of conjugated matrices with determinants $+1$ and -1 correspondingly

$$a^{(\pm)}(n) = 2g^{(\pm)}(n)\frac{\ln n}{n}\,, \tag{6.5}$$

where $g^{(\pm)}(n)$ is the number of periodic orbits with determinant ± 1 and trace equals n. The factor 2 is introduced for the convenience. $\varepsilon = -1$ for the Dirichlet boundary conditions and $\varepsilon = 1$ for the Neumann ones. The function $\tilde{d}(E)$ contains all other terms.

The arithmetic nature of the modular billiard group leads to the conclusion that

$$\langle a^{(\pm)}(n)\rangle = \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} a^{(\pm)}(n) = 1\,. \tag{6.6}$$

Similarly to Appendix A one concludes that quantities

$$a^{(\pm)}(q,r) = \lim_{N\to\infty}\frac{1}{N}\sum_{m=0}^{N-1} a^{(\pm)}(mq + r) \tag{6.7}$$

are equal to

$$a^{(\pm)}(q,r) = \frac{qM^{(\pm)}_{q,r}}{|M_q|}\,, \tag{6.8}$$

where $M^{(\pm)}_{q,r}$ is the number of matrices with integer entries modulo q whose determinant equals ± 1 and whose trace equals r, $|M_q|$ is the total number of matrices modulo q with determinant 1. (The total number of matrices with determinant -1 will be the same.)

After merely rephrasing arguments of previous sections we find that the two-point correlation form factor of the modular billiard can be written in the following form:

$$K(t) = \frac{1}{4\pi^2 w}(f^{(++)}(w) + 2f^{(+-)}(w) + f^{(--)}(w))\,, \tag{6.9}$$

where $w = 2k\exp(-kt/2)$ and

$$f^{(\varepsilon_1,\varepsilon_2)}(x) = 2\pi \sum_{(p,q)=1} \beta^{(\varepsilon_1)}(p,q)\beta^{(\varepsilon_2)}(p,q)\delta\left(x - 2\pi\frac{p}{q}\right)\,, \tag{6.10}$$

where $\varepsilon = \pm$ and

$$\beta^{(\varepsilon)}(p,q) = \frac{1}{q}\sum_{r=0}^{q-1} a^{(\varepsilon)}(q,r)\exp\left(\frac{2\pi i p}{q}r\right)\,.$$

The explicit formulas for $\beta^{(-)}(p,q)$ are given in Appendix C.

To compute the average behaviour of $K(t)$ we have to know the average values of $f^{(--)}(x)$ and $f^{(+-)}(x)$ as $x \to 0$. As in the previous section one should first find the mean values of the product of two $\beta^{(\varepsilon)}$ over all values of p:

$$\beta^{(\varepsilon_1,\varepsilon_2)}(q) = \frac{1}{\varphi(q)} \sum_{p:(p,q)=1} \beta^{(\varepsilon_1)}(p,q)\beta^{(\varepsilon_2)}(p,q)\,. \tag{6.11}$$

The later sum is connected to functions $A_0^{(\varepsilon_1,\varepsilon_2)}(q)$ defined in Appendix D. If $q = \omega_1^{n_1}\omega_2^{n_2}\cdots\omega_k^{n_k}$ is the canonical representation of q as the product of different primes ω then

$$\beta^{(\varepsilon_1,\varepsilon_2)}(q) = \frac{1}{\varphi(q)}\prod_{\omega_i|q} A_0^{(\varepsilon_1,\varepsilon_2)}(\omega_i^{n_i})\,. \tag{6.12}$$

Using this value it is shown in Appendix F that as $x \to 0$,

$$\langle f^{(--)}(x)\rangle = \langle f^{(++)}(x)\rangle = \frac{\pi}{12}x, \quad \langle f^{(+-)}(x)\rangle = O(x^{3/2})\,. \tag{6.13}$$

Therefore as $t \gg \ln k/k$,

$$K(t) \to \frac{1}{24\pi}\,, \tag{6.14}$$

which coincides with the Poisson value for this quantity ($K(t) = A/(2\pi)^2$ and $A = \pi/6$).

Note that the last of relations (6.13) means that in the universal limit eigenvalues of different symmetry classes (odd–even with respect to the inversion) are uncorrelated. This property is usually taken for granted on "general considerations." But the problem does not seem to be trivial because the same periodic orbits enter the trace formulas for both odd and even states. Only their phases are different.

A priori it is unclear that the cross term will vanish. To our knowledge the modular billiard is the only dynamical system where the absence of correlation between states of different symmetry can be checked analytically. (On this subject see also [40].)

At small values of $t \sim \ln k/k$ the billiard form factor has peaks similar to the ones already found in the case of the full modular group. Note that the largest peak with $p/q = 1/2$ is absent in the billiard with the Dirichlet boundary conditions because $-1 \equiv 1 \mod 2$.

7. Concluding Remarks

In this paper we have computed the two-point correlation function for the energy levels of the modular group and the modular billiard. From the point of view of classical motion these systems are ergodic with strong chaotic properties. But their arithmetical nature leads to a very large degeneracy of lengths of periodic orbits and, as a consequence, to the fact that their two-point correlation function tends to the Poisson value typical for the integrable systems and not for chaotic ones. At large scale the two-point correlation function has prominent number-theoretical oscillations.

To clarify the (tedious) derivation we very briefly repeat the essential steps.

- The Selberg trace formula allows to write the density of states as a sum over classical periodic orbits (the function $\langle d(E)\rangle$ is explicitly known):

$$d(E) = \langle d(E)\rangle + \frac{2}{\pi k}\sum_{n=n_0}^{\infty} \tilde{\alpha}(n)\cos(2k\ln n)\,, \tag{7.1}$$

  where $\tilde{\alpha}(n) = \alpha(n) - 1$,

$$\alpha(n) = g(n)\frac{\ln n}{n}$$

  and $g(n)$ is the number of periodic orbits with trace equals n.
- The two-point correlation form factor can be expressed as follows:

$$K(t) = \frac{1}{\pi^2 w} f(w)\,, \tag{7.2}$$

  where $w = 2k\exp(-kt/2)$,

$$f(x) = \sum_{r=-\infty}^{+\infty} \gamma(r)e^{irx}\,,$$

  and

$$\gamma(r) = \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\tilde{\alpha}(n)\tilde{\alpha}(n+r)$$

  is the two-point correlation function for the multiplicities of periodic orbits.
- Using a probabilistic approach we show that

$$\alpha(q;r) = \lim_{N\to\infty}\frac{1}{N}\sum_{m=0}^{N-1}\alpha(qm+r) = \frac{q|M_{q,r}|}{|M_q|}\,, \tag{7.3}$$

  where $|M_q|$ is the total number of matrices with entries which are taken as integers modulo q and $|M_{q,r}|$ is the number of such matrices having trace equal to r modulo q.

- A generalization of the Hardy–Littlewood method permits to find an explicit expression for $\gamma(r)$:

$$\gamma(r) = \sum_{(p,q)=1} |\beta(p,q)|^2 \exp\left(2\pi i \frac{p}{q} r\right) , \tag{7.4}$$

where the summation is taken over all q and all $p < q$ coprime to q and

$$\beta(p,q) = \frac{1}{q} \sum_{r=0}^{q-1} \alpha(q;r) \exp\left(2\pi i \frac{p}{q} r\right) . \tag{7.5}$$

- Introducing the Kloosterman sums

$$S(n,m,c) = \sum_{d=0}^{c-1} \exp\left(\frac{2\pi i}{c}(nd + md^{-1})\right) ,$$

$\beta(p,q)$ can be written as follows:

$$\beta(p,q) = \frac{1}{q^2 \prod_{\omega|q}(1-\omega^{-2})} S(p,p;q) , \tag{7.6}$$

where ω are the prime divisors of q.
- These formulae give the explicit expression for the two-point correlation form factor

$$K(t) = \frac{1}{\pi^3 k} \sum_{(p,q)=1} \left| \frac{q}{p} \beta(p,q) \right|^2 \delta(t - t_{p,q}) , \tag{7.7}$$

where

$$t_{p,q} = \frac{2}{k} \ln \frac{kq}{\pi p} .$$

In the limit $k \to \infty$ and t fixed, the dominant contribution comes from terms with $p/q \ll 1$. Smoothing over such values we show that in this limit $K(t)$ has the constant Poisson value:

$$K(t) = \frac{A}{(2\pi)^2} . \tag{7.8}$$

Here $A = \pi/3$ is the area of the fundamental region of the modular group. For small t (of order of $\ln k/k$) $K(t)$ has number-theoretical oscillations due to cumulative contributions of degenerate periodic orbits. For very small values of t (of order of $1/k$) the two point form factor has δ function peaks connected with short periodic orbits.

Analogous formulas can be obtained for the modular billiards. As a byproduct we proved that the energy levels of different symmetry are uncorrelated. It seems that practically all such formulae can be generalized to other arithmetical groups. Though the modular group is by no means a generic system, it is the first ergodic dynamical system for which it is possible to compute explicitly the distribution of the energy levels.

Appendix A

Here we want to show how the expression for $\alpha(q;r)$ is derived. One notes first that the modular group is generated by two elements s and t defined as follows:

$$s = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \qquad t = \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix} . \tag{A.1}$$

Since $s^2 = t^3 = 1$, any element of the modular group can be represented as a sequence of alternating s and t^σ, where $\sigma = \pm 1$. Further, such representations are in fact unique. From this follows that to each conjugacy class there corresponds uniquely a word beginning with s and ending with a t^σ, up to cyclic permutations. That is to say, cyclically equivalent words generate identical conjugacy classes, but no other words do. To this end, it is essential, however, to enforce the condition that the word begin with an s and end with a t^σ. In the following we will therefore only consider words generated by the matrices

$$m_1 = st \qquad m_2 = st^{-1} . \tag{A.2}$$

Let us now define $\alpha(q;r)$ more precisely: Up to a normalization factor of q, it is the probability that a conjugacy class belonging to a given value of n should have a trace equal to r modulo q, after averaging over n. Here instead of averaging over n, we shall average over all conjugacy classes which require k symbols to generate and eventually consider the average over many and sufficiently large values of k. It seems highly reasonable to claim that these two averages should be equivalent. This is in fact the major assumption in this Appendix.

It is obvious that there are exactly 2^k different words generated by k symbols. Each of those words denotes a different conjugacy class, up to cyclic invariance. For the overwhelming majority of words, this simply introduces a fixed factor of k which does not matter in this discussion. Those words for which the degeneracy factor is different from k are exponentially rare and can be neglected. We can thus view the words (or conjugacy classes) as arising from the following Markov process: Starting from the identity, at each step multiply the matrix to the right randomly either by m_1 or by m_2 with probability 1/2. In this way all conjugacy classes are generated with equal probability. If one now projects this Markov process down onto the set M_q by simply considering the entries of each matrix as integers modulo q, then the Markov process described above projects to a Markov chain on a finite space, namely M_q. Under such circumstances, very general theorems ([33]) guarantee the existence of an approach to equilibrium. For this purpose, one needs to show that the process is ergodic, that is, that every element of M_q can indeed be reached by an appropriate combination of m_1 and m_2. This is seen as follows: Consider an arbitrary element of M_q as an element of the modular group. As such, it can be represented as a product of s and t^σ, though not necessarily of m_1 and m_2. To obtain the latter, note that there is a number k such that

$$(st)^k = 1 \,(\mathrm{mod}\, q) . \tag{A.3}$$

In fact k can be chosen equal to $q-1$. Inserting this representation of the identity either before or after the word, one can always bring it into the required form.

As a final point, we need to show that the equilibrium attained on M_q is indeed the uniform distribution. This is seen by noting that m_1 and m_2 are always different and invertible, so that the Markov process always connects an arbitrary matrix a with two different matrices a_1 and a_2 with probability 1/2. Thus the probability of finding matrix a after k steps satisfies the equation

$$P_k(a) = \frac{1}{2}(P_{k-1}(am_1^{-1}) + P_{k-1}(am_2^{-1})) . \tag{A.4}$$

This has the uniform solution as an equilibrium (k-independent) solution. Since in a finite space the uniqueness of equilibrium is guaranteed, one can show that uniform distribution is indeed approached. From this follows the desired claim: Indeed, since $\alpha(q;r)/q$ is identified with the probability that a word fall upon a matrix of trace r modulo q, this probability is clearly given by the ratio of the number of matrices in M_q having trace r to the total number.

There is still a slight caveat, however: It is well-known that cyclic behaviour cannot be excluded on general grounds. Indeed, odd-even oscillations are in fact observed for q equal to two. Nevertheless, it is readily seen that the effect of such oscillations dies out when one averages over k. Thus, in the absence of oscillations, it would be sufficient to take the average over one set of conjugacy classes defined by a sufficiently large value of k. In the presence of oscillations we must still average over different values of k. In fact, it could probably be shown that cyclic behaviour for cycles larger than two cannot exist. It certainly has never been observed up to now in the specific cases we have looked at. One might add that the transition matrix of the above Markov chain has some very remarkable properties: Its eigenvalues are highly degenerate and appear to lie on very specific loci of the complex plane. A great deal of this unexpected structure can be traced back to the fact that this matrix is invariant under the action of the modulary group (the modular group taken modulo q) and is related to the regular representation of the latter which is always reducible. We do not go any further into those details, however, because they are unnecessary to our immediate purpose.

Appendix B

Here we give another derivation of the Hardy–Littlewood method, which has the advantage of highlighting the approximations involved. We need to evaluate the integral on the left-hand side of Eq. (3.4) which expresses the function $\gamma(r)$. We therefore divide the unit circle in intervals $I_{p,q}$ centered around $\exp(2\pi ip/q)$, where p and q run over all relatively prime numbers with $p < q$ and q less than some prescribed upper bound Q which later goes to infinity. If one now divides the integral in this way and denotes the interval $I_{p,q}$ shifted so as to be centered around $z = 1$ by $J_{p,q}$, one obtains:

$$\begin{aligned}\gamma(r) =& \lim_{u\to 0}(2u) \sum_{(p,q)=1} \sum_{n,n'} \alpha(n)\alpha(n')e^{-(n+n')u} \\ & \int_{J_{p,q}} \frac{d\psi}{2\pi} \exp\left(\frac{2\pi ip}{q}(n-n'-r) - i(n-n'-r)\psi\right) .\end{aligned} \tag{B.1}$$

To simplify the expression, one rewrites the sums over n and n' as sums over m',r' and m'',r'' respectively, where $n = m'q + r'$ and r' is between 0 and $q-1$, and similarly for n'. In this case, the integral over $J_{p,q}$ is strongly oscillatory for

$m' + m''$ and one finds:

$$\gamma(r) = \lim_{u \to 0} \frac{2u}{q} \sum_{(p,q)=1} \sum_{m'} \sum_{r,r'=0}^{q-1} \alpha(m'q + r')\alpha(m'q + r'')$$

$$\exp(-(2m'q + r' + r'')u) e^{\frac{2\pi i p}{q}(r' - r'' - r)} . \tag{B.2}$$

At this stage we make a fairly tricky approximation: In essence, we assume that there are no other correlations between the $\alpha(n)$ than those implied by the dependence on q and r of its average over numbers which are equal to r modulo q. In that sense, we rewrite Eq. (7.2) as:

$$\gamma(r) = q^{-2} \sum_{(p,q)=1} \sum_{r,r'=0}^{q-1} \alpha(q; r')\alpha(q; r'') \exp\left(\frac{2\pi i p}{q}(r' - r'' - r)\right)$$

$$= \sum_{(p,q)=1} q^{-2} \left| \sum_{r'=0}^{q-1} \alpha(q; r') \exp\left(\frac{2\pi i p}{q} r'\right) \right|^2 \exp\left(-\frac{2\pi i p}{q} r\right) . \tag{B.3}$$

Perhaps some examples of sequences $\alpha(n)$ for which the method works, and others for which it fails may be helpful. For example, let us consider the sequence $\alpha_p(n)$ which is equal to 1 when $n \equiv 0 \pmod{2p}$ and -1 when $n \equiv p \pmod{2p}$, where p is a given prime, and zero otherwise. It is easy to check that in this case $\alpha(q; r)$ is zero unless q is a multiple of $2p$ and r a multiple of p. The complete calculation of $\gamma(r)$ for this sequence by the formulae given in the text gives complete agreement with the exact result. On the other hand, let us now randomize this sequence in the following way: The numbers $\alpha((2k+1)p)$ and $\alpha(2kp)$ are randomly interchanged with probability 1/2. The resulting sequence is clearly still correlated. In particular, $\gamma(p)$ is non-zero. Nevertheless it is seen that all $\alpha(q; r)$ of this sequence are zero. Thus the Hardy–Littlewood method clearly fails to take into account mere short-range correlations. It is seen that adding this sequence to another, say the $\alpha(n)$ we have been studying in the text, clearly modifies the correlations $\gamma(r)$ but in no way affects the $\alpha(q; r)$, so that one really must assume that there is no short-range correlation of this type present. Can this assumption be sustained for the $\alpha(n)$ we have been studying?

At first sight, this appears very strange, since we are dealing with correlations over quite short ranges. However, from Appendix A, we see that the natural variable in which to study the development of correlations is not the trace n but rather the number of symbols necessary to generate a given conjugacy class. This number, however, does not vary smoothly at all with trace and the overwhelming majority of conjugacy classes with nearby traces have quite a different number of symbols that generate them. Thus it is allowable to consider them as decorrelated as the Hardy–Littlewood method implicitly does.

Appendix C

In the following we develop some basic tools to compute expressions for $|M_q|$ and $|M_{q,r}|$. The fundamental identity we shall be using is the following easily verified fact:

$$q^{-1} \sum_{r=0}^{q-1} \exp\left(\frac{2\pi i r x}{q}\right) = \delta_{x,0} , \tag{C.1}$$

where x is an integer between 0 and $q-1$, or else an arbitrary integer taken modulo q. We therefore find, for $\beta(p,q)$:

$$\beta(p,q) = \frac{1}{q}\sum_{r=0}^{q-1}\alpha(q;r)\exp\left(\frac{2\pi ip}{q}r\right)$$
$$= \frac{1}{|M_q|}\sum_{r=0}^{q-1}|M_{q,r}|\exp\left(\frac{2\pi ip}{q}r\right) . \qquad (C.2)$$

$|M_{q,r}|$ is the number of matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ where a,b,c,d are taken modulo q such that $ad-bc=1$ and $a+d=r$ modulo q.

Therefore

$$\beta(p,q) = \frac{1}{|M_q|}\sum_{r=0}^{q-1}\sum_{abcd=0}^{q-1}\delta_{ad-bc-1,0}\delta_{a+d-r,0}\exp\left(\frac{2\pi ip}{q}r\right)$$
$$= \frac{1}{q^2|M_q|}\sum_{r=0}^{q-1}\sum_{abcd=0}^{q-1}\sum_{mm'=0}^{q-1}\exp\left(\frac{2\pi i}{q}[m(ad-bc-1)+m'(a+d-r)+pr]\right) . \qquad (C.3)$$

Note that, in the following, the Kronecker deltas must *always* be interpreted as being 1 if the two indices are equal modulo q. We keep the evaluation of $|M_q|$ for the end and perform the sums over m' and r first. This gives:

$$\beta(p,q) = \frac{1}{|M_q|}\sum_{bcd=0}^{q-1}\sum_{m=0}^{q-1}\delta_{md+p,0}\exp\left(\frac{2\pi i}{q}(dp-mbc-m)\right) . \qquad (C.4)$$

Now we use the fact that $(p,q)=1$ from which follows that the equation $md \equiv -p \,(\mathrm{mod}\, q)$ cannot hold unless $(d,q)=(m,q)=1$. Under these circumstances, on the other hand, the equation has a unique solution, namely $m \equiv -pd^{-1}\ (\mathrm{mod}\, q)$, where the inverse is to be taken modulo q. It is well-known that if $(d,q)=1$, this inverse exists and is unique. From this follows, after performing all sums:

$$\beta(p,q) = \frac{q}{|M_q|}\sum_{(d,q)=1}\exp\left(\frac{2\pi ip}{q}(d+d^{-1})\right) . \qquad (C.5)$$

Note that such a formula cannot be used to obtain, say, $|M_q|$ by setting p to zero, since the assumption that $(p,q)=1$ is essential in deriving Eq. (C.5).

This formula gives an explicit expression of the Fourier transformation of the number of matrices modulo q with determinant equals 1. For billiard problems we shall need also the same quantity but for matrices with determinant -1. Generalizing the previous arguments we obtain:

$$\beta^{(\varepsilon)}(p,q) = \frac{q}{|M_q|}\sum_{(d,q)=1}\exp\left(\frac{2\pi ip}{q}(d+\varepsilon d^{-1})\right) , \qquad (C.6)$$

where $\varepsilon=\pm 1$ corresponds to matrices with determinant ± 1.

If one introduces the so-called Kloosterman sums [36, 38],

$$S(n,m,c) = \sum_{(d,c)=1} \exp\left(\frac{2\pi i}{c}(nd + md^{-1})\right), \tag{C.7}$$

then $\beta^{(\varepsilon)}(p,q)$ can be written as

$$\beta^{(\varepsilon)}(p,q) = \frac{q}{|M_q|} S(p, \varepsilon p, q). \tag{C.8}$$

It is easy to see that $\beta^{(\varepsilon)}(p,q) \to 0$ as $q \to \infty$. More specifically, it has been shown that [37]:

$$|S(n,m,c)| < d(c)\sqrt{c}(m,n,c)^{1/2}, \tag{C.9}$$

where $d(c)$ is the number of divisors of c and (m,n,c) is the largest common divisor of m,n,q. Such estimates yield sharp bounds for the $\beta(p,q)$.

To finish, we need the value of $|M_q|$. This is found as follows: First, we can limit ourselves to the case of $q = s^n$, where s is a prime number, because $|M_q|$ is a multiplicative function of q by the Chinese Remainder Theorem. Consider first the case q equal to s. In this case, we are dealing with matrices over a field. The number of singular matrices is easily seen to be $s^3 + s^2 - s$. Indeed, there are $s^2 - 1$ different ways of choosing the first row to be a non-zero vector, and to each of those correspond s different vector for the second row, which must be taken proportional to the first. If the first row is zero, on the other hand, any choice for the second row will yield a singular matrix. Thus the number of regular matrices is found to be $(s-1)s(s^2-1)$. These are distributed uniformly over $s-1$ different non-zero values of the determinant, so that

$$|M_s| = s(s^2 - 1) = s^3(1 - s^{-2}). \tag{C.10}$$

Now consider the case $q = s^n$. Any numbers a, b, c and d satisfying $ad - bc \equiv 1 \pmod{s^n}$ will also satisfy the same equation modulo s^{n-1}. Let us then define

$$a \equiv a_1 + \alpha s^{n-1} \pmod{s^n}, \tag{C.11}$$

where a_1 is a taken modulo s^{n-1} and α is an appropriate number taken modulo s. Define similar numbers for b, c and d. One then has

$$ad - bc = a_1 d_1 - b_1 c_1 + s^{n-1}(a_1\delta + d_1\alpha - b_1\gamma - c_1\beta) \pmod{s^n}. \tag{C.12}$$

Hence, to obtain a solution modulo s^n, we need an arbitrary solution modulo s^{n-1} and a set of numbers α, β, γ and δ satisfying

$$a_1\delta + d_1\alpha - b_1\gamma - c_1\beta \equiv 0 \pmod{s}. \tag{C.13}$$

But it is impossible that a_1, b_1, c_1 and d_1 should all be zero. Therefore, let us assume, say, that $a_1 \neq 0$. Then all possible sets of values for α, β and γ yield exactly one value for δ. Therefore to each solution modulo s^{n-1} there correspond exactly s^3 solutions modulo s^n, from which follows, for $q = s^n$

$$|M_q| = s^3|M_{q/s}| = s^{3n}(1 - s^{-2}). \tag{C.14}$$

From this follows the general relation (see e.g. [21, 42])

$$|M_q| = q^3 \prod_{p|q} (1 - p^{-2}) \,. \tag{C.15}$$

Appendix D

In this Appendix, we give an expression for $A_r(q)$ which is defined by the following relation:

$$A_r(q) = \sum_{p:(p,q)=1} |\beta(p,q)|^2 \exp\left(\frac{2\pi i p}{q} r\right) \,. \tag{D.1}$$

Using the above expression for $\beta(p,q)$ one obtains:

$$A_r(q) = \frac{q^2}{|M_q|^2} a_r(q) \,, \tag{D.2}$$

where

$$a_r(q) = \sum_{p:(p,q)=1} \sum_{(d,q)=1} \sum_{(\delta,q)=1} \exp\left(\frac{2\pi i p}{q}(d + d^{-1} - \delta - \delta^{-1} - r)\right) \,. \tag{D.3}$$

Now we can limit ourselves to evaluating $a_r(q)$ for q equal to s^n and s a prime, since $a_r(q)$ is in fact a multiplicative function of q. Let us first consider the case where q is equal to s and s is odd. In this case we have:

$$\begin{aligned} a_r(s) &= \sum_{p=1}^{s-1} \sum_{d=1}^{s-1} \sum_{\delta=1}^{s-1} \exp\left(\frac{2\pi i p}{s}(d + d^{-1} - \delta - \delta^{-1} - r)\right) \\ &= \sum_{p=0}^{s-1} \sum_{d=1}^{s-1} \sum_{\delta=1}^{s-1} \exp\left(\frac{2\pi i p}{s}(d + d^{-1} - \delta - \delta^{-1} - r)\right) - (s-1)^2 \,. \end{aligned} \tag{D.4}$$

Now one notices that the sum appearing in the last equation of Eq. (D.4) denotes, up to a factor of s, the number of solutions of the equation

$$d + d^{-1} - \delta - \delta^{-1} \equiv r \,(\mathrm{mod}\, s) \,. \tag{D.5}$$

One sees further that x can be represented in the form $d + d^{-1}$ modulo s if and only if $x^2 - 4$ can be represented as a square. This follows from the fact that the equation

$$d + d^{-1} \equiv x \,(\mathrm{mod}\, s) \tag{D.6}$$

has the solution

$$d \equiv \frac{1}{2}(x \pm \sqrt{x^2 - 4}) \,(\mathrm{mod}\, s) \,. \tag{D.7}$$

Here again 1/2 must be taken modulo s, but since s is an odd prime this is always possible. Further, we see that the number of different representations of x as $d + d^{-1}$

is $1 + ((x^2 - 4)/q)$, where we are using the Legendre symbol already defined in the text. From this we obtain the following representation of $a_r(s)$:

$$\begin{aligned}
a_r(s) &= s\sum_{x=0}^{s-1}\left(1+\left(\frac{(x^2-4)}{s}\right)\right)\left(1+\left(\frac{((x+r)^2-4)}{s}\right)\right)-(s-1)^2\\
&= s(s-2)+s\sum_{x=0}^{s-1}\left(\frac{(x^2-4)((x+r)^2-4)}{s}\right)-(s-1)^2\\
&= s\sum_{x=0}^{s-1}\left(\frac{(x^2-4)((x+r)^2-4)}{s}\right)-1\,,
\end{aligned} \tag{D.8}$$

where we have made use of the following standard identity

$$\sum_{x=0}^{s-1}\left(\frac{x(x-a)}{s}\right) = -1\,, \tag{D.9}$$

which we prove for completeness in Appendix E which gives some elementary properties of the Legendre symbol. For all practical purposes, this representation is sufficiently explicit, and no more specific one could be found for general values of r. For $r = 0$ and $r = \pm 4$, however, closed expressions can indeed be found: We shall only require the case r equal to zero. In this case the remaining sum involving Legendre symbols becomes $s - 2$, since these are identically equal to one except for $x = \pm 2$. From this we finally obtain for s an odd prime

$$A_0(s) = \frac{s^2-2s-1}{s^4(1-s^{-2})^2}\,. \tag{D.10}$$

Now let us consider the case of q equal to s^n. In this case, we reexpress Eq. (D.3) in the following way: Let $B_r(q)$ be the number of roots of the equation

$$d + d^{-1} - \delta - \delta^{-1} \equiv r\,(\mathrm{mod}\,q)\,. \tag{D.11}$$

In other words:

$$B_r(q) = q^{-1}\sum_{p=0}^{q-1}\sum_{(d,q)=1}\sum_{(\delta,q)=1}\exp\left(\frac{2\pi ip}{q}(d+d^{-1}-\delta-\delta^{-1}-r)\right)\,. \tag{D.12}$$

$a_r(s^n)$ was defined as a sum over p coprime to s^n. It is evident that it is equivalent to the sum over all p minus sum over p divisible on s. From this follows that we can express $a_r(q)$ in terms of $B_r(q)$:

$$\begin{aligned}
a_r(s^n) &= s^nB_r(s^n) - s^2s^{n-1}B_r(s^{n-1})\\
&= s^n(B_r(s^n) - sB_r(s^{n-1}))\,.
\end{aligned} \tag{D.13}$$

The factor s^2 in the second term comes from the fact that the summation over d and δ in $a_r(s^n)$ runs from 0 to $s^n - 1$, whereas in $B_r(s^{n-1})$ it only runs from 0 to $s^{n-1} - 1$.

We must therefore evaluate the number of roots of Eq. (D.11). There is an important simplification, however: Since we must eventually evaluate the difference between the number of roots of Eq. (D.11) for s^n and s times the number of those roots for s^{n-1}, we can neglect all roots modulo s^{n-1} which generate exactly

s roots modulo s^n, since these then cancel exactly. In the following we shall always associate to a root modulo s^n the corresponding number modulo s^{n-1}, which is also a root of the equation.

This means that we can immediately discard all roots in which we do not have both $d \equiv \pm 1 \pmod{s}$ and $\delta \equiv \pm 1 \pmod{s}$. Indeed, let either of those two be different from 1 modulo s; for definiteness let it be d. In this case we have

$$d = d_1 + \alpha s^{n-1} \qquad \delta = \delta_1 + \beta s^{n-1} , \tag{D.14}$$

where d_1 is a number modulo s^{n-1}, α is a number modulo s and similarly for δ_1 and β. In fact, we shall systematically use suffixes and Greek letters according to this convention. Putting this into Eq. (D.11) and developing the terms in s^{n-1} as first order infinitesimals (which one may do since they have the same property modulo s^n of being different from zero but vanishing in any power higher than the first) one obtains

$$\alpha(1 - d_1^{-2}) - \beta(1 - \delta_1^{-2}) \equiv 0 \pmod{s} . \tag{D.15}$$

From this follows that any value of β determines α uniquely and that β can be chosen arbitrarily. Thus there are exactly s solutions modulo s^n to every such solution modulo s^{n-1}, so that only solutions with both $d \equiv \pm 1 \pmod{s}$ and $\delta \equiv \pm 1 \pmod{s}$ need be considered.

In this case we can express d and δ in the following way:

$$d = \pm 1 + \sum_{k=1}^{n-1} \alpha_k s^k \qquad \delta = \pm 1 + \sum_{k=1}^{n-1} \beta_k s^k . \tag{D.16}$$

Again, we can put these expressions in the equation to be solved treating the expression as a formal power series which is cut off at order n. Two cases appear: Either all α_k and β_k which would appear in terms higher than linear modulo s^{n-1} vanish or they do not. In the latter case, a straightforward generalization of the above argument shows that every solution modulo s^{n-1} generates s solutions modulo s^n, so that these can again be discarded. Finally, there remains the case in which no terms but the linear ones are non-zero. In this case all these terms cancel, so that we are led to the conditions $r \equiv 0 \pmod{s^{n-1}}$ or $r \equiv \pm 4 \pmod{s^{n-1}}$ for $a_r(s^n)$ to be different from zero.

We must now distinguish between the case where n is even or odd. Let us first consider the even case. We then define n to be $2k$ and consider first the case r equal to zero. This means we can set

$$d = \pm 1 + \sum_{l=k}^{n-1} \alpha_l s^l \qquad \delta = \pm 1 + \sum_{l=k}^{n-1} \beta_l s^l , \tag{D.17}$$

where the signs of the leading ± 1 must now be taken equal, so that r will be zero. In this case the solutions modulo s^n are simply all possible choices of α_l and β_l over the prescribed range, since they automatically cancel. That is, there are $2s^n$ solutions, where the factor 2 comes from the possibility to choose both signs for the leading term. Similarly, one obtains for the same numbers taken modulo s^{n-1} that there are $2s^{n-2}$ solutions, from which follows

$$a_0(s^n) = 2s^n(1 - s^{-1}) . \tag{D.18}$$

Similarly, for n even, it is easy to see that

$$a_{\pm 4}(s^n) = s^n(1 - s^{-1}), \tag{D.19}$$

since we have no possibility of choosing two signs. The rest of the argument is exactly as above, however.

If we now consider the case where r is zero or ± 4 modulo s^{n-1} but not modulo s^n, we find the following: It is not possible to satisfy the equation (D.11) modulo s^n with numbers of the form given by Eq. (D.17). This implies that for $t \neq 0$:

$$a_{ts^{n-1}}(s^n) = -2s^n s^{-1}, \tag{D.20}$$

as well as

$$a_{\pm 4 + ts^{n-1}}(s^n) = -s^n s^{-1}. \tag{D.21}$$

Let us now consider n odd, that is, equal to $2k + 1$. In this case we must take into account all solutions of the form:

$$d = \pm 1 + \sum_{l=k}^{n-1} \alpha_l s^l \qquad \delta = \pm 1 + \sum_{l=k}^{n-1} \beta_l s^l. \tag{D.22}$$

If we now compute d^{-1} from this expression, we find

$$d^{-1} = 1 - \sum_{l=k}^{2k} \alpha_l s^l + \alpha_k^2 s^{2k}. \tag{D.23}$$

Putting the expression in Eq. (D.23) into the equation, we find that

$$d + d^{-1} - \delta - \delta^{-1} = (\alpha_{k-1}^2 - \beta_{k-1}^2) s^{n-2}. \tag{D.24}$$

Therefore, the number of solutions of this equation is the number of zeroes of the expression $\alpha_k^2 - \beta_k^2$, which is $2s - 1$ multiplied by the number of possible choices of the remaining parameters. We therefore have

$$B_0(s^{2k+1}) - sB_0(s^{2k}) = 2(2s - 1)s^{2k} - 2s^{2k+1}. \tag{D.25}$$

From this one obtains again

$$a_0(s^n) = 2s^n(1 - s^{-1}), \tag{D.26}$$

so that the expression in the case of r equal to zero is unchanged. In the case of r equal to ± 4, we must consider the number of times the expression $\alpha_k^2 + \beta_k^2$ becomes zero. This either occurs $2s - 1$ times or once only, depending on whether -1 can be expressed as a square modulo s, which, as is well-known, depends on whether s is equal to 1 or -1 modulo 4. Thus, if $s \equiv 1 \pmod 4$ the above formulae are also recovered in the case of odd n, whereas in the case of $s \equiv -1 \pmod 4$ one obtains

$$B_{\pm 4}(s^{2k+1}) - sB_{\pm 4}(s^{2k}) = s^{2k} - s^{2k+1}, \tag{D.27}$$

from which follows that generally speaking

$$a_{\pm 4}(s^n) = \varepsilon(n, s) s^n (1 - s^{-1}), \tag{D.28}$$

where $\varepsilon(n, s)$ is 1 either if n is even or s of the type $4m + 1$ and -1 in the other case.

It now remains to check the formulae in the only remaining case, which is when r is equal to zero or ± 4 modulo s^{n-1} but not modulo s^n, still with n odd. Under these circumstances the entire reasoning shown above can be repeated word for word, except that at the point where we evaluate how often $\alpha_k^2 - \beta_k^2$ takes the value zero, we must now ask how many times it takes on a value different from zero. As we shall see, this last is independent of the value considered and is $s-1$. This is verified as follows: Let N be the number of ways in which w can be expressed as the difference of two squares. One finds

$$\begin{aligned} N &= \sum_{x=0}^{s-1} \left(1 + \left(\frac{x}{s}\right)\right) \left(1 + \left(\frac{x+w}{s}\right)\right) \\ &= s-1 \,, \end{aligned} \tag{D.29}$$

where we have again used the identity (C.8) to simplify the expression. From this we obtain

$$B_{ts^{n-1}}(s^{2k+1}) - sB_0(s^{2k}) = 2(s-1)s^{2k} - 2s^{2k+1} = -2s^{2k} \,, \tag{D.30}$$

whereas for $\pm 4 + ts^{n-1}$ we must ask how often the number $\alpha_k^2 + \beta_k^2$ takes on a fixed non-zero value. Again, depending on whether $s \equiv 1 \pmod 4$ or not, it takes the value $s-1$ times or $s+1$ times as is seen by looking at the sum

$$\sum_{x=0}^{s-1} \left(1 + \left(\frac{x}{s}\right)\right) \left(1 + \left(\frac{a-x}{s}\right)\right) . \tag{D.31}$$

The remaining case $s = 2$ can be treated analogously. The results are presented in Eqs. (4.8)–(4.12).

The above-discussed values of $A_r(q)$ correspond to matrices with determinant $+1$. For billiard problems we need also matrices with determinant -1. The corresponding formulae for $\beta^{(\pm)}(p,q)$ are presented in Appendix B. Instead of one function $A_r(q)$ one has 3 functions

$$A_r^{(\varepsilon_1,\varepsilon_2)}(q) = \sum_{p:(p,q)=1} \beta(p,q)^{(\varepsilon_1)} \beta(p,q)^{(\varepsilon_2)} \exp\left(\frac{2\pi i p}{q} r\right) . \tag{D.32}$$

$A_r^{(++)}(q) = A_r(q)$ and the two other functions can be computed in the same fashion as $A_r(q)$. We omit the details of the calculations and present only final results.

When $q = s$ is a prime $(s \neq 2)$

$$A_r^{(\varepsilon_1,\varepsilon_2)}(q) = \frac{s^2}{|M_s|^2}\left[s \sum_{x=0}^{s-1} \left(\frac{(x^2 - 4\varepsilon_1)((x+r)^2 - 4\varepsilon_2)}{s} \right) - 1 \right] . \tag{D.33}$$

For later applications the value of $A_0^{(\varepsilon_1,\varepsilon_2)}(s)$ is important. One obtains:

$$\begin{aligned} A_0^{(--)}(s) &= A_0^{(++)} \quad \text{if } s = 4m+1 \\ &= \frac{1}{s^2} \quad \quad \text{if } s = 4m+3 \,, \end{aligned} \tag{D.34}$$

$$A_0^{(+-)}(s) = \frac{1}{s^4(1-s^{-2})}(sH_s(1) - s - 1) \,, \tag{D.35}$$

where

$$H_s(k) = \sum_{x=0}^{s-1} \left(\frac{x(x^2 - k)}{s} \right),$$

$H_s(k) = 0$ if $(-1/s) = -1$.

The explicit formula for $H_s(1)$ is not known but it is possible to show [34] that if $(-1/s) = 1$,

$$(H_s(1))^2 + (H_s(\alpha))^2 = s,$$

where α is a number for which $(\alpha/s) = -1$. Therefore for any s

$$|H_s(1)| < \sqrt{s}.$$

For $q = s^n$ $(n \geqq 2)$ it follows that if $(-1/s) = -1$ (i.e. there is no solution of the equation $\chi^2 \equiv -1 \bmod s$) then

$$A_r^{(--)}(s^n) = A_r^{(+-)}(s^n) = 0.$$

If $(-1/s) = +1$ so that there is a number χ such that $\chi^2 \equiv -1 \bmod s$, then

$$A_r^{(--)}(s^n) = A_{\chi r}^{(++)}(s^n).$$

And finally, if $r \neq \pm 1 \pm \chi$, then $A_r^{(+-)}(s^n) = 0$.

An important consequence of these relations is that for all s

$$A_0^{(+-)}(s^n) = 0.$$

Appendix E

In this Appendix we give some standard properties of the Legendre symbol. Since multiplication modulo q is a group when q is a prime, one has the well known identity

$$x^{q-1} = 1 \quad (x \neq 0). \tag{E.1}$$

From this follows that if x is a square and is non-zero modulo q one has $x^{(q-1)/2}$ equal to one, whereas for x arbitrary it can only take on the values 1 and -1. On the other hand, one finds that the squares of all positive numbers between 1 and $(q-1)/2$ are in fact distinct, so that there are at least $(q-1)/2$ different squares. But a polynomial of degree n (with non-zero leading coefficient) cannot have more than n different roots. We therefore have the general result

$$\left(\frac{x}{q} \right) = x^{(q-1)/2} \,(\bmod q). \tag{E.2}$$

From this follows

$$\left(\frac{xy}{q} \right) = \left(\frac{x}{q} \right) \left(\frac{y}{q} \right) \tag{E.3}$$

as well as the result that -1 has a square root if and only if q is of the form $4k+1$. Since we have exactly $(q-1)/2$ squares and the same number of non-squares we find the simple result

$$\sum_{x=0}^{q-1} \left(\frac{x}{q} \right) = 0. \tag{E.4}$$

Further, one finds, for $a \neq 0$:

$$\sum_{x=0}^{q-1}\left(\frac{x}{q}\right)\left(\frac{x+a}{q}\right)=\sum_{x=1}^{q-1}\left(\frac{x(x+a)x^{-2}}{q}\right)$$
$$=\sum_{x=1}^{q-1}\left(\frac{1+ax^{-1}}{q}\right)=\sum_{x=0}^{q-1}\left(\frac{x}{q}\right)-1=-1\,. \qquad \text{(E.5)}$$

These are the only properties we have used in the text. Much more is known, however, and can be found in [34, 35].

Appendix F

In this Appendix we shall compute the function

$$G_N(x)=2\pi\sum_{\substack{(p,q)=1\\ 0<p/q<x/(2\pi)\\ q>N}}\beta(q)\,, \qquad \text{(F.1)}$$

where $\beta(q)$ is defined as follows: For prime powers ω^n it has the value $A_0(\omega^n)$, where $A_0(p^n)$ is defined by Eqs. (4.5) and (4.6) in the body of the text. $\beta(q)$ is then determined everywhere by the multiplicative property. Let

$$q=\omega_1^{n_1}\omega_2^{n_2}\cdots\omega_k^{n_k} \qquad \text{(F.2)}$$

is a canonical representation of integer q into a product of primes ω_i then

$$\beta(q)=\frac{1}{q^3}\prod_{\omega_i}M\frac{B(\omega_i,n_i)}{(1-\omega_i^{-1})(1-\omega_i^{-2})^2}\,, \qquad \text{(F.3)}$$

where if $\omega \neq 2$,

$$B(\omega,1)=1-2/\omega-1/\omega^2, \qquad B(\omega,n)=2(1-1/\omega)\,.$$

For $\omega=2$,

$$B(2,1)=1/4, \qquad B(2,2)=1/2, \qquad B(2,3)=0\,,$$
$$B(2,4)=1, \qquad B(2,5)=0, \qquad B(2,n)=2 \quad \text{if } n\geqq 6\,. \qquad \text{(F.4)}$$

To take into account the inequalities $p/q<x/2\pi$ and $q>N$ we find it convenient to use the following identity:

$$\frac{1}{2\pi i}\int_{\varepsilon-i\infty}^{\varepsilon+i\infty}\frac{u^{-s}}{s}ds=\begin{cases}0 & \text{if } u>1\\ 1 & \text{if } 0<u<1\end{cases}\,, \qquad \text{(F.5)}$$

where $\varepsilon>0$. We therefore introduce the function

$$\hat{G}(s,s')=\left(\frac{x}{2\pi}\right)^s\frac{1}{N^{s'}}\sum_{(p,q)=1}\frac{q^{s+s'}\beta(q)}{p^s}\,. \qquad \text{(F.6)}$$

One then obtains ($\varepsilon_1, \varepsilon_2 > 0$)

$$G_N(x) = \frac{1}{(2\pi i)^2} \int_{\varepsilon_1 - i\infty}^{\varepsilon_1 + i\infty} \frac{ds}{s} \int_{\varepsilon_2 - i\infty}^{\varepsilon_2 + i\infty} \frac{ds'}{s'} \hat{G}(s, s') . \tag{F.7}$$

The summation in $\hat{G}(s, s')$ is taken over all q and all p coprime. The latter sum can be computed using the standard formula (see e.g. [23])

$$\sum_{(p,q)=1} f(p) = \sum_{k=1}^{\infty} \sum_{\delta | q} f(k\delta)\mu(\delta) , \tag{F.8}$$

where $\mu(\delta)$ is the Möbius function defined as a multiplicative function which is zero on all numbers which are divisible by a square and satisfies

$$\mu(1) = 1 \quad \mu(p) = -1$$

for all primes p. This gives

$$\sum_{(p,q)=1} \frac{1}{p^s} = \sum_{k=1}^{\infty} \frac{1}{k^s} \sum_{\delta | q} \frac{\mu(\delta)}{\delta^s} = \zeta(s) \prod_{\omega | q} (1 - \omega^{-s}) ,$$

where $\zeta(s)$ is the Riemann zeta function and the product extends over all prime factors.

Therefore

$$\hat{G}(s, s') = \left(\frac{x}{2\pi}\right)^s \frac{1}{N^{s'}} \zeta(s) \sum_q \frac{1}{q^{3-s-s'}} \prod_i \frac{1 - \omega_i^{-s}}{(1 - \omega_i^{-1})(1 - \omega_i^{-2})^2} B(\omega_i, n_i) , \tag{F.9}$$

where ω_i and n_i are as in Eq. (F.1). Let us define

$$\frac{C_k(\omega)}{\omega^k} = \sum_{n=1}^{\infty} \frac{1}{\omega^{nk}} B(\omega, n) , \tag{F.10}$$

where $k = 3 - s - s'$. For $\omega \neq 2$ a direct computation gives

$$\begin{aligned} \frac{C_k(\omega)}{\omega^k} &= \frac{B(\omega, 1)}{\omega^k} + \sum_{n=2}^{\infty} \frac{1}{\omega^{nk}} B(\omega, n) \\ &= \frac{1}{\omega^k} \left(1 - \frac{2}{\omega} - \frac{1}{\omega^2}\right) + 2\left(1 - \frac{1}{\omega}\right) \sum_{n=2}^{\infty} \frac{1}{\omega^{nk}} \\ &= \frac{1}{\omega^k} \left(1 - \frac{1}{\omega^2} - 2\frac{\omega^{k-1} - 1}{\omega^k - 1}\right) . \end{aligned} \tag{F.11}$$

Therefore if $\omega \neq 2$

$$C_k(\omega) = 1 - \frac{1}{\omega^2} - 2\frac{\omega^{k-1} - 1}{\omega^k - 1} . \tag{F.12}$$

Similarly

$$C_k(2) = \frac{1}{4} + \frac{1}{2^{k+1}} + \frac{1}{2^{3k}} + \frac{1}{2^{4k-1}(2^k - 1)} . \tag{F.13}$$

We now rewrite Eq. (F.8) as a sum over all ω_i and n_i. One can then perform the sums over the n_i and obtains a factor $C_k(\omega_i)$ for each ω_i. Finally one obtains

$$\hat{G}(s,s') = \left(\frac{x}{2\pi}\right)^s \frac{1}{N^{s'}}\zeta(s)\prod_{\omega}\left(1+\frac{1}{\omega^{3-s-s'}}\frac{1-\omega^{-s}}{(1-\omega^{-1})(1-\omega^{-2})^2}C_{3-s-s'}(\omega)\right), \tag{F.14}$$

where the product is taken over all primes.

This product converges when $1 < \Re s < 2 - \Re s'$. To obtain the function $G(x)$ one should compute the integral (F.6) along the line parallel to the imaginary axis real part of which lies in the above interval. Because $N \gg 1$ it is possible to shift the contour of integration over s' right until it reaches the first singularity of $\hat{G}(s,s')$. It is easy to see that this singularity is a pole at $s' = 2 - s$ coming from the product over all primes. Putting $s' = 2 - s - \varepsilon$ and assuming that $\varepsilon \to 0$ one has

$$\hat{G}(s,s') = \left(\frac{x}{2\pi}\right)^s \zeta(s)\prod_{\omega}\left(1+\frac{1}{\omega^{1+\varepsilon}}\frac{1-\omega^{-s}}{(1-\omega^{-1})(1-\omega^{-2})}C_1(\omega)\right). \tag{F.15}$$

But from Eqs. (F.11) and (F.12) it follows that

$$C_1(\omega) = 1 - \omega^{-2} \tag{F.16}$$

for all ω including $\omega = 2$. Therefore

$$\hat{G}(s,2-s-\varepsilon) = \left(\frac{x}{2\pi}\right)^s \frac{1}{N^{2-s}}\zeta(s)\prod_{\omega}\left(1+\frac{D(\omega)}{\omega^{1+\varepsilon}}\right). \tag{F.17}$$

where

$$D(\omega) = \frac{1-\omega^{-s}}{(1-\omega^{-1})(1-\omega^{-2})}$$

and as $\varepsilon \to 0$

$$\hat{G}(s,2-s-\varepsilon) \to \left(\frac{xN}{2\pi}\right)^s N^{-2}\zeta(s)\zeta(1+\varepsilon)K_s\,, \tag{F.18}$$

where

$$K_s = \prod_{\omega}\left(1+\frac{D(\omega)}{\omega}\right)\left(1-\frac{1}{\omega}\right)$$

because

$$\zeta(s) = \prod_{\omega}(1-\omega^{-s})^{-1}\,.$$

It is well known that the Riemann zeta function $\zeta(s)$ has a pole at $s = 1$ with unit residue (see e.g. [41]) and consequently as $s' \to 2 - s$,

$$\hat{G}(s,s') \to N^{-2}\frac{(xN)^s}{(2\pi)^s}\zeta(s)\frac{K_s}{2-s-s'}\,. \tag{F.19}$$

Integrating over s' one concludes that in the leading order of $1/N$,

$$G_N(x) = \frac{N^{-2}}{2\pi i}\int_{\varepsilon-i\infty}^{\varepsilon+i\infty}\frac{ds}{s(2-s)}\zeta(s)\frac{(xN)^s}{(2\pi)^s}K_s\,. \tag{F.20}$$

For convergence it is necessary that $1 < \varepsilon < 2$. Rewriting $(xN)^s$ as $\exp(-s\ln(xN))$ one easily concludes that if $xN \ll 1$ one can shift the contour of integration right up to the pole at $s = 2$ and

$$G_N(x) \to \frac{x^2}{48} \tag{F.21}$$

because $\zeta(2) = \pi^2/6$ and $K_2 = 1$. If $xN \gg 1$ one can move the contour of integration only left up to the pole $s = 1$ coming from $\zeta(s)$ and

$$G_N(x) \to \frac{x}{N} K_1 \,, \tag{F.22}$$

where

$$K_1 = \prod_\omega \left(1 - \frac{1}{\omega}\right)\left(1 + \frac{1}{\omega(1-\omega^{-2})}\right) = \prod_\omega \left(1 - \frac{1}{\omega(\omega+1)}\right) ,$$

and the product is taken over all primes.

In Fig. 5 we present the plot of $g(x)$. On a double logarithmic scale, the x^2 behaviour for $x \ll 1$ is in evidence. The prefactor is also found to be correct, as shown by the theoretical line. As indicated in the figure captions, we then show $g(x)$ in which the peaks with denominators less than N have been subtracted. In these one sees an approximately linear behaviour over a broad range of x, which is in complete agreement with the behaviour predicted in Sect. 5 and in this Appendix.

For billiard problems one has to compute functions $G(x)$ with $\beta^{(\varepsilon_1,\varepsilon_2)}(q)$ defined in Eq. (6.11). Using the values of $A_0^{(\varepsilon_1,\varepsilon_2)}(q)$ presented in Appendix D one concludes that

$$B^{(--)}(\omega, n) = B^{(++)}(\omega, n)$$

if $(-1/\omega) = 1$ and consequently

$$C_k^{(--)}(\omega) = C_k^{(++)}(\omega)$$

for such primes in particular (see Eq. (F.15))

$$C_1^{(--)}(\omega) = 1 - \omega^{-2} .$$

For primes with $(-1/\omega) = -1$,

$$B^{(--)}(\omega, 1) = 1 - \omega^{-2}, \qquad B^{(--)}(\omega, n) = 0 \quad \text{if } n \geqq 2 ,$$

It means that for all primes the value of $C_1^{(--)}(\omega)$ is the same as for $C_1^{(++)}(\omega)$ and the asymptotics of the function $G^{(--)}(x)$ coincides with that of $G^{(++)}(x)$.

In particular as $x \ll 1/N$,

$$G^{(--)}(x) \to \frac{x^2}{48} , \tag{F.23}$$

and when $x \gg 1/N$

$$G^{(--)}(x) \to \frac{x}{N} K_1 . \tag{F.24}$$

For the function $G^{(+-)}(x)$ one obtains from Eqs. (D.34),

$$B^{(+-)}(\omega, 1) = \frac{1}{\omega^2}(sH_s(1) - s - 1)$$

and $B^{(+-)}(\omega, n) = 0$ for $n \geqq 2$. Therefore

$$C_k^{(+-)}(\omega) = \frac{1}{\omega^2}(sH_s(1) - s - 1)\,.$$

But we have mentioned that

$$|H_s(1)| < \sqrt{s}\,.$$

It means that the product (F.13) will have a singularity only when $s = 2.5$ and as $x \ll 1/N$,

$$G_N^{(+-)}(x) = \mathcal{O}(x^{2.5})\,. \tag{F.25}$$

Acknowledgements. We would like to acknowledge O. Bohigas for useful discussions. One of the authors (F.L.) gratefully acknowledges the warm hospitality during his stay at the IPN Orsay as well as the financial support of DGAPA project IN 100491.

References

1. Hopf, E.: Ergodentheorie. Berlin, Heidelberg, New York: Springer 1937
2. Encyclopedia of Mathematical Sciences, Vol. 1 Dynamical Systems, eds. D.V. Arnold, V.I. Anosov, Berlin, Heidelberg, New York: Springer 1988
3. Balazs, N.L., Voros, A.: Phys. Rep. **143**, 109 (1986)
4. Proceedings of the Les Houches Summer School (1989) Chaos and Quantum Physics. Giannoni, M.J., Voros, A., Zinn-Justin, J. (eds.) Amsterdam: North Holland, 1991
5. Gutzwiller, M.C.: Chaos in Classical and Quantum Mechanics. Berlin, Heidelberg, New York: Springer 1990
6. Selberg, A.: J. Ind. Math. Soc. **20**, 47 (1956)
7. Hejhal, D.: The Selberg Trace Formula for PSL (2, R). Vol. **1** Lecture Notes in Mathematics **548** (1979); Vol. **2**, ibid. **1001** (1983)
8. Gutzwiller, M.C.: J. Math. Phys. **12**, 343 (1971)
9. Bohigas, O., Giannoni, M.-J., Schmit, C.: Phys. Rev. Lett. **52**, 1 (1984); J. Physique Lettres **45**, L1015 (1984)
10. Mehta, M.L.: Random Matrices and the Statistical Theory of Energy Levels. New York Academic Press, 1967
11. Schmit, C.: In: [4]
12. Schmit, C.: Triangular Billiards on the Hyperbolic Plane: Spectral Properties. Report IPNO/TH (1991) (unpublished)
13. Aurich, R., Steiner, F.: Physica D **39**, 169 (1989)
14. Aurich, R., Steiner, F.: Physica D **43**, 155 (1990)
15. Berry, M.V., Tabor, M.: Proc. R. Soc. London **A 356**, 375 (1977)
16. Bogomolny, E., Georgeot, B., Giannoni, M.-J., Schmit, C.: Phys. Rev. Lett. **69**, 1477 (1992)
17. Bolte, J., Steil, G., Steiner, F.: Phys. Rev. Lett. **69**, 2188 (1992)
18. Bogomolny, E., Georgeot, B., Giannoni, M.-J., Schmit, C.: Arithmetical Chaos. Preprint IPNO/TH 93-51
19. Bolte, J.: Ph.D. Thesis DESY (1993)
20. Gelfand, I.M., Graev, M.I., Pyatetskii-Shapiro, I.I.: Representation Theory and Automorphic Functions. Philadelphia-London-Toronto: W.B. Saunders 1969
21. Katok, S.: Fuchsian Group. Chicago Lecture in Mathematics, 1992
22. Berry, M.V.: Proc. R. Soc. London **A400**, 229 (1985)
23. Hardy, G.H., Littlewood, J.E.: Acta Mathematica **44**, 1 (1922)

24. Magnus, W.: Non-Euclidean Tesselations and their Groups. New York: Academic Press (1974)
25. Terras, A.: Harmonic Analysis on Symmetric Spaces and Applications. Berlin, Heidelberg, New York, Springer 1979
26. Sarnak, P.: J. Number Theory **15**, 229 (1982)
27. Matthies, C., Steiner, F.: Phys. Rev. **A44**, R7877 (1991)
28. Huber, H.: Math. Ann. **138**, 1 (1959)
29. Berry, M.V.: In: [4]
30. Bogomolny, E., Schmit, C.: Nonlinearity **6**, 523 (1993)
31. Aurich, R., Sieber, M.: J. Phys. **A27**, 1967 (1994)
32. Karamata, J.: J. reine und angewandte Mathematik **164**, 27 (1931)
33. Feller, W.: An Introduction to Probability Theory and its Applications. vol. **1**, New-York: John Wiley and Sons, 1957; vol. **2**, 1971
34. Vinogradov, I.M.: An Introduction to the Theory of Numbers. Oxford, London, New-York, Paris: Pergamon Press, Paris 1961
35. Hardy, G.H., Wright, E.M.: An Introduction to the Theory of Numbers. (4th ed.) Oxford: Clarendon Press, 1959
36. Kloosterman, H.D.: Acta Math. **49**, 407 (1926)
37. Weil, A.: Proc. Nat. Acad. Sci. USA **34**, 204 (1948)
38. Kusnetsov, N.V.: J. Sov. Math. **29**, 1131 (1985)
39. Venkov, A.B.: Spectral Theory of Automorphic Functions. Proc. Steklov Institute of Math. **4** (1982)
40. Bogomolny, E., Leboeuf, P.: Nonlinearity **7**, 1155 (1994)
41. Titchmarsh, E.C.: The Theory of the Riemann Zeta-Function. Oxford: Clarendon Press, 1951
42. Rankin, R.A.: Modular Forms and Functions. Cambridge: Cambridge University Press, 1977

Communicated by Ya. G. Sinai