ON THE DISTRIBUTION OF CONTINUED FRACTIONS AND MODULAR SYMBOLS

Yuri I. Manin, Matilde Marcolli

Max–Planck–Institut für Mathematik, Bonn, Germany

Abstract. Using the techniques introduced by D. Mayer, we prove an extension of the classical Gauss–Kuzmin theorem about the distribution of continued fractions, which in particular allows one to take into account some congruence properties of consecutive convergents. We then study some averages related to modular symbols.

§0. Introduction and summary

0.1. Continued fractions. We start by fixing the notation which will be used throughout the paper. Considering first $k_1, \dots, k_n$ as independent variables, put for $n \geq 1$

$$[k_1, \dots, k_n] := \frac{1}{k_1 + \frac{1}{k_2 + \dots \frac{1}{k_n}}} = \frac{P_n(k_1, \dots, k_n)}{Q_n(k_1, \dots, k_n)} \tag{0.1}$$

where P_n, Q_n are polynomials with integral coefficients which can be calculated inductively from the relations

$$Q_{n+1}(k_1, \dots, k_n, k_{n+1}) = k_{n+1} Q_n(k_1, \dots, k_n) + Q_{n-1}(k_1, \dots, k_{n-1}),$$

$$P_n(k_1, \dots, k_n) = Q_{n-1}(k_2, \dots, k_n) \tag{0.2}$$

It is convenient also to put formally $Q_{-1} = 0, Q_0 = 1$ which is compatible with (0.2), (0.1).

From (0.2) one readily sees that

$$[k_1, \dots, k_{n-1}, k_n + x_n] =$$

$$\frac{P_{n-1}(k_1, \dots, k_{n-1})\, x_n + P_n(k_1, \dots, k_n)}{Q_{n-1}(k_1, \dots, k_{n-1})\, x_n + Q_n(k_1, \dots, k_n)} = \begin{pmatrix} P_{n-1} & P_n \\ Q_{n-1} & Q_n \end{pmatrix} (x_n) \tag{0.3}$$

where we use the standard matrix notation for fractional linear transformations defining the action of $GL(2)$ on $\mathbf{P}^1$:

$$z \mapsto \frac{az+b}{cz+d} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} (z)$$

If $\alpha \in (0,1)$ is an irrational number, there is a unique sequence of integers $k_n(\alpha) \geq 1$ such that α is the limit of $[\,k_1(\alpha), \dots, k_n(\alpha)\,]$ as $n \to \infty$. Moreover, there is a unique sequence $x_n(\alpha) \in (0,1)$ such that

$$\alpha = [\,k_1(\alpha), \dots, k_{n-1}(\alpha), k_n(\alpha) + x_n(\alpha)\,]$$

for each $n \geq 1$. Rational numbers in $(0,1]$ can be accommodated by allowing finite sequences of $k_i \geq 1$ complemented by zeroes, and similarly for $x_n(\alpha)$.

We can specialize (0.3) at the point α and get by induction

$$\alpha = \begin{pmatrix} 0 & 1 \\ 1 & k_1(\alpha) \end{pmatrix} \dots \begin{pmatrix} 0 & 1 \\ 1 & k_n(\alpha) \end{pmatrix} (x_n(\alpha)). \tag{0.4}$$

Put

$$p_n(\alpha) := P_n(k_1(\alpha), \dots, k_n(\alpha)),\ q_n(\alpha) := Q_n(k_1(\alpha), \dots, k_n(\alpha))$$

so that $p_n(\alpha)/q_n(\alpha)$ is the sequence of convergents to α. Denote also

$$g_n(\alpha) := \begin{pmatrix} p_{n-1}(\alpha) & p_n(\alpha) \\ q_{n-1}(\alpha) & q_n(\alpha) \end{pmatrix}$$

For further use, we reproduce a description of the total set of matrices $g_n(\alpha)$. Put

$$\mathrm{Red}_n = \left\{ \begin{pmatrix} 0 & 1 \\ 1 & k_1 \end{pmatrix} \dots \begin{pmatrix} 0 & 1 \\ 1 & k_n \end{pmatrix} \,|\, k_1, \dots, k_n \geq 1;\ k_i \in \mathbf{Z} \right\}, \tag{0.5}$$

and $\mathrm{Red} := \cup_{n\geq 1} \mathrm{Red}_n \subset GL(2, \mathbf{Z})$.

In the terminology of [LewZa1], Red_n consists of reduced matrices of length n. The following properties of Red are proved in that paper:

(i) *A matrix in $GL(2, \mathbf{Z})$ is reduced iff it has non–negative entries which are non–decreasing downwards and to the right.*

(ii) *The length $l(g)$ of a reduced matrix g and its representation in the form (0.5) are uniquely defined.*

(iii) *Reduced matrices are hyperbolic and have two distinct fixed points on the real axis. Every conjugacy class g of hyperbolic matrices in $GL(2, \mathbf{Z})$ contains reduced representatives. They all have the same length $l(g)$, and there are exactly $l(g)/k(g)$ of them where $k(g)$ is the maximal integer such that $g = h^{k(g)}$ for some h.*

0.1.1. Generalized Gauss problem. Consider a subgroup of finite index $G \subset GL(2, \mathbf{Z})$ and the coset space $\mathbf{P} = GL(2, \mathbf{Z})/G$ with the left transitive action of $GL(2, \mathbf{Z})$ on it:

$$t \mapsto \begin{pmatrix} a & b \\ c & d \end{pmatrix} (t)$$

For any $x \in [0,1]$, $t \in \mathbf{P}, n \geq 0$ put

$$m_n(x,t) := \text{measure of the set } \{\alpha \in (0,1) \,|\, x_n(\alpha) \leq x,\, g_n(\alpha)^{-1}(t_0) = t\} \tag{0.6}$$

where t_0 is the base point of $\mathbf{P}$, the coset of G. Notice that $x_n(\alpha) = g_n(\alpha)^{-1}(\alpha)$ so that m_n is essentially the pullback of the Lebesgue measure on $(0,1) \times \mathbf{P}$ with respect to the operator $g_n(\alpha)$ acting upon α and t simultaneously. Notice also that $g_n(\alpha)^{-1}$ is another notation for the n–th power of the shift operator

$$T:\, (\alpha, t) \mapsto \left(\frac{1}{\alpha} - \left[\frac{1}{\alpha}\right], \begin{pmatrix} -[1/\alpha] & 1 \\ 1 & 0 \end{pmatrix}(t)\right)$$

The first result of this paper is the following generalization of the Gauss–Kuzmin–Lévy formula:

0.1.2. Theorem. *Assume that* $\mathrm{Red}\,(t) = \mathbf{P}$ *for each* $t \in \mathbf{P}$. *Then the limit* $m(x,t) = \lim_{n\to\infty} m_n(x,t)$ *exists and equals*

$$m(x,t) = \frac{1}{|\mathbf{P}|\log 2}\log(1+x). \tag{0.7}$$

A proof is given in §1 below. It starts with a straightforward generalization of the Gauss–Kuzmin inductive expression of m_{n+1} through m_n. To write it down, consider first the following sets: for $y \in (0,1)$, $s \in \mathbf{P}$ put

$$M_n(y,s) := \{\beta \in (0,1) \,|\, x_n(\beta) \leq y,\, g_n(\beta)^{-1}t_0 = s\}.$$

Then we have, using (0.4) and neglecting the rationals which have measure zero:

$$M_{n+1}(x,t) = \coprod_{k=1}^{\infty} \left\{ M_n\left(\frac{1}{k}, \begin{pmatrix} 0 & 1 \\ 1 & k \end{pmatrix}(t)\right) - M_n\left(\frac{1}{x+k}, \begin{pmatrix} 0 & 1 \\ 1 & k \end{pmatrix}(t)\right)\right\}.$$

Therefore,

$$m_{n+1}(x,t) = \sum_{k=1}^{\infty} \left\{ m_n\left(\frac{1}{k}, \begin{pmatrix} 0 & 1 \\ 1 & k \end{pmatrix}(t)\right) - m_n\left(\frac{1}{x+k}, \begin{pmatrix} 0 & 1 \\ 1 & k \end{pmatrix}(t)\right)\right\}. \tag{0.8}$$

Derivating (0.8) in x, we get the following equation for the densities, which introduces an important operator L, the generalized Gauss–Kuzmin operator:

$$m'_{n+1}(x,t) = (Lm'_n)(x,t) := \sum_{k=1}^{\infty} \frac{1}{(x+k)^2} m'_n\left(\frac{1}{x+k}, \begin{pmatrix} 0 & 1 \\ 1 & k \end{pmatrix}(t)\right). \tag{0.9}$$

In the classical case, when $|\mathbf{P}| = 1$, at least four ways to deduce (0.7) from (0.9) are known: two "elementary" deductions going back to R. Kuzmin and P. Lévy respectively (see e.g. [Sch]), and two functional analytic methods, based upon the spectral analysis of the operator L formally defined by (0.9): see [Ba], [BaYu] and [May].

Clearly, the limiting measure must be an eigenfunction of L corresponding to the eigenvalue 1, and both analytic proofs show the convergence of $L^n m_0'$ to this limiting measure by establishing that 1 is the multiplicity one eigenvalue with maximal modulus for the extension of L to an appropriate function space. K. Babenko realizes L as a self–adjoint operator in a Hilbert space, whereas D. Mayer works in the context of nuclear and trace class operators in Banach spaces.

Of these four proofs, we were able to generalize to our context only Mayer's method. Babenko's representation seems to be inadequate for proving convergence. However, it might still be useful for numerical calculations, and we present it in §1.3.

The operator L and its deformation L_s ((1.1) below) were introduced and studied also in a recent paper [ChMay], of which we became aware only after the first draft of this paper was written.

0.2. Modular curves and geodesics. Our study of the generalized Gauss measure (0.7) was motivated by the relationship between continued fractions and one–dimensional homology of modular curves. We will start by recalling the basic features of this relationship in the form given in [Man1]; see also [Mer] for additional information.

Let G_0 be a subgroup of finite index in the fractional linear group $PSL(2,\mathbf{Z}) = SL(2,\mathbf{Z})/(\pm 1)$. It determines the noncompact modular curve $G_0 \setminus H$ (where H is the upper complex half–plane). This curve admits a smooth compactification by a finite number of cusps which are in a natural bijection with the set $G_0 \setminus \mathbf{P}^1(\mathbf{Q})$. Let X_{G_0} (or more precisely, $X_{G_0}(\mathbf{C})$) denote this compactification, φ the respective covering map.

For any two points $\alpha, \beta \in \overline{H} := H \cup \mathbf{P}^1(\mathbf{Q})$ we can define a real homology class ("modular symbol") $\{\alpha,\beta\}_{G_0} \in H^1(X_{G_0}, \mathbf{R})$ by integrating the lifts of differentials ω of the first kind on X_{G_0} along the geodesic path connecting α to β:

$$\int_{\{\alpha,\beta\}} \omega := \int_\alpha^\beta \varphi^*(\omega). \tag{0.10}$$

Modular symbols satisfy the basic additivity and invariance properties:

$$\{\alpha,\beta\}_{G_0} + \{\beta,\gamma\}_{G_0} = \{\alpha,\gamma\}_{G_0},$$

$$\forall g \in G_0,\ \{g\alpha, g\beta\}_{G_0} = \{\alpha, \beta\}_{G_0}. \tag{0.11}$$

The integrals (0.10) can be related to finite, stably periodic, or general infinite continued fractions, depending on the arithmetical nature of the ends α, β. We will briefly treat these three cases separately.

(i) *Finite continued fractions.* Assume first that α, β are cusps. It is known that in this case the modular symbol represents a rational homology class (Manin–Drinfeld's theorem).

By additivity, it suffices to look at the integrals of the form $\int_0^\alpha$, $\alpha \in \mathbf{Q}$. Let

$$g_k := \begin{pmatrix} p_{k-1}(\alpha) & p_k(\alpha) \\ q_{k-1}(\alpha) & q_k(\alpha) \end{pmatrix},\ k = 1, \dots, n,\ \alpha = \frac{p_n(\alpha)}{q_n(\alpha)}.$$

Then, again by additivity, we have

$$\int_0^\alpha \varphi^*(\omega) = \sum_{k=1}^n \int_{p_{k-1}(\alpha)/q_{k-1}(\alpha)}^{p_k(\alpha)/q_k(\alpha)} \varphi^*(\omega) = -\sum_{k=1}^n \int_{\{g_k(0), g_k(i\infty)\}} \omega. \tag{0.12}$$

Finally, in view of (0.11), the k–th integral in (0.12) depends only on the class of g_k in $G_0 \setminus PSL(2, \mathbf{Z})$.

Thus, (0.12) establishes a connection between the distribution of modular symbols $\{\alpha, \beta\}_{G_0}$ and the distribution of pairs of consecutive convergents to α and β in $\mathbf{P}_0 := G_0 \setminus PSL(2, \mathbf{Z})$.

Notice that there is a slight discrepancy with 0.1.1 where we dealt instead with $G \subset GL(2, \mathbf{Z})$ and $GL(2, \mathbf{Z})/G$. In order to reduce the current problem to the former one, we can first replace G_0 by its lift to $SL(2, \mathbf{Z})$ and then denote by G the subgroup generated by this lift and $\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$. We will have a natural identification $\mathbf{P}_0 = G \setminus GL(2, \mathbf{Z})$, and this set in turn identifies with $\mathbf{P}$ in 0.1.1 under the map $g \mapsto g^{-1}$ which is implicit in (0.6).

If α and/or β in (0.10) is real irrational, the integral diverges at this end. In this case, it is natural to study its asymptotic behavior. We will define the "limiting modular symbol" by the following expression whenever it makes sense:

$$\{\{*, \beta\}\}_{G_0} := \lim \frac{1}{T(x, y)} \{x, y\}_{G_0} \in H^1(X_{G_0}, \mathbf{R}). \tag{0.13}$$

Here $x, y \in H$ are two points on the geodesic joining α to β, x is arbitrary but fixed, $T(x, y)$ is the geodesic distance between them, and the limit is taken as y tends to β. In §2 we will prove that if the limit does exist, it depends neither on x nor on α, hence the first argument is replaced by $*$, whereas the double curly brackets remind about the limit. We will discuss the relation of this symbol to continued fractions in two situations.

(ii) *Stably periodic continued fractions.* Among geodesics with two irrational ends there is an important subclass consisting of geodesics that connect two fixed points of a hyperbolic element in G_0. More precisely, any hyperbolic $g \in G_0$ has two fixed points $\alpha^{\pm}$, repelling and attracting, on the real line. Let $\Lambda_g^{\pm}$ be the respective eigenvalues, $0 < \Lambda_g^+ < 1$. The oriented geodesic in H connecting α_g^- to α_g^+ is g–invariant, and the action of g induces on it the shift by the geodesic distance $\lambda(g) := \log \Lambda_g^-$.

For any point x on this geodesic, the image of its segment $[x, gx]$ is a parameterized closed loop on X_{G_0} missing the cusps. (The supporting set–theoretic loop is ran over exactly $k(g)$ times where $k(g)$ is the maximal k such that g is a k–th power in G_0). The homology class of this loop is $\{0, g(0)\}$. When we integrate from a fixed x to $g^n x$, we run over the parameterized loop n times, the geodesic length of the path is $n\,\lambda(g)$, and its homology class is $n\,\{0, g(0)\}$. Therefore the limiting modular symbol (0.13) exists and equals

$$\{\{*, \alpha_g^+\}\} = \frac{\{0, g(0)\}}{\lambda(g)}. \tag{0.14}$$

The most important generating function for closed geodesics is the Selberg zeta function. However, it encodes only the lengths of closed geodesics in the hyperbolic metric. The usual modular symbol in the numerator of (0.14) depends only on the class of g modulo $[G_0, G_0]$ and is additive in g (see [Man1], Prop. 1.4). Perhaps, one can construct a generating function for (0.14) as a combination of Selberg's zetas with abelian characters.

The usual Selberg's zetas were studied in [May], and then in [LewZa1] for GL and SL separately. It turned out that they could be represented as Fredholm determinants $\det(1 - L_s)$ and $\det(1 - L_s^2)$ respectively. Literally the same is true in our generalized setting, when subgroups G or G_0 are introduced. This is proved in [ChMay], and we supply a brief discussion of this in §3.

As we have briefly explained, the distribution of continued fractions and modular symbols at cusps is encoded in the eigenvalues of L_1, 1 being the dominant value producing the distribution (0.7).

From the identity $\det(1 - L_s) = Z(s)$ it follows that the zeroes of $Z(s)$ are exactly those values for which the deformed operator L_s has eigenvalue 1.

J. Lewis and D. Zagier produced also an in–depth study of the respective eigenfunctions for the full modular groups $GL(2, \mathbf{Z})$ and $SL(2, \mathbf{Z})$.

(iii) *General infinite continued fractions.* For this case, we will prove in §2 two results. Namely, we will establish that with an additional assumption the limiting modular symbol (0.13) exists only in a weak sense and and is zero.

0.2.1. Theorem. *Assume that* $\mathrm{Red}(t) = \mathbf{P}_0$ *for each* $t \in \mathbf{P}_0$ *(transitivity condition). Then (0.13) weakly converges to zero.*

For the precise description of the sense in which this convergence holds, see §2.3 and in particular, (2.21).

This vanishing can be compared with a well known interpretation of the Selberg trace formula for compact surfaces: quantum mechanical averages for the geodesic flow can be calculated as if this flow were classically concentrated on closed geodesics.

The transitivity assumption for Red will be checked in §2 in the case $G_0 = \Gamma_0(N)$.

The next result of §2 concerns a series of averaging formulas of a different kind. Drawing on a lemma of P. Lévy, we will explain in §2 how to calculate averages (over $[0,1]$) of some functions of α defined by the sums over all pairs of consecutive convergents of α. Here we will state an interesting particular case, providing averages of weighted modular symbols.

Fix a prime number $N > 0$ and put $G_0 = \Gamma_0(N)$. We will assume that the genus of $X_{G_0} = X_0(N)$ is ≥ 1, otherwise our identities become trivial. Consider a $\Gamma_0(N)$–invariant differential $\varphi^*(\omega)$ on H which is a cusp eigenform for all Hecke operators and denote by $L_\omega^{(N)}(s)$ (resp. $\zeta^{(N)}(s)$) its Mellin transform (resp. Riemann's zeta) with omitted Euler N–factor. More precisely, the coefficients of $L_\omega^{(N)}(s)$ are Hecke eigenvalues of $\varphi^*(\omega)/dz$.

0.2.2. Theorem. *We have for* $\mathrm{Re}\, t > 0$*:*

$$\int_0^1 d\alpha \sum_{n=0}^{\infty} \frac{q_{n+1}(\alpha) + q_n(\alpha)}{q_{n+1}(\alpha)^{1+t}} \int_{\{0,q_n(\alpha)/q_{n+1}(\alpha)\}} \omega =$$

$$\left[\frac{\zeta(1+t)}{\zeta(2+t)} - \frac{L_\omega^{(N)}(2+t)}{\zeta^{(N)}(2+t)^2} \right] \int_0^{i\infty} \varphi^*(\omega). \tag{0.16}$$

0.3. Related problems and comments. (i) In [Man2], it was shown that after a choice of Schottky uniformization, the Arakelov geometry of a complex curve X at arithmetical infinity can be described in terms of the hyperbolic geometry of geodesics not on X itself, but rather in the hyperbolic handlebody having X as the boundary at infinity. It would be interesting to clarify the statistical aspects of the closed and infinite geodesics in the same vein as above and to relate them to Arakelov geometry. We hope to return to this question later.

(ii) The quotient $PGL(2,\mathbf{Z}) \setminus \mathbf{P}^1(\mathbf{R})$ can be identified with the space of classes of continued fractions modulo the equivalence relation "$k_{n+n_0}(\alpha) = k_{n+n_1}(\beta)$ *for* *some* n_0, n_1 *and all* n". Classical results on various averages like

$$\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} f(k_i(\alpha))$$

state that these averages are almost everywhere constant functions on this space.

On the other hand, this space and its finite coverings corresponding to subgroups $G \subset PGL(2,\mathbf{Z})$ constitute the boundary of the analytic moduli stack of elliptic curves whose irrational points are invisible in algebraic geometry (only cusps admit classical algebraic interpretation). According to the emerging general philosophy, this boundary is a bridge to the world of non–commutative geometry, in particular, the limiting geometric objects parameterized by this boundary must be "non–commutative elliptic curves", perhaps endowed with an additional structure and represented by objects like triangulated categories of their coherent sheaves. For some explanations, see [Soi], [Man4].

Our study concerns rather the geometry of the boundary itself, which also must be considered as (a tower of) "non–commutative modular curves" in Connes' spirit. From this perspective, the modular complex introduced in [Man1] and further studied in [Mer], [Gon] and other papers, provides a definition of the (co)homology of this boundary compatible with passage to the limit. In the general framework of noncommutative geometry this definition looks however somewhat *ad hoc.* It would be interesting to relate it to the standard homological constructions of non–commutative geometry.

In the same spirit, the identity (0.16) and its generalizations considered in §2.1–2.2 show that a considerable part of the theory of automorphic functions in the upper half plane can be recast as the study of averages of certain functions defined on the boundary $\mathbf{R}$ as sums of the type (2.1) representing a class of boundary functions with automorphic behavior.

For another family of similar phenomena, see [Za] and [LawZa].

It is interesting to remark in this context that the Gauss–Kuzmin operator L_s vaguely looks like a "Hecke operator at the arithmetical infinity", and has some properties that might be expected of such an operator.

§1. Gauss–Kuzmin operator

1.1. Operator L_s. Consider the operator L_s acting on functions of two variables (x,t) and formally given by

$$(L_s f)(x,t) := \sum_{k=1}^{\infty} \frac{1}{(x+k)^{2s}} f\left(\frac{1}{x+k}, \begin{pmatrix} 0 & 1 \\ 1 & k \end{pmatrix} (t)\right). \tag{1.1}$$

The variable x here varies in a subset of $\mathbf{C}$ stable with respect to all maps $x \mapsto (x+k)^{-1}$, $k = 1, 2, \dots$. In our context this subset will always contain $[0,1]$. The variable t belongs to a finite $GL(2,\mathbf{Z})$–set $\mathbf{P}$ endowed with a base point t_0. The parameter s here is real and $> 1/2$. We are mostly interested in the case $s = 1$. In §3 we will allow s to take complex values. Finally, f will vary in a linear space of functions stable with respect to L_s and containing the function $m_0'(x,t) = \delta_{t,t_0}$ (cf. 0.1.1).

For the proof of Theorem 0.1.2, we want to create a functional analytic context in which the machinery of Krasnoselskii's theorem as stated in [May], 7.25, becomes applicable. To this end we make the following choices, slightly generalizing Mayer. (They were also made in [ChMay]).

(i) *Definition domain.* It will be $\mathbf{D} \times \mathbf{P}$ where $\mathbf{D} := \{z \in \mathbf{C} \,|\, |z-1| < 3/2\}$. We will call the subsets $\mathbf{D} \times \{t\}$ *sheets.* Notice that each map $z \mapsto (z+k)^{-1}$ transforms $\mathbf{D}$ strictly into itself.

(ii) *Functional spaces.* We shall consider the complex Banach space $B_{\mathbf{C}} := V_{\mathbf{C}}(\mathbf{D} \times \mathbf{P})$ consisting of functions holomorphic on each sheet and continuous on its boundary. We shall also consider the real Banach space $B := V(\mathbf{D} \times \mathbf{P})$ of functions holomorphic on each sheet and continuous on its boundary, real at the real points of each sheet. Both spaces are endowed with the supremum norm. These spaces obviously contain δ_{t,t_0} and are stable with respect to L_s for real $s > 1/2$. The space $B_{\mathbf{C}}$ is also stable with respect to L_s with $\mathrm{Re}\, s > 1/2$.

(iii) *Positive cone.* Denote by $K \subset B$ the cone consisting of functions taking non–negative values at real points of each sheet. We have $K \cap -K = 0$ (K is proper) because a nonzero analytic function cannot vanish on an interval. We also have $B = K - K$ (K is reproducing) because $f = (f+r) - r$, and if r is large and positive, $f + r, r \in K$. Finally, functions positive at all real points of all sheets form the interior of K.

We write $f \leq g$ if $f - g \in K$.

(iv) L_s *is K–positive.* This means that $L_s(K) \subset K$ which is obvious.

1.1.1. Lemma. *Assume that* $\mathbf{P}$ *contains no proper invariant subsets with respect to the operators* Red *(see (0.5)).*

Then for each nonzero $f \in K$ there exist two real positive constants a, b and an integer $p \geq 1$ such that $a \leq L_s^p f \leq b$.

Proof. The upper bound is trivial. Assume that for some f and all p the lower bound is zero. This means that for each $p \geq 1$, $L_s^p f$ vanishes at some point (x_p, t_p) with x_p real in the closure of $\mathbf{D}$. Since in (1.1) all summands are non–negative at real points, when $f \in K$, we see that f must vanish at all points contained in the set $\cup_{p\geq 1}\mathrm{Red}_p(x_p, t_p)$.

From our assumption it follows that for some q and any $t \in \mathbf{P}$, $\mathrm{Red}_q(t) = \mathbf{P}$. By downward induction we deduce first that for any p, $L_s^p f$ has real zeroes on all sheets. Then, again by downward induction, one sees that for each t there exists a sequence of integers $q_n \to \infty$ and real points y_n in $\mathbf{D}$ such that $f(x,t)$ vanishes at all $x \in \cup_n \mathrm{Red}_{q_n}(y_n)$. But the intersection of the latter set with $[0,1]$ is dense in $[0,1]$. A nonvanishing holomorphic function cannot have as many zeroes. This contradiction proves our assertion.

1.1.2. Example. Take for G the subgroup generated by the lift of $\Gamma_0(N)$ and the sign change. We will show that $\mathrm{Red}_3(t) = \mathbf{P}$ for any t. In fact, elements of $\mathbf{P}$ can be thought of as points of the projective line over $\mathbf{Z}/N$, that is, formal quotients of residues $\bmod N$ that can be represented by pairwise prime integers ([Man1]). Moreover, this encoding can be chosen compatible with the usual action of $GL(2)$ upon $\mathbf{P}^1$.

Let us break these points into three groups:

(I) $\{u/1 \,|\, (u,N) = 1\}$.

(II) $\{du/1 \,|\, d/N, d > 1, (u,N) = 1\}$.

(III) $\{1/du\} \,|\, d/N, d > 1, (u,N) = 1$.

Let us say that s can be obtained from t in one step, if $s \in \mathrm{Red}_1(t)$, that is, $s = (t+k)^{-1}$ for some k. The following statements are straightforward, and taken together, prove our claim.

From any single element of I∪II we can obtain in one step all elements of the set I∪III.

From any single element of III one can obtain in one step an element of II, by adding zero and inverting. Hence from the total III one can obtain the total II in one step.

1.1.3. Lemma. *$L_s : B_{\mathbf{C}} \to B_{\mathbf{C}}$ is a nuclear operator of order zero, in particular compact and trace class for* $\mathrm{Re}\, s > 1/2$.

Proof. The reasoning is the same as in [May], and we only sketch it. Denote the k–th summand in (1.1) by $(\pi_{s,k} f)(x,t)$. Each $\pi_{s,k}$ is nuclear, and $\sum_k \|\pi_{s,k}\|$ converges for $\mathrm{Re}\, s > 1/2$. In fact, the spectrum of $\pi_{s,k}$ can be easily calculated.

Let z_k be the unique fixed point of $\gamma_k := \begin{pmatrix} 0 & 1 \\ 1 & k \end{pmatrix}$ in $\mathbf{D}$, and $\mu_i^{(k)}$ the spectrum of the permutation induced by this matrix on $\mathbf{P}$. Then the spectrum of $\pi_{s,k}$ is $\{(-1)^n (z_k + k)^{-2(s+n)} \mu_i^{(k)}\}$, $n \geq 0$.

1.2. Proof of the generalized Gauss–Kuzmin Theorem. We have now checked all the conditions for the applicability of the Theorem 7.25 of [May]; see also [KraLS] for more details.

Using this theorem, we conclude that there exists exactly one eigenfunction of norm one f_s of L_s in the interior of K and its eigenvalue $\lambda_{0,s}$ is positive and simple. All other eigenvalues have strictly lesser modulus. For any $f \in B$ and $\epsilon > 0$, we have

$$L_s^n f = \text{const}\, \lambda_{0,s}^n\, f_s + O(c(\epsilon)(q+\epsilon)^n \lambda_{0,s}^n) \tag{1.2}$$

as $n \to \infty$, where $q = q(L_s) < 1$ is the spectral margin of L_s.

In the case $s = 1$ we know a positive eigenfunction: this is Gauss' density $\dfrac{1}{x+1}$, independent of t. The normalization constant is straightforward.

This argument completes the proof of the generalized Gauss–Kuzmin theorem. We will, however, provide some more details about the deduction of (1.2), because this technique can be useful also in the treatment of lower eigenvalues.

The basic result that ensures the existence of eigenfunctions in certain invariant cones is the following ([KraLS], Theorem 9.2):

If K is a cone in a real Banach space B satisfying $\overline{K-K} = B$ and L is a compact operator with $LK \subseteq K$, and with positive spectral radius $r(L)$, then $r(L)$ is an eigenvalue of L with a corresponding eigenfunction in K.

We also recall some results which enable us to establish that the top eigenvalue of an operator with an invariant cone is simple, and when the rest of the spectrum is of strictly smaller absolute value.

Following [May], we say that an operator L is u–bounded, with respect to a function $u \in K$, if for any $f \in K$ there exists some $n > 0$ and $a, b > 0$ such that

$$au \leq L^n f \leq bu$$

(Since we allow a power of L, this is a weaker definition of u–boundedness than the one on p. 110 of [KraLS], cf. their remark on p. 111.)

Lemma 1.1.1 shows that L_s is u–bounded with respect to the constant function $u(x,t) = 1$. The lower bound guarantees the positivity of the spectral radius $r(L_s)$ and hence the applicability of the Theorem 9.2 of [KraLS]. This fact follows from Lemma 9.2 of [KraLS].

We then have the following result ([KraLS], Theorem 11.1). Assume that the cone K is reproducing and the K–positive operator L is u–bounded. Assume moreover that L has an eigenvalue $\lambda_0 > 0$ with an eigenvector $f \in K$. Then the eigenvalue λ_0 is simple. A proof of this fact can be obtained using Lemma 11.1 of [KraLS]. This is actually a simple result of linear algebra which uses only the fact that f is an interior point of the cone (nowhere vanishing), and that iterates of the operator L map boundary points of the cone different from $\{0\}$ to interior points of the cone.

Furthermore, Theorem 11.4 of [KraLS] shows that with the same hypothesis as in the previous result, every eigenvalue λ of L different from λ_0 satisfies $|\lambda| < \lambda_0$. This result follows from the observation that if the operator L is u–bounded then it is also f–bounded, where again f is the eigenfunction in K with eigenvalue λ_0. Then, if h is an eigenfunction with eigenvalue λ the estimate

$$-\alpha(\lambda_0 - \epsilon) f \leq \lambda h \leq \alpha(\lambda_0 - \epsilon) f$$

for some $\epsilon > 0$ follows easily, where $\alpha > 0$ is the smallest positive number such that $-\alpha f \leq h \leq \alpha f$ is satisfied. This gives $|\lambda| \leq \lambda_0 - \epsilon$.

Since in our case we know that the operator L_s is compact, the previous result implies that all the other eigenvalues λ (hence all the points in the spectrum of L_s) satisfy the estimate $|\lambda| < q\lambda_0$, for some $q < 1$.

Finally, we have a result on the convergence of iterates, cf. Theorem 15.4 of [KraLS].

The cone K in our case contains some ball of positive radius. In this case, Theorem 9.11 on p. 97 of [KraLS] ensures that the adjoint operator L^* acting on the dual Banach space B' has an eigenfunctional f^* in the adjoint cone K^* of linear K–positive functionals, with eigenvalue $\lambda \leq r(L)$ where $r(L)$ is the spectral radius of L. In our case L_s has an eigenvalue $\lambda_0 = r(L_s)$ and a corresponding eigenfunction f which is an interior point of the cone K. Thus, if $L^* f^* = \lambda f^*$, for a non-trivial $f^* \in K^*$, and if moreover $f^*(f) > 0$, then $\lambda = \lambda_0$, because

$$\lambda_0 f^*(f) = f^*(\lambda_0 f) = f^*(Lf) = (L^* f^*)(f) = \lambda f^*(f).$$

Assume that the operator L has a simple eigenvalue equal to the spectral radius, $\lambda_0 = r(L)$, and the remaining part of the spectrum lies in the disk $|\lambda| < q\, r(L)$ for some $q < 1$. Let $f \in K$ be the eigenfunction of the eigenvalue λ_0. Let f^* be an eigenfunctional for L^* in K^*, with eigenvalue λ_0 as above, satisfying $f^*(f) = 1$. Then the sequence of iterates

$$f_{n+1} = L f_n$$

converges to the eigenfunction f in the following sense.

Define the operators $Uh := f^*(h)f$, and $U^\perp h := h - f^*(h)f$. We have

$$\lim_n \frac{\|U^\perp f_n\|}{\|Uf_n\|} = 0$$

and the rate of convergence is estimated by

$$\frac{\|U^\perp f_n\|}{\|Uf_n\|} \le c(\epsilon)(q+\epsilon)^n \frac{\|U^\perp f_1\|}{\|Uf_1\|},$$

for arbitrarily small $\epsilon > 0$. In other words, the iterates converge as fast as a geometric progression with ratio arbitrarily close to the spectral margin $q = q(L)$, cf. §15.2 of [KraLS], and in particular Theorem 15.3, where a more refined estimate of the coefficient $c(\epsilon)$ is also given.

Sometimes similar techniques may be applied to the study of the second eigenvalue, by supplying a suitable real Banach space with an invariant cone for the operator $h \mapsto L_s h - \lambda_{0,s} f^*(h) f$, cf. [May].

1.3. The integral kernel operator. In this subsection, we consider the formal operator (1.1) in another functional space, and show that it admits there a representation with integral kernel, generalizing that of [Ba], [BaYu].

We will have to assume additionally that the $GL(2,\mathbf{Z})$–set $\mathbf{P}$ (see 1.1) is such that the action

$$t \mapsto \begin{pmatrix} 0 & 1 \\ 1 & k \end{pmatrix}(t)$$

depends only on $k \bmod N$ for an appropriate integer N. This assumption is satisfied for instance when G is a congruence subgroup. In the following, we fix such N.

With this assumption, (1.1) can be written as

$$(L_s f)(z,t) = \sum_{p=-N+1}^{0} \sum_{k:\, k\ge 1,\, k\equiv p\,(N)} \frac{1}{(k+z)^{2s}} f\left(\frac{1}{k+z}, \begin{pmatrix} 0 & 1 \\ 1 & p \end{pmatrix}(t)\right) \qquad (1.3)$$

Our new functional space $\mathbf{H}$ will consist of functions $f(z,t)$ holomorphic on the sheets $\{\mathrm{Re}\, z > -1/2\} \times t$.

We first recall a useful identity which is the essential ingredient in the arguments of [Ba], [BaYu], [May], namely

$$\sum_{k\ge 1} \frac{1}{(k+z)^{2s}} \exp\left(\frac{-\xi}{k+z}\right) = \xi^{-s+1/2} \int_0^\infty \eta^{s-1/2} e^{-\eta(z+1/2)} \frac{J_{2s-1}(2\sqrt{\xi\eta})}{2\sinh(\eta/2)} d\eta. \quad (1.4)$$

(This is the formula (111) of [May] rewritten in a way that looks more similar to the corresponding formula in [Ba], [BaYu].)

For our purpose, it is useful to consider also the following corollary of (1.4).

1.3.1. Lemma. *Let $p \in \{-N+1, \dots, 0\}$. We have*

$$\sum_{k\colon k\geq 1,\, k\equiv p\ (N)} \frac{1}{(k+z)^{2s}} \exp\left(-\frac{\xi}{k+z}\right) =$$

$$N^{2s-2}\xi^{-s+1/2} \int_0^\infty \eta^{s-1/2} \exp\left(-\eta\left(z+p+N/2\right)\right) \frac{J_{2s-1}(2\sqrt{\xi\eta})}{2\sinh\left(N\eta/2\right)}\, d\eta. \tag{1.5}$$

Proof. We have

$$\sum_{l\geq 1} \frac{1}{(p+lN+z)^2} \exp\left(-\frac{\xi}{p+lN+z}\right) =$$

$$\frac{1}{N^2}\sum_{l\geq 1} \frac{1}{(l+(z+p)/N)^2} \exp\left(-\frac{\xi/N}{l+(z+p)/N}.\right).$$

Plugging (1.4) in and redenoting $N\eta$ as the new η, we get (1.5).

It is convenient to write the functions $f(z,t)$ in the form

$$f(z,t) = \sum_{j=0}^{|\mathbf{P}|-1} f_j(z)\delta_j(t),$$

with respect to a basis of delta functions. We have

$$\delta_j\left(\begin{pmatrix} 0 & 1 \\ 1 & p \end{pmatrix}(t)\right) = \sum_l A_{jl}(p)\,\delta_l(t),$$

where $\mathbf{A}(p)$ is the matrix representing the permutation of the sheets.

With this notation, we can write L_s in the form

$$(L_s f)(z,t) = \sum_{j=0}^{|\mathbf{P}|-1} \sum_{p=-N+1}^{0} \sum_{k\colon k\geq 1,\, k\equiv p\ (N)} f_j\left(\frac{1}{k+z}\right) \frac{1}{(k+z)^{2s}}\, \delta_j(\begin{pmatrix} 0 & 1 \\ 1 & p \end{pmatrix}(t)) =$$

$$\sum_{j,l=0}^{|\mathbf{P}|-1} \sum_{p=-N+1}^{0} \sum_{k\colon k\geq 1,\, k\equiv p\ (N)} f_j\left(\frac{1}{k+z}\right) \frac{1}{(k+z)^{2s}} A_{jl}(p)\delta_l(t).$$

We now introduce the following operators. Let $\mathcal{L}$ denote the Fourier–Laplace transform

$$(\mathcal{L}g)(z,t) := \int_0^\infty e^{-\xi z} g(\xi,t)\, d\xi.$$

We also define the multiplication operator

$$(Tg)(\xi,t) := e^{-\xi/2} g(\xi,t),$$

as in [Ba], [BaYu], and

$$(Sg)(\xi,t) := S(\xi)\, g(\xi,t),$$

with

$$S(\xi) = \frac{(1-e^{-N\xi})^{1/2}}{\xi^{-1/2+s}}.$$

The operator $\mathcal{L}T$ is an isometric isomorphism between the space $\mathcal{L}_2((0,\infty)\times \mathbf{P})$ and $\mathbf{H}$ endowed with the norm

$$\|f\|^2 := \frac{1}{2\pi}\sum_{j=0}^{|\mathbf{P}|-1}\int_{-\infty}^{\infty} |f_j(0+iy)|^2 dy.$$

Define the matrix function

$$\Theta_{jl}(\xi) := e^{-(N-1)\xi/2}\sum_{p=-N+1}^{0} e^{-\xi p} A_{jl}(p). \tag{1.6}$$

We can and will choose its square root with complex valued real analytic entries $\Theta_{jl}^{1/2}(\xi)$:

$$\Theta_{jl}(\xi) = \sum_{k=0}^{|\mathbf{P}|-1} \Theta_{jk}^{1/2}(\xi)\Theta_{kl}^{1/2}(\xi),$$

for all $\xi \in (0,\infty)$. We define

$$\hat{\Theta}_{jl}(\xi) := e^{-(N-1)\xi/4}\Theta_{jl}^{1/2}(\xi). \tag{1.7}$$

The results of the following Lemma and Proposition represent the analog in our setting of the main result of [Ba], see also 7.4.2 of [May].

1.3.2. Lemma. *For a function $f(z,t) = \sum_j f_j(z)\,\delta_j(t)$ such that*

$$f_j = \mathcal{L}TS^{-1}h_j,$$

we can write the operator L_s in the form

$$(L_s f)(z,t) = \sum_{l=0}^{|\mathbf{P}|-1}(\sum_{j=0}^{|\mathbf{P}|-1}\int_0^{\infty} e^{-\eta z} S(\eta)^{-1}\Theta_{jl}(\eta) e^{-\eta/2} e^{-N\eta/4}\times$$

$$\int_0^\infty \tilde{\kappa}(\xi,\eta)e^{N\xi/4}e^{-\xi/2}h_j(\xi)d\xi d\eta)\ \delta_l(t) \tag{1.8}$$

with a function $\tilde{\kappa}(\xi,\eta)$ *satisfying* $\tilde{\kappa}(\xi,\eta)=\tilde{\kappa}(\eta,\xi)$.

Proof. Assume that the functions $f_j(z)$, for $j=0,\ldots,|\mathbf{P}|-1$, are in the range of the operator $\mathcal{L}T$, namely

$$f_j=\mathcal{L}Tg_j,$$

for some function g_j in $\mathcal{L}_2(0,\infty)$. This means that we can write

$$f_j(z)=\int_0^\infty e^{-\xi z}e^{-\xi/2}g_j(\xi)d\xi.$$

We now apply (1.5) and obtain

$$=N^{2s-2}\int_0^\infty \xi^{-s+1/2}\int_0^\infty \eta^{s-1/2}e^{-\eta(z+p+N/2)}\frac{J_{2s-1}(2\sqrt{\xi\eta})}{2\sinh(N\eta/2)}e^{-\xi/2}g_j(\xi)d\xi d\eta.$$

With our previous definition of $\Theta(\xi)$ as in (1.6) we can write

$$\sum_p\sum_{k:\,k\geq 1,\,k\equiv p\,(N)} f_j\left(\frac{1}{k+z}\right)\frac{1}{(k+z)^{2s}}A_{jl}(p)=$$

$$N^{2s-2}\int_0^\infty e^{-\eta z}\int_0^\infty (\frac{\eta}{\xi})^{1/2-s}\frac{J_{2s-1}(2\sqrt{\xi\eta})}{2\sinh(N\eta/2)}e^{-(\xi+\eta)/2}\Theta_{jl}(\eta)g_j(\xi)d\xi d\eta.$$

Now, for $g_j(\xi)=S(\xi)^{-1}h_j(\xi)$, this can be rewritten as

$$\int_0^\infty e^{-\eta z}S(\eta)^{-1}\Theta_{jl}(\eta)\int_0^\infty \frac{N^{2s-2}J_{2s-1}(2\sqrt{\xi\eta})}{\sqrt{2}\sinh(N\eta/2)^{1/2}}e^{-(\xi+\eta)/2}e^{-N\eta/4}\frac{g_j(\xi)}{\xi^{1/2-s}}d\xi\, d\eta=$$

$$\int_0^\infty e^{-\eta z}S(\eta)^{-1}\Theta_{jl}(\eta)\int_0^\infty \tilde{\kappa}(\xi,\eta)e^{-(\xi+\eta)/2}e^{-N\eta/4}e^{N\xi/4}h_j(\xi)d\xi\, d\eta,$$

where we have set

$$\tilde{\kappa}(\xi,\eta):=\frac{N^{2s-2}J_{2s-1}(2\sqrt{\xi\eta})}{2\sinh(N\eta/2)^{1/2}\sinh(N\xi/2)^{1/2}}.$$

Now the final step.

1.3.3. Proposition. *On the range of* $\mathcal{R} := \mathcal{L}TS^{-1}\hat{\Theta}$*, the operator* L_s *satisfies*

$$\mathcal{R}^{-1}L_s\mathcal{R} = \mathcal{M},$$

where $\mathcal{M}$ *is the integral kernel operator*

$$(\mathcal{M}\zeta)(\eta,t) = \sum_{i=0}^{|\mathbf{P}|-1}\int_0^\infty \sum_{j=0}^{|\mathbf{P}|-1} \mathcal{M}_{ij}(\eta,\xi)\zeta_j(\xi)d\xi\ \delta_i(t),$$

with $\zeta(\xi,t) = \sum_j \zeta_j(\xi)\delta_j(t)$*. The integral kernel is of the form*

$$\mathcal{M}_{ij}(\eta,\xi) = \kappa(\eta,\xi)\sum_r \hat{\Theta}_{jr}(\xi)\hat{\Theta}_{ri}(\eta),$$

where the function

$$\kappa(\xi,\eta) = e^{(N/4-1/2)(\xi+\eta)}\tilde{\kappa}(\xi,\eta)$$

still satisfies $\kappa(\xi,\eta)=\kappa(\eta,\xi)$*, but the integral kernel is in general not symmetric.*

Proof. For a function

$$f(z,t) = \sum_{j=0}^{|\mathbf{P}|-1} f_j(z)\delta_j(t) = \sum_{j=0}^{|\mathbf{P}|-1}\int_0^\infty e^{-\xi z}e^{-\xi/2}S(\xi)^{-1}\sum_{i=0}^{|\mathbf{P}|-1}\hat{\Theta}_{ij}(\xi)\zeta_i(\xi)d\xi\,\delta_j(t),$$

we have

$$(L_sf)(z,t) = \sum_{l=0}^{|\mathbf{P}|-1}(L_sf)_l(z)\delta_l(t),$$

with $(L_sf)_l(z)$ of the form

$$\sum_{j=0}^{|\mathbf{P}|-1}\int_0^\infty e^{-\eta z}\Theta_{jl}(\eta)e^{-\eta/2}S(\eta)^{-1}\int_0^\infty e^{-N\eta/4}\tilde{\kappa}(\xi,\eta)e^{N\xi/4}e^{-\xi/2}\sum_{i=0}^{|\mathbf{P}|-1}\hat{\Theta}_{ij}(\xi)\zeta_i(\xi)d\xi d\eta.$$

We can write this equivalently as

$$\sum_{k,i=0}^{|\mathbf{P}|-1}\int_0^\infty e^{-\eta z}\hat{\Theta}_{kl}(\eta)e^{-\eta/2}S(\eta)^{-1}\int_0^\infty \mathcal{M}_{ik}(\eta,\xi)\zeta_i(\xi)d\xi d\eta,$$

where we define

$$\mathcal{M}_{ki}(\eta,\xi) := e^{(N/4-1/2)(\eta+\xi)}\tilde{\kappa}(\eta,\xi)\sum_{j=0}^{|\mathbf{P}|-1}\hat{\Theta}_{ij}(\xi)\hat{\Theta}_{jk}(\eta).$$

Thus, we have obtained

$$L_s\,\mathcal{L}TS^{-1}\hat{\Theta}\,\zeta = \mathcal{L}TS^{-1}\hat{\Theta}\,\mathcal{M}\,\zeta.$$

1.4. Remarks about the l–adic case. Let l be a prime number, $\mathbf{Z}_l$ (resp. $\mathbf{Q}_l$) the ring of l–adic integers (resp. the field of all l–adic numbers). Put also

$$\overline{\mathcal{Z}}_l := \left\{ \frac{n}{l^r} \mid n, r \in \mathbf{Z}, r \geq 0 \right\} \cap (0, l) \subset \mathbf{Q},\ \mathcal{Z}_l := \overline{\mathcal{Z}}_l \setminus \{1, 2, \ldots, l-1\}. \qquad (1.9)$$

Every irrational $\alpha \in \mathbf{Q}_l$ has a unique representation in the form

$$\alpha = \frac{a_{-r}}{l^r} + \cdots + \frac{a_{-1}}{l} + a_0 + a_1 l + \cdots + a_s l^s + \ldots,\ \ a_i \in \{0, 1, \ldots, l-1\}.$$

The l–adic norm of this α is $|\alpha|_l = l^r$ if $a_{-r} \neq 0$. We then put

$$[\alpha]_l := \frac{a_{-r}}{l^r} + \cdots + \frac{a_{-1}}{l} + a_0 \in \overline{\mathcal{Z}}_l. \qquad (1.10)$$

In more invariant terms, $[\alpha]_l$ is the unique element in $\overline{\mathcal{Z}}_l$ such that $|\alpha - [\alpha]_l| < 1$.

Repeating the usual reasoning, one sees that each irrational l-adic α with $|\alpha|_l < 1$ determines a unique sequence of $k_n(\alpha) \in \mathcal{Z}_l$ and $x_n(\alpha) \in l\mathbf{Z}_l \setminus \{0\}$ such that

$$\alpha = [\,k_1(\alpha), \ldots, k_{n-1}(\alpha), k_n(\alpha) + x_n(\alpha)\,] \qquad (1.11)$$

for each $n \geq 1$. We get thus the formalism of the theory of continued fractions in the l–adic setting, in which $\mathbf{R}, \mathbf{Z}, [0,1)$ are replaced respectively by $\mathbf{Q}_l, \overline{\mathcal{Z}_l}, l\mathbf{Z}_l$. Notice that the successive convergents are still rational numbers, but the incomplete quotients $k_n(\alpha)$ generally are not integral.

It will be convenient to restrict ourselves to irrational α in $\mathbf{Q}_l^* \setminus \mathbf{Z}_l^*$. In this case all $k_n(\alpha)$ will belong to $\mathcal{Z}_l$.

The shift operator is given by $T:\ \alpha \mapsto \alpha^{-1} - [\alpha^{-1}]_l$, and it transforms $\mathbf{Q}_l^* \setminus \mathbf{Z}_l^*$ into itself.

The definition of a (deformed) transfer operator, however, presents interesting new problems. We can consider two basic options.

(A) We can try to define the formal transfer operator by the classical formula

$$(L_s f)(x) := \sum_{k \in \mathcal{Z}_l} \frac{1}{(x+k)^{2s}} f\left(\frac{1}{x+k}\right) \qquad (1.12)$$

in which $\mathbf{Z}$ is replaced by $\mathcal{Z}_l$.

We could have included a second argument t, but did not do it in order to focus on the peculiarity of (1.12) apparent already in this straightforward version of the classical setting. Namely, we should not imagine f as a function taking real or complex values. In fact, otherwise $\dfrac{1}{(x+k)^{2s}}$ will be defined only at rational points x and will not tend to zero as k runs over $\mathcal{Z}_l$ since from the archimedean viewpoint, $\mathcal{Z}_l$ is a dense set inside $[0,l)$ so that (1.12) will tend to diverge, unless f is highly discontinuous and tends to zero when the denominator of the argument tends to infinity. But this last property will be lost after an application of L_s.

However, $\mathcal{Z}_l$ is discrete and unbounded in the l–adic sense, so that (1.12) still makes sense as an operator in various l–adic function spaces. For example, one can consider the space of analytic functions on $l\mathbf{Z}_l$ represented by convergent series $\sum_{n\geq 0} a_n x^n$.

This remark, and parallels with the theories of l–adic uniformization and Drinfeld modules, suggest the following problems.

(i) Find a natural Banach space of l–adic functions in which (1.12) would define a compact operator.

Since compact operators are nuclear in the l–adic theory, this would allow us to define l–adic Selberg's zeta values at integral points $2s > 0$ as $\det(1 - L_{2s})$.

(ii) Define the set of l–adic reduced matrices by the same prescription as (0.5), but this time with k_i running over $\mathcal{Z}_l$.

Can one find a characterization of this set similar to that given in [LewZa1] and reproduced in 0.1? Assuming we know an l–adic zeta, can one find an Euler product for it similar to (3.1) below?

(iii) Again assuming a positive answer to the first question, is there an eigenfunction with eigenvalue 1 of L_1? Can one find its measure–theoretic interpretation in terms of Mazur's theory of l–adic integration? (See e. g. [Man5], §8 and 9).

At this point, we may notice that the passage from T to L_1 in the classical case implicitly involves integration with respect to the additively invariant measure, since these operators are adjoint via the obvious bilinear form determined by this integration, cf. formula (2.19) below.

It is well known that this invariant measure becomes inadequate for integrating l–adic valued functions, for the simple reason that the smaller is, say, an l–adic ball, the larger is its invariant measure l–adically.

Instead, the l–adic integration invented by B. Mazur for treating l–adic L–functions, utilizes finitely additive functions on open/closed subsets of, say, $\mathbf{Z}_l$ which take values in bounded subsets of finite–dimensional l–adic spaces. Such measures produce linear functionals on the spaces of functions satisfying the Lipschitz condition. See [Man5], §8 and §9, where a more general class of measures of moderate growth is introduced and studied as well.

Hence another option, perhaps more natural than the formal prescription (1.12) is:

(B) Define and study the transfer operators on various spaces of l–adic measures, as well as appropriate modifications of the questions (i)–(iii).

§2. Calculation of certain averages

2.1. P. Lévy's lemma. Let f be a complex valued function defined on pairs of coprime integers (q,q') such that $q \geq q' \geq 1$ and $f(q,q') = O(q^{-\varepsilon})$ for some $\varepsilon > 0$. Put for $\alpha \in (0,1]$

$$l(f,\alpha) = \sum_{n=1}^{\infty} f(q_n(\alpha), q_{n-1}(\alpha)). \tag{2.1}$$

2.1.1. Proposition (P. Lévy, 1929). *We have*

$$\int_0^1 l(f,\alpha)d\alpha = {\sum}' \, \frac{f(q,q')}{q(q+q')}. \tag{2.2}$$

Sums and integrals in (2.1), (2.2) converge absolutely and uniformly.

A notational convention: prime at the summation sign as in (2.2) refers to the domain $q \geq q' \geq 1$, $(q,q') = 1$.

Proof. This Proposition is an immediate consequence of the following statement.

For any $q \geq q' \geq 1$ with $(q,q') = 1$ there exists a unique $n \geq 0$ such that one can find $\alpha \in (0,1]$ with $q_n(\alpha) = q$, $q_{n-1}(\alpha) = q'$. Moreover, all such α form a semi-interval of length $\dfrac{1}{q(q+q')}$.

In fact, assume that such an α and n exist, and let $p_n(\alpha)/q_n(\alpha)$, $p_{n-1}(\alpha)/q_{n-1}(\alpha)$ be the respective convergents to α, then we have

$$p_{n-1}(\alpha)\, q_n(\alpha) - p_n(\alpha)\, q_{n-1}(\alpha) = (-1)^n.$$

Together with the conditions $p_k(\alpha) \leq q_k(\alpha)$ this allows us to reconstruct n uniquely by induction and shows that all α with this property fill the semi-interval

$$\frac{p_{n-1}(\alpha)z + p_n(\alpha)}{q_{n-1}(\alpha)z + q_n(\alpha)}, \quad z \in (0,1].$$

(cf. (0.3)). Conversely, for any (q,q') we can start with complementing this line by $p \leq q, p' \leq q'$ to a reduced $(2,2)$–matrix with determinant ± 1, and then produce the continued fraction for p/q with neighboring convergents p/q, p'/q'. This proves the lemma.

It is often convenient to have the summation domain in (2.2) extended to all $q \geq q' \geq 1$. One can do this by first extending f to this domain in the following way: choose a function $\kappa : \mathbf{N} \to \mathbf{Z}$ and a number t and put

$$F(q,q') := \kappa(d)\, d^{-t} f(q,q'),\ d := \text{g. c. d.}\,(q,q'). \tag{2.3}$$

Then

$$\sum{}' \frac{f(q,q')}{q(q+q')} = \zeta(\kappa,t)^{-1} \sum_{q\geq q'\geq 1} \frac{F(q,q')}{q(q+q')}, \tag{2.4}$$

where $\zeta(\kappa,t) := \sum_{d\geq 1} \kappa(d)\, d^{-t}$. (This remark is also contained in [L].)

We will now combine this with the results of [Man1] in order to prove Theorem 0.2.2.

2.2. Averaging weighted modular symbols. In this subsection we keep the notation explained before the statement of the Theorem 0.2.2. In particular, modular symbols refer to the group $\Gamma_0(N)$. We start with the identity (20) in [Man1]:

$$\sum_{d/m} \sum_{b=1}^{d} \int_{\{0,b/d\}} \omega = (\sigma(m) - c_m) \int_0^{i\infty} \varphi^*(\omega). \tag{2.5}$$

Here $(m,N) = 1$, $\phi^*(\omega)/dz$ is a cusp form for $\Gamma_0(N)$ with eigenvalue c_m with respect to the Hecke operator T_m, $\sigma(m)$ is the sum of the divisors of m.

Multiply this identity by m^{-2-t} and sum over all m prime to N:

$$\sum_{m:\,(m,N)=1} \frac{1}{m^{2+t}} \sum_{d/m} \sum_{b=1}^{d} \int_{\{0,b/d\}} \omega = \left[\sum_{m:\,(m,N)=1} \frac{\sigma(m)}{m^{2+t}} - L_\omega^{(N)}(2+t) \right] \int_0^{i\infty} \varphi^*(\omega). \tag{2.6}$$

Any symbol $\{0, \frac{q'}{q}\}$, $(q,q') = 1$, occurs in the guise $\{0, \frac{b}{d}\}$ for some d/m only when q divides m, and then exactly $\tau(mq^{-1})$ times where τ is the number of divisors. Hence the integration path in the left hand of (2.6) can be rewritten in the following way:

$$\sum_{m:\,(m,N)=1} \sum_{q/m} \frac{\tau(mq^{-1})}{m^{2+t}} \sum_{q'\leq q,\,(q,q')=1} \{0, \frac{q'}{q}\} =$$

$$\sum_{n:\,(n,N)=1} \frac{\tau(n)}{n^{2+t}} \sum_{q:\,(q,N)=1} \frac{\sum_{q'\leq q,\,(q,q')=1} \{0, \frac{q'}{q}\}}{q^{2+t}} =$$

$$\zeta^{(N)}(2+t)^2 \left[\sum_{q:\,(q,N)=1} \frac{1}{q^{2+t}} \sum_{q'\leq q,\,(q,q')=1} \{0, \frac{q'}{q}\} \right]. \tag{2.7}$$

Moreover, the first series inside the square brackets in (2.6) equals

$$\zeta^{(N)}(1+t)\, \zeta^{(N)}(2+t).$$

Hence (2.6) divided by $\zeta^{(N)}(2+t)^2$ can be rewritten as

$$\sum_{q:\,(q,N)=1} \frac{1}{q^{2+t}} \sum_{q'\le q,\,(q,q')=1} \int_{\{0,\frac{q'}{q}\}} \omega =$$

$$\left[\frac{\zeta^{(N)}(1+t)}{\zeta^{(N)}(2+t)} - \frac{L_\omega^{(N)}(2+t)}{\zeta^{(N)}(2+t)^2}\right] \int_0^{i\infty} \varphi^*(\omega). \tag{2.8}$$

The left hand side of (2.8) can be represented as the right hand side of (2.2) with the function $f(q,q')$ which vanishes for q/N and otherwise equals

$$f(q,q') = \frac{q+q'}{q^{1+t}}\,\{0,\frac{q'}{q}\}. \tag{2.9}$$

Let us define a new function $\tilde{f}(q,q')$ by the same formula (2.9) *for all* relatively prime (q,q'). Since N is prime, we have $\{0,\frac{q'}{q}\} = \{0,i\infty\}$ for N/q. Therefore, writing (2.2) for this $\tilde{f}$, and integrating $\varphi^*(\omega)$, we get on the left hand side the same expression as on the left hand side of (0.13). The right hand side becomes the sum of the right hand side of (2.8) and

$$\sum_{d=1}^{\infty} \frac{\phi(Nd)}{(Nd)^{2+t}} \int_0^{i\infty} \varphi^*(\omega) = \left[\frac{\zeta(1+t)}{\zeta(2+t)} - \frac{\zeta^{(N)}(1+t)}{\zeta^{(N)}(2+t)}\right] \int_0^{i\infty} \varphi^*(\omega) \tag{2.10}$$

where ϕ is the Euler function. This sum equals the right hand side of (0.16). This completes the proof.

2.2.1. Comments and variations. The distribution of modular symbols was studied by D. Goldfeld ([Gol1], [Gol2]) who has found interesting connections between the conjectural asymptotic behavior of certain sums involving such symbols and other number–theoretical problems, e. g. the *abc*–conjecture. One of Goldfeld's conjectures reads:

$$\sum_{\substack{c^2M^2+d^2\le X\\ c\equiv 0\,(N)}} \{0,\frac{b}{d}\}_N \sim R(iM)\,X$$

as $X\to\infty$, where the sum is taken over matrices in $\Gamma_0(N)$, $R(iM) := \int_{iM}^{i\infty}$, and both sides are considered as functionals on the space of $\Gamma_0(N)$ cusp forms of weight two.

D. Goldfeld and C. O'Sullivan introduced a class of Eisenstein series twisted by modular symbols and established their analytic properties. The simplest series of this kind can be represented as the right hand side of (2.2) if one chooses for f the following function (depending on z, s as parameters):

$$\frac{f(q,q')}{q(q+q')} := \chi(q)\,\{0,\frac{q'}{q}\}_N \sum_{g\in A_{q,q'}} \mathrm{Im}\,(gz)^s$$

where $A_{q,q'}$ is the set of matrices in $\Gamma_0(N)$ with the second column $(q',q)^t$.

Here is another class of quite simple functions f that might produce interesting specializations of (2.2):

$$\frac{f(q,q')}{q(q+q')} := \frac{\chi_1(q)\,\chi_2(q')}{q^{s_1}q'^{s_2}}.$$

They lead to some identities involving double logarithms at roots of unity at the right hand side of (2.2). As Goncharov has shown in [Gon], relations between these numbers can be described in terms of the modular complex for $\Gamma_1(N)$. This stresses the relevance of the modular symbols in the study of the distribution of continued fractions.

Our last example is a function that was introduced in [AlZa]:

$$\frac{f(q,q')}{q(q+q')} := x^{\sum k_j(q/q')-1}\,q\log_2 q.$$

2.3. Proof of the Theorem 0.2.1. We now return to the notation and conventions explained in the paragraph around formula (0.13). In particular, β is real irrational. We start with proving that whenever the limit (0.13) exists, it does not depend on α (independence on $x\in H$ with fixed α is obvious).

We will compare the behavior of (0.13) for two geodesics Γ_1,Γ_2 ending at β. It suffices to consider the case when Γ_1 starts from $i\infty$, whereas Γ_2 starts from some real $\alpha<\beta$.

Denote by $p_n(\beta)/q_n(\beta) = p_n/q_n$ the convergents to β. If n is large enough and has the appropriate parity, the respective convergents will have the following positions on the real line:

$$\frac{\alpha+\beta}{2} < \frac{p_{n-1}}{q_{n-1}} < \beta < \frac{p_n}{q_n}\,. \qquad (2.11)$$

Besides, we always have

$$\left|\frac{p_{n-1}}{q_{n-1}}-\beta\right| > \left|\frac{p_n}{q_n}-\beta\right|\,. \qquad (2.12)$$

Let $[a, b]$ generally denote the geodesic joining a to b. Define two sequences of points in H by

$$z_n := \Gamma_1 \cap \left[\frac{p_{n-1}}{q_{n-1}}, \frac{p_n}{q_n}\right], \ \zeta_n := \Gamma_2 \cap \left[\frac{p_{n-1}}{q_{n-1}}, \frac{p_n}{q_n}\right]. \tag{2.13}$$

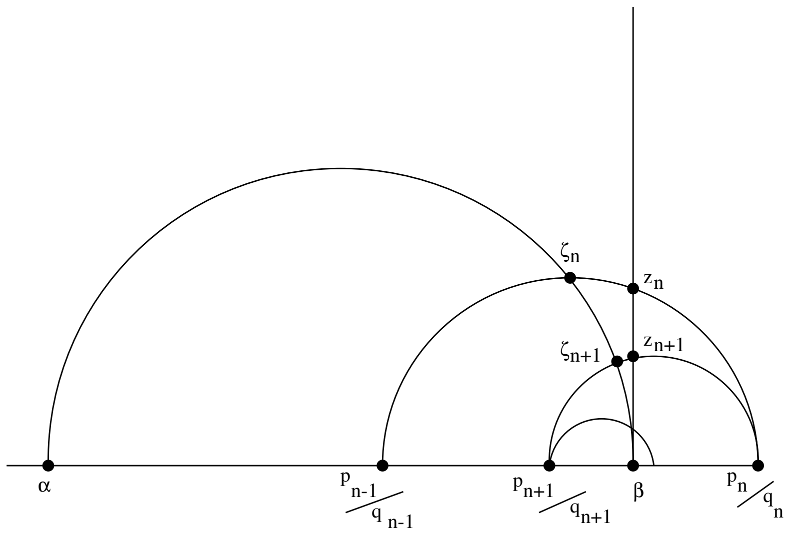
From (2.11)–(2.13) it follows that

$$\frac{1}{2q_n q_{n+1}} < \operatorname{Im} z_n < \frac{1}{2q_{n-1}q_n},$$

and if moreover n is large enough,

$$\frac{\theta}{2q_{n+2}q_{n+1}} < \theta \operatorname{Im} z_{n+1} < \operatorname{Im} \zeta_n \leq \frac{1}{2q_{n-1}q_n}$$

where θ is some fixed constant between 0 and 1. The easiest way to convince oneself of this is to look at a picture containing all the relevant geodesics.



The geodesic distance from any fixed $x_1 \in \Gamma_1$ to $z \in \Gamma_1$ equals $-\log \operatorname{Im} z + O(1)$ as $z \to \beta$. The similar distance from a fixed $x_2 \in \Gamma_2$ to $\zeta \in \Gamma_2$ to β equals $-\log \operatorname{Im} \zeta + O(1)$ as $\zeta \to \beta$.

Taking into account our inequalities and the additivity of modular symbols, we obtain

$$\frac{1}{T(x_1, z_n)} \{x_1, z_n\} = \frac{1}{T(x_2, \zeta_n) + O(1)} [\{x_2, \zeta_n\} + O(1)]. \tag{2.15}$$

From this and (2.14) it follows that both limits exist or otherwise simultaneously, and have a common value whenever they both exist.

Moreover, according to the Khintchin–Lévy theorem we have for almost all β

$$\log q_n = Cn(1+o(1)),\ C = \frac{\pi^2}{12\log 2} \tag{2.16}$$

as $n \to \infty$. Hence for almost all β we can replace the limit (0.13) by

$$\lim_{n\to\infty} \frac{1}{2Cn}\{i\infty, z_n\} = \lim_{n\to\infty} \frac{1}{2Cn}\sum_{i=1}^{n}\left\{\frac{p_{i-1}(\beta)}{q_{i-1}(\beta)}, \frac{p_i(\beta)}{q_i(\beta)}\right\}. \tag{2.17}$$

Temporarily fixing n, we will consider the sum of modular symbols on the right hand side of (2.17) as a function of β, and then prove that the resulting sequence of functions weakly converges to zero in the $\mathcal{L}_2$–sense. For this, we need the following lemma.

2.3.1. Lemma. *Let $\mathbf{P}_0$ be a finite left $GL(2,\mathbf{Z})$–set such that $\mathrm{Red}^{-1}(t) = \mathbf{P}_0$ for each $t \in \mathbf{P}_0$. Let $\varphi : \mathbf{P}_0 \to H$ be a function with values in an $\mathbf{R}$–vector space, $t_0 \in \mathbf{P}_0$. Then we have*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\varphi(g_i(x)^{-1}t_0) = \frac{1}{|\mathbf{P}_0|}\sum_{s\in\mathbf{P}_0}\varphi(s), \tag{2.18}$$

where the limit is taken in the sense of weak convergence in $\mathcal{L}_2([0,1]\times\mathbf{P}_0)$, and

$$g_k(x) = \begin{pmatrix} p_{k-1}(x) & p_k(x) \\ q_{k-1}(x) & q_k(x)\end{pmatrix}.$$

Proof. Consider the shift

$$T(x,t) = \left(\frac{1}{x} - \left[\frac{1}{x}\right], \begin{pmatrix} -[1/x] & 1 \\ 1 & 0\end{pmatrix}(t)\right)$$

We denote by λ the measure on $[0,1]\times\mathbf{P_0}$ given by the standard Lebesgue measure on $[0,1]$ and the counting measure on $\mathbf{P_0}$.

The Gauss–Kuzmin operator $L = L_1$ that we discussed in §1 is the adjoint of this shift T, in the sense that, for any function $h \in \mathcal{L}_1([0,1]\times\mathbf{P_0},\lambda)$ and any $f \in B_{\mathbf{C}}$ we have

$$\int_{[0,1]\times\mathbf{P_0}} f\cdot Lh\,d\lambda = \int_{[0,1]\times\mathbf{P_0}} (f\circ T)\,h\,d\lambda. \tag{2.19}$$

The eigenfunctional of of L^* denoted $f^* \in K^*$ in §1.2 can be taken as

$$h \mapsto \int_{[0,1]\times \mathbf{P_0}} h(x,t)d\lambda(x,t).$$

From 1.2 it follows that for any $h \in B_{\mathbf{C}}$ we have strong convergence

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n}(L^k h)(x,t) = \frac{1}{|\mathbf{P}_0|\log 2}\,\frac{1}{1+x}\int_{[0,1]\times \mathbf{P_0}} h\,d\lambda. \tag{2.20}$$

According to (2.19), this is equivalent to the convergence

$$\int_{[0,1]\times \mathbf{P_0}} \frac{1}{n}\sum_{k=1}^{n} f(T^k(x,t))h(x,t)d\lambda(x,t) \to$$

$$\frac{1}{|\mathbf{P}_0|\log 2}\left(\int_{[0,1]\times \mathbf{P_0}} \frac{f(x,t)}{1+x} d\lambda(x,t)\right)\int_{[0,1]\times \mathbf{P_0}} h\,d\lambda,$$

for any $f \in B_{\mathbf{C}}$ and any test function $h \in B_{\mathbf{C}}$. If we consider a function $f(x,t) = \varphi(t)$, independent of $x \in [0,1]$, we obtain that for any t

$$\frac{1}{n}\sum_{k=1}^{n}\varphi(g_k(x)^{-1}t) \to \frac{1}{|\mathbf{P_0}|}\sum_{s\in \mathbf{P_0}}\varphi(s)$$

weakly in $\mathcal{L}_2$, because among the test functions we have all polynomials.

This is equivalent to (2.18) for $\mathbf{C}$–valued functions and therefore also for vector–valued ones.

We can now conclude the proof of the Theorem 0.2.1. If our modular curve is $G_0 \setminus \overline{H}$, we put $\mathbf{P}_0 = PSL(2,\mathbf{Z})/G_0$ and consider $\mathbf{P}_0$ as a left $GL(2,\mathbf{Z})$–set as explained in the Introduction.

Since modular symbols are left G_0–invariant, we can find a function φ and $t_0 \in \mathbf{P}_0$ such that

$$\varphi(g_k(\beta)^{-1}\,t_0) = \{g_k(\beta)(0), g_k(\beta)(i\infty)\} = \left\{\frac{p_{k-1}(\beta)}{q_{k-1}(\beta)}, \frac{p_k(\beta)}{q_k(\beta)}\right\}.$$

It follows from (2.18) that the weak limit (2.17) is

$$\frac{1}{2C\,|\mathbf{P}_0|}\sum_k \{h_k(0), h_k(i\infty)\}$$

where h_k now run over a complete set of representatives of $\mathbf{P}_0$. But this last sum vanishes. In fact, let

$$\sigma = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Then $\{h_k\sigma\}$ as well is a complete system of representatives, and

$$\{\sigma(0), \sigma(i\infty)\} = -\{0, i\infty\}.$$

Let us stress that the pointwise behavior of (2.17) might be wildly oscillating. We proved only that for any measurable set E we have

$$\lim_{n\to\infty} \frac{1}{n} \int_E \sum_{i=1}^{n} \left\{ \frac{p_{i-1}(\beta)}{q_{i-1}(\beta)}, \frac{p_i(\beta)}{q_i(\beta)} \right\} d\beta = 0. \qquad (2.21)$$

§3. Selberg's zeta function

3.1. Notation. In this section we explain the definition of Selberg's zeta for subgroups of finite index $G \subset GL(2,\mathbf{Z})$ and $G_0 \subset SL(2,\mathbf{Z})$, their representation as Fredholm determinant, and relations to geodesics on modular curves. We closely follow [LewZa1], pp. 3–6, whose version requires only minor modifications. For a much more comprehensive treatment, see [ChMay].

As in [LewZa1], for $g \in GL(2,\mathbf{Z})$ put

$$D(g) = \mathrm{Tr}\,(g)^2 - 4\det(g),\ \ N(g) = \left(\frac{\mathrm{Tr}\,(g) + D(g)^{1/2}}{2}\right)^2,$$

and call g hyperbolic if $\mathrm{Tr}\,(g)$ and $D(g)$ are positive. A hyperbolic matrix is primitive if it is not a nontrivial power of an element of $GL(2,\mathbf{Z})$. For a hyperbolic g set

$$\chi_s(g) = \frac{N(g)^{-s}}{1-\det(g)\,N(g)^{-1}}.$$

As above, put $\mathbf{P} := GL(2,\mathbf{Z})/G$ and denote by $\rho_{\mathbf{P}}$ the natural representation of $GL(2,\mathbf{Z})$ in the space of functions on $\mathbf{P}$. Finally, put

$$Z_G(s) := \prod_{g\in \mathrm{Prim}} \prod_{m=0}^{\infty} \det\,\left[1-\det(g)^m\, N(g)^{-s-m}\,\rho_{\mathbf{P}}(g)\right] \tag{3.1}$$

where Prim is a set of representatives of all $GL(2,\mathbf{Z})$–conjugacy classes of primitive hyperbolic elements of $GL(2,\mathbf{Z})$.

For $G_0 \subset SL(2,\mathbf{Z})$, we define $Z_{G_0}(s)$ in the same way, replacing Prim by Prim_0, a set of representatives of all $SL(2,\mathbf{Z})$–conjugacy classes of primitive hyperbolic elements of $SL(2,\mathbf{Z})$, and $\mathbf{P}$ by $\mathbf{P}_0$.

3.2. Theorem. *We have*

$$Z_G(s) = \det(1-L_s),\ \ Z_{G_0}(s) = \det(1-L_s^2) \tag{3.2}$$

where L_s is given by (1.1) and considered as a nuclear operator in the space $B_{\mathbf{C}}$.

We give only a sketch of formal calculations for $Z_G(s)$. Using notation as in the proof of Lemma 1.1.2, we have

$$-\log\det(1-L_s) = \sum_{l=1}^{\infty} \frac{\mathrm{Tr}\,L_s^l}{l} = \mathrm{Tr}\,\left(\sum_{l=1}^{\infty}\frac{1}{l}\left(\sum_{n=1}^{\infty}\pi_{s,n}\right)^l\right) =$$

$$\mathrm{Tr}\left(\sum_{g\in \mathrm{Red}} \frac{1}{l(g)}\pi_s(g)\right) = \sum_{g\in \mathrm{Hyp}} \frac{1}{k(g)}\,\chi_s(g)\,\tau_g. \tag{3.3}$$

Here we define the operator $\pi_s(g)$ for a reduced matrix g as in (0.5) as the product of the respective π_{s,k_i}, and $l(g)$ means its length. Hyp denotes a set of representatives of all conjugacy classes of hyperbolic matrices, and $k(g)$ the maximal integer such that $g = h^{k(g)}$. The last piece of notation is

$$\tau_g := \mathrm{Tr}\,(\rho_{\mathbf{P}}(g)) = \mathrm{card}\,\{t \in \mathbf{P} \mid g(t) = t\}. \tag{3.4}$$

The appearance of τ_g is explained by the fact that our $\pi_s(g)$ acts as the tensor product of the similar operator $\pi_s(g)$ for the case $G = GL(2,\mathbf{Z})$ and of $\rho_{\mathbf{P}}$, and our trace is the product of the respective traces.

Using the properties (i)–(iii) of Red summarized at the end of 0.1, we can keep rewriting (3.3):

$$\sum_{g\in \mathrm{Hyp}} \frac{1}{k(g)}\,\chi_s(g)\,\tau_g = \sum_{g\in \mathrm{Prim}}\sum_{k=1}^{\infty}\frac{1}{k}\,\frac{N(g)^{-ks}\,\tau_{g^k}}{1-\det\,(g)^k\,N(g)^{-k}}. \tag{3.5}$$

On the other hand, from (3.1) we find:

$$-\log Z_G(s) = \sum_{g\in \mathrm{Prim}}\sum_{m=0}^{\infty}\mathrm{Tr}\sum_{k=1}^{\infty}\frac{1}{k}\det\,(g)^{mk}\,N(g)^{-(s+m)k}\,\rho_{\mathbf{P}}(g^k) =$$

$$\sum_{g\in \mathrm{Prim}}\sum_{k=1}^{\infty}\frac{1}{k}\frac{N(g)^{-ks}\,\tau_{g^k}}{1-\det\,(g)^k\,N(g)^{-k}} \tag{3.6}$$

This finishes the formal argument which differs from that of [LewZa1], §1, only by the presence of τ_g. The subsequent check of convergence in [LewZa1] and the argument of §3 concerning SL generalize in the same straightforward way.

Finally, in order to interpret (3.6) in the language of closed geodesics, it remains only to notice that if $G \subset GL(2,\mathbf{Z})$ is the lift of $G_0 \subset PSL(2,\mathbf{Z})$ as in the Introduction, then $G_0 \setminus H$ can be naturally identified with $GL(2,\mathbf{Z}) \setminus (H \times \mathbf{P})$, and any closed geodesic on the respective modular curve is covered by the geodesics $[\alpha_g^-, \alpha_g^+]$ lying on those sheets which are left invariant by the respective hyperbolic matrix $g \in G$, in agreement with (3.4).

References

[AlZa] J. C. Alexander, D. Zagier. *The entropy of a certain infinitely convolved Bernoulli measure.* J. Lond. Math. Soc., (2), 44 (1991), 121–134.

[Ba] K. I. Babenko, *On a problem of Gauss.* Dokl. Akad. Nauk SSSR, Tom 238 (1978) No. 5, 1021–1024. (English translation: Soviet Math. Dokl. Vol.19 (1978) N.1, 136–140.)

[BaYu] K. I. Babenko, S. P. Yurev, *On a problem of Gauss.* Sel. Math. Sov. Vol. 2 (1982) N.4, 331–378.

[ChMay] C.-H. Chang, D. Mayer. *Thermodynamic formalism and Selberg's zeta function for modular groups.* Regular and Chaotic Dynamics, vol. 5, No 3 (2000), 281–312.

[Gol1] D. Goldfeld. *Zeta functions formed with modular symbols.* In: Proc. Symp. Pure Math.. vol. 66, 1 (1999), 111–121.

[Gol2] D. Goldfeld. *The distribution of modular symbols.* In: Number Theory in Progress (A. Schinzel's Festschrift, vol. 2, Walter de Gruyter, Berlin–New York, 1999, 849–865.

[Gon] A. Goncharov. *The double logarithm and Manin's complex for modular curves.* Math. Res. Letters, 4(1997), 617–636.

[KraLS] M. Krasnosel'skij, Je. Lifshits, A. Sobolev. *Positive linear systems.* Heldermann Verlag, 1989.

[L] P. Lévy. *Sur les lois de probabilité dont dépendent les quotients complets et incomplets d'une fraction continue.* Bull. Soc. Math. France, 557 (1929), 178–194.

[LawZa] R. Lawrence, D. Zagier. *Modular forms and quantum invariants of 3–manifolds.* In: Asian J. Math. 3 (1999) (Atiyah's Festschrift), No. 1, 93–107.

[LewZa1] J. Lewis, D. Zagier. *Period functions and the Selberg zeta function for the modular group.* In: The Mathematical Beauty of Physics, Adv. Series in Math. Physics 24, World Scientific, Singapore, 1997, pp. 83–9.

[LewZa2] J. Lewis, D. Zagier. *Period functions for Maass wave forms.* Preprint.

[Man1] Yu. Manin. *Parabolic points and zeta-functions of modular curves.* Math. USSR Izvestija, vol. 6, No. 1 (1972), 19–64, and Selected Papers, World Scientific, 1996, 202–247.

[Man2] Yu. Manin. *Explicit formulas for the eigenvalues of Hecke operators.* Acta Arithmetica, 24 (1973), 239–249.

[Man3] Yu. Manin. *Three–dimensional hyperbolic geometry as ∞–adic Arakelov geometry.* Inv. Math., 104 (1991), 223–244.

[Man4] Yu. Manin. *Mirror symmetry and quantization of abelian varieties.* Preprint math.AG/0005143

[Man5] Yu. Manin. *Periods of parabolic forms and p–adic Hecke series.* Math. USSR Sbornik, 21:3 (1973), 371–393, and Selected Papers, World Scientific, 1996, 268–297.

[May] D. Mayer. *Continued fractions and related transformations.* In: Ergodic Theory, Symbolic Dynamics and Hyperbolic Spaces, Eds. T. Bedford et al., Oxford University Press, Oxford 1991, pp. 175–222.

[Mer] L. Merel. *Intersections sur les courbes modulaires.* Manuscripta Math., 80 (1993), 283–289.

[O'S] C. O'Sullivan. *Properties of Eisenstein series formed with modular symbols.* J. reine u. angew. Math., 518 (2000), 163–186.

[Sch] F. Schweiger. *Ergodic theory of fibred systems and metric number theory.* Clarendon Press, 1995.

[Soi] Y. Soibelman. *Quantum tori, mirror symetry and deformation theory.* Preprint math.QA/0011162

[Wa] M.S. Waterman, *Remarks on invariant measures for number theoretic transformations.* Monatshefte für Mathematik, 79 (1975) 157–163.

[Za] D. Zagier. *Vassiliev invariants and a strange identity related to the Dedekind eta–function.* Preprint MPIM 99–78.