PERIOD FUNCTIONS AND THE SELBERG ZETA FUNCTION FOR THE MODULAR GROUP

John Lewis and Don Zagier

Framingham State College
Department of Mathematics
Framingham, MA 01701

USA

Max-Planck-Institut
für Mathematik
Gottfried-Claren-Str. 26
53225 Bonn

Germany

MPI 96-112

PERIOD FUNCTIONS AND THE SELBERG ZETA FUNCTION FOR THE MODULAR GROUP

JOHN LEWIS AND DON ZAGIER

The Selberg trace formula on a Riemann surface X connects the discrete spectrum of the Laplacian with the length spectrum of the surface, that is, the set of lengths of the closed geodesics of on X. The connection is most strikingly expressed in terms of the Selberg zeta function, which is a meromorphic function of a complex variable s that is defined for $\Re(s) > 1$ in terms of the length spectrum and that has zeros at all $s \in \mathbb{C}$ for which $s(1-s)$ is an eigenvalue of the Laplacian in $L^2(X)$. We will be interested in the case when X is the quotient of the upper half-plane $\mathcal{H}$ by either the modular group $\Gamma_1 = \mathrm{SL}(2,\mathbb{Z})$ or the extended modular group $\Gamma = \mathrm{GL}(2,\mathbb{Z})$, where $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$ acts on $\mathcal{H}$ by $z \mapsto (az+b)/(cz+d)$ if $\det(\gamma) = +1$ and $z \mapsto (a\bar{z}+b)/(c\bar{z}+d)$ if $\det(\gamma) = -1$. In this case the length spectrum of X is given in terms of class numbers and units of orders in real quadratic fields, while the eigenfunctions of the Laplace operator are the non-holomorphic modular functions usually called Maass wave forms. (Good expositions of this subject can be found in [6] and [7]).

A striking fact, discovered by D. Mayer [4, 5] and for which a simplified proof will be given in the first part of this paper, is that the Selberg zeta function $Z_\Gamma(s)$ of $\mathcal{H}/\Gamma$ can be represented as the (Fredholm) determinant of the action of a certain element of the quotient field of the group ring $\mathbb{Z}[\Gamma]$ on an appropriate Banach space. Specifically, let $\mathbf{V}$ be the space of functions holomorphic in $\mathbb{D} = \{z \in \mathbb{C} \mid |z-1| < \frac{3}{2}\}$ and continuous in $\overline{\mathbb{D}}$. The semigroup $\{\gamma \in \Gamma \mid \gamma(\mathbb{D}) \subseteq \mathbb{D}\}$ acts on the right by $\pi_s\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) f(z) = (cz+d)^{-2s} f\left(\frac{az+b}{cz+d}\right)$. In particular, for all $n \geq 1$ the element $\left(\begin{smallmatrix} 0 & 1 \\ 1 & n \end{smallmatrix}\right)$, which can be written in terms of the generators $\sigma = \left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)$ and $\rho = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 1 \end{smallmatrix}\right)$ of Γ as $\sigma^{n-1}\rho$, acts on $\mathbf{V}$. It turns out (cf. §2) that the formal expression

$$\mathcal{L} = (1-\sigma)^{-1}\rho = \sum_{n=1}^{\infty} \left[\begin{smallmatrix} 0 & 1 \\ 1 & n \end{smallmatrix}\right] \tag{1}$$

defines an operator $L_s = \pi_s(\mathcal{L})$ of trace class on $\mathbf{V}$ (first for $\Re(s) > \frac{1}{2}$, and then by analytic continuation to all s). This implies that the operator $1 - L_s$ has a determinant in the Fredholm sense; and the result then is:

Theorem 1. *The Selberg zeta function of $\mathcal{H}/\Gamma$ is given by*

$$Z_\Gamma(s) = \det\bigl(1 - L_s\bigr)\,. \tag{2}$$

(Actually, Mayer's result is that the Selberg zeta function of $\mathcal{H}/\Gamma_1$ equals $\det(1 - L_s^2)$, but everything works in much the same way for the two groups Γ and Γ_1. We will discuss both cases, but in our exposition have given precedence to the larger group Γ.)

On the other hand, as we already mentioned, the function $Z_\Gamma(s)$ has a meromorphic continuation with zeros corresponding to the eigenvalues of even Maass wave forms on $SL(2,\mathbb{Z})$. Formally, equation (2) says that these zeros correspond to the fixed points of L_s, i.e., to the functions $h \in \mathbf{V}$ such that $h(z) = \sum_{n=1}^{\infty}(z+n)^{-2s}h(1/(z+n))$. Adding $z^{-2s}h(1/z)$ to both sides we find that $h(z) + z^{-2s}h(1/z) = h(z-1)$, or equivalently, that the shifted function $\psi(z) = h(z-1)$ satisfies the *three-term functional equation*

$$\psi(z) = \psi(z+1) + z^{-2s}\,\psi(1+1/z)\,. \tag{3}$$

It is therefore natural to ask whether there is a direct connection between the spectrum of the Laplace operator Δ on $\mathcal{H}/\Gamma$ and the solutions of the three-term functional equation. Such a connection was discovered (independently of Mayer's work) in [2], whose main result, in a slightly strengthened form, can be stated as follows:

Theorem 2. *Let s be a complex number with $0 < \Re(s) < 1$. Then there is a canonical bijection between square integrable solutions of $\Delta u = s(1-s)\,u$ in $\mathcal{H}/\Gamma$ and holomorphic solutions of* (3) *in the cut plane $\mathbb{C}' = \mathbb{C} \setminus (-\infty, 0]$ satisfying the growth condition $\psi(x) = \mathrm{O}(1/x)$ as $x \to \infty$.*

The formula for the correspondence $u \mapsto \psi$ in [2] was completely explicit (eq. (12) below), but its proof was indirect and did not make the reasons for its properties at all transparent. Other proofs and several other formulas for ψ in terms of u were found in [3], where it was also observed that this correspondence is exactly analogous to the relationship between a holomorphic modular form and its period polynomial in the sense of Eichler, Shimura, and Manin. We will call the function $\psi(z)$ the *period function* of the wave form u.

Taken together, these two theorems give another point of view on the Selberg trace formula: Theorem 1 relates the "length spectrum" definition of the Selberg zeta function to the fixed points of the operator L_s and hence, by implication, to the solutions of the functional equation (3), and Theorem 2 relates the solutions of (3) to the "discrete spectrum of the Laplacian" definition of Z_Γ. In this paper (which, except for the simplifications in the proof of Theorem 1, is mostly expository) we will discuss both aspects. Part I uses reduction theory to establish the connection between the Selberg zeta function and the operator L_s. In §1 we outline a proof of Theorem 1. The details (e.g. the proofs of various assertions needed from reduction theory, verification of convergence, etc.) are filled in in §2, while the following section gives various complements: the modifications when Γ is replaced by Γ_1, a reformulation of some of the ideas of the proof in terms of group algebras, and a brief description of Mayer's original approach via the symbolic dynamics of the continued fraction map. Part II describes the connection between the solutions of the functional equation (3) and the eigenfunctions of the Laplacian in $\mathcal{H}/\Gamma$. We will give several formulas for the $u \leftrightarrow \psi$ correspondence, sketch some the ideas involved in the proof, describe the analogy with the theory of periods of modular forms, and discuss some other properties of solutions of (3) on $\mathbb{C}'$ or on $\mathbb{R}^+$. Here we will give fewer details than in Part I and omit all proofs, referring the reader to [2] and [3] for more information.

§1. The formal calculation. The basic conjugacy invariants of an element $\gamma \in \Gamma$ are the numbers $\mathrm{Tr}(\gamma)$, $\det(\gamma)$, and $\Delta(\gamma) = \mathrm{Tr}(\gamma)^2 - 4\det(\gamma)$ (trace, determinant, discriminant). We will call γ *hyperbolic* if $\Delta(\gamma)$is positive and (to distinguish between γ and $-\gamma$, which act in the same way on $\mathcal{H}$) also $\mathrm{Tr}(\gamma) > 0$. If γ is hyperbolic, we set

$$\mathcal{N}(\gamma) = \left(\frac{\mathrm{Tr}(\gamma) + \Delta(\gamma)^{\frac{1}{2}}}{2}\right)^2, \qquad \chi_s(\gamma) = \frac{\mathcal{N}(\gamma)^{\frac{1}{2}-s}}{\Delta(\gamma)^{\frac{1}{2}}} = \frac{\mathcal{N}(\gamma)^{-s}}{1 - \det(\gamma)\mathcal{N}(\gamma)^{-1}} \qquad (s \in \mathbb{C})$$

and define $k(\gamma)$ as the largest integer k such that $\gamma = \gamma_0^k$ for some $\gamma_0 \in \Gamma$ (which is then hyperbolic and *primitive*, i.e. $k(\gamma_0) = 1$). The Selberg zeta function $Z_\Gamma(s)$ for Γ is defined by

$$Z_\Gamma(s) = \prod_{\substack{\{\gamma\} \text{ in } \Gamma \\ \gamma \text{ primitive}}} \prod_{m=0}^{\infty} \left(1 - \det(\gamma)^m \mathcal{N}(\gamma)^{-s-m}\right) \qquad (\Re(s) > 1),$$

where the notation "$\{\gamma\}$ in Γ" means that the product is taken over all (primitive hyperbolic) elements of Γ up to Γ-conjugacy. That the function $Z_\Gamma(s)$ extends meromorphically to all complex values of s is one of the standard consequences of the Selberg trace formula, which expresses its logarithmic derivative as a sum over the eigenvalues of the (hyperbolic) Laplacian in $\mathcal{H}/\Gamma$.

For $\Re(s) > 1$ we have the simple computation

$$\begin{aligned} -\log Z_\Gamma(s) &= \sum_{\substack{\{\gamma\} \text{ in } \Gamma \\ \gamma \text{ primitive}}} \sum_{m=0}^{\infty} \sum_{k=1}^{\infty} \frac{1}{k} \det(\gamma)^{km} \mathcal{N}(\gamma)^{-k(s+m)} \\ &= \sum_{\substack{\{\gamma\} \text{ in } \Gamma \\ \gamma \text{ primitive}}} \sum_{k=1}^{\infty} \frac{1}{k} \frac{\mathcal{N}(\gamma)^{-ks}}{1 - \det(\gamma)^k \mathcal{N}(\gamma)^{-k}} \\ &= \sum_{\substack{\{\gamma\} \text{ in } \Gamma \\ \gamma \text{ hyperbolic}}} \frac{1}{k(\gamma)} \chi_s(\gamma), \end{aligned} \tag{4}$$

where the last step just expresses the fact that every hyperbolic element of Γ can be written uniquely as γ^k with γ primitive and $k \geq 1$.

To get further we use a version of reduction theory. This theory is usually presented for quadratic forms, but is translatable into the language of matrices by the standard observation that there is a 1:1 correspondence between conjugacy classes of matrices of trace t and determinant n and equivalence classes of quadratic forms of discriminant $t^2 - 4n$. We define the set of *reduced* elements of Γ by

$$\mathsf{Red} = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma \;\middle|\; 0 \leq a \leq b, c \leq d \right\},$$

i.e. matrices with non-negative entries which are non-decreasing downwards and to the right. (We will explain in §3B where this definition comes from.) Then we have the following facts, whose proofs will be indicated in §2:

(I) Every reduced matrix can be written uniquely as a product $\left(\begin{smallmatrix}0&1\\1&n_1\end{smallmatrix}\right)\cdots\left(\begin{smallmatrix}0&1\\1&n_\ell\end{smallmatrix}\right)$ with $n_1,\dots,n_\ell \geq 1$ for a unique positive integer $\ell = \ell(\gamma)$.

(II) Every conjugacy classes of hyperbolic matrices in Γ contains reduced representatives γ; they all have the same value of $\ell(\gamma)$ and there are $\ell(\gamma)/k(\gamma)$ of them.

(III) If γ is reduced, then the operator $\pi_s(\gamma)$ is of trace class and $\operatorname{Tr}\pi_s(\gamma) = \chi_s(\gamma)$.

Combining these assertions with (4), we find

$$\begin{aligned} -\log Z_\Gamma(s) &\underset{(II)}{=} \sum_{\gamma\in\mathsf{Red}} \frac{1}{\ell(\gamma)}\chi_s(\gamma) \\ &\underset{(III)}{=} \operatorname{Tr}\left(\sum_{\gamma\in\mathsf{Red}} \frac{1}{\ell(\gamma)}\pi_s(\gamma)\right) \\ &\underset{(I)}{=} \operatorname{Tr}\left(\sum_{\ell=1}^{\infty}\frac{1}{\ell}\left(\sum_{n=1}^{\infty}\pi_s\left(\begin{smallmatrix}0&1\\1&n\end{smallmatrix}\right)\right)^{\ell}\right), \end{aligned} \tag{5}$$

and this is equivalent to (2) by the definition of $\mathcal{L}$ and the Fredholm determinant formula $\log\det(1-L_s) = -\sum_{\ell=1}^{\infty}\operatorname{Tr}(L_s^\ell)/\ell$.

§2. Details. In this section we verify the assertions (I)–(III) and establish the validity of the formal calculations of §1 for $\Re(s) > 1$; (2) then holds for all s by analytic continuation.

A. Proof of (I). Suppose that $\gamma = \left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right) \in \mathsf{Red}$. If $a = 0$ then $\gamma = \left(\begin{smallmatrix}0&1\\1&d\end{smallmatrix}\right)$ with $d \geq 1$ and we are already finished with $\ell(\gamma) = 1$. If $a > 0$, we set $n = \lceil d/b\rceil - 1$ (i.e. n is the unique integer $n < d/b \leq n+1$). One easily checks that this is the only $n \in \mathbb{Z}$ for which $\gamma = \left(\begin{smallmatrix}0&1\\1&n\end{smallmatrix}\right)\gamma^*$ with $\gamma^* \in \mathsf{Red}$. Moreover, the sum of the entries of γ^* is smaller than that of γ, so we can assume by induction that γ^* has the form claimed, and then so does γ with $\ell(\gamma) = \ell(\gamma^*) + 1$.

B. Proof of (II). This is essentially equivalent to the theory of periodic continued fractions: to each hyperbolic matrix $\gamma = \left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right)$ we associate the roots $\frac{a-d\pm\sqrt{\Delta(\gamma)}}{2c}$ of $\gamma x = x$, which are quadratic irrationalities; two γ's are conjugate if and only if the corresponding roots are Γ-equivalent; each quadratic irrationality has a continued fraction expansion $1/(m_1 + 1/(m_2 + 1/\cdots))$ which is eventually periodic; and if the fixed point of γ has a continued fraction expansion with period $(n_1,\dots,n_\ell)$ then γ is conjugate to the reduced matrix $\left(\begin{smallmatrix}0&1\\1&n_1\end{smallmatrix}\right)\cdots\left(\begin{smallmatrix}0&1\\1&n_\ell\end{smallmatrix}\right)$ with $n_1,\dots,n_\ell \geq 1$. However, one can also do the reduction procedure directly on the matrix level. We define a conjugacy class preserving map F from the set of hyperbolic matrices to itself by $\gamma = \left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right) \mapsto F(\gamma) = \left(\begin{smallmatrix}0&1\\1&n\end{smallmatrix}\right)^{-1}\gamma\left(\begin{smallmatrix}0&1\\1&n\end{smallmatrix}\right)$ where n is the unique integer for which the interval $[n, n+1]$ contains both d/b and c/a. (This definition must be modified slightly if $a = 0$.) Notice that if γ is reduced then this is the same n as was used in the proof of (I) and $F(\gamma)$ is simply $\gamma^*\left(\begin{smallmatrix}0&1\\1&n\end{smallmatrix}\right)$ in the notation

above. The effect of F on a reduced matrix $\gamma = \left(\begin{smallmatrix} 0 & 1 \\ 1 & n_1 \end{smallmatrix}\right) \cdots \left(\begin{smallmatrix} 0 & 1 \\ 1 & n_\ell \end{smallmatrix}\right)$ is thus simply to replace it by the cyclically permuted product $F(\gamma) = \left(\begin{smallmatrix} 0 & 1 \\ 1 & n_2 \end{smallmatrix}\right) \cdots \left(\begin{smallmatrix} 0 & 1 \\ 1 & n_\ell \end{smallmatrix}\right)\left(\begin{smallmatrix} 0 & 1 \\ 1 & n_1 \end{smallmatrix}\right)$. It is clear that under this "internal conjugation" the exact period of γ will be the number $\ell(\gamma)/k(\gamma)$. (*Proof.* If γ is the kth power of another matrix γ' with $k \geq 1$, then γ' is also reduced and hence also a product of matrices $\left(\begin{smallmatrix} 0 & 1 \\ 1 & n \end{smallmatrix}\right)$, and the cycle $(n_1, \ldots, n_\ell)$ for γ is just the k-fold concatenation of the cycle for γ'; and conversely if the cycle of γ is a k-fold concatenation then γ is a kth power. Hence $\ell(\gamma)/k(\gamma)$ is the exact period of the sequence of integers $\{n_i\}$.) The assertion of (II) is thus proved if we show that (i) iterating F often enough eventually sends an arbitrary hyperbolic element of Γ to an element of Red, and (ii) two elements of Red are Γ-conjugate only if they are already "internally" conjugate, i.e., if and only if one is mapped to the other by a power of F. Both steps are proved by a series of elementary inequalities which show that each application of F "improves things" (i.e. either makes a non-reduced matrix more nearly reduced in the sense that some positive integer measuring the failure of the inequalities defining Red gets smaller, or else reduces the size of the entries of the matrix conjugating one reduced γ into another). We omit the details, which are exactly parallel to the proofs of the corresponding assertions in the usual reduction theory of quadratic forms as carried out in standard books, e.g. in §13 of [8].

C. Proof of (III). If γ is reduced, then γ maps the closed interval $[-\frac{1}{2}, \frac{5}{2}]$ into the half-open interval (0,2] and hence maps the closed disk $\overline{\mathbb{D}}$ into the open disk $\mathbb{D}$. Standard results from the theory of composition operators on spaces of holomorphic functions (cf. [5], Thm. 7.9 and Lemma 7.10 and the papers cited there) then imply that the operator $\pi_s(g)$ is of trace class and that its trace equals $\chi_s(g)$.

D. Verification of convergence. The operator $L_s = \pi_s(\mathcal{L})$ sends $h \in \mathbf{V}$ to

$$(L_s h)(z) = \sum_{n=1}^{\infty} \frac{1}{(z+n)^{2s}}\, h\Big(\frac{1}{z+n}\Big)\,.$$

Since h is holomorphic at 0, the sum converges absolutely for s in the half-plane $\Re(s) > \frac{1}{2}$ to a function which again belongs to $\mathbf{V}$, and the absolute convergence also implies that this operator is of trace class. We have to show that in the smaller half-plane $\Re(s) > 1$ all of the steps of the calculations in (4) and (5) are justified. But this follows from the calculations themselves: The absolute convergence of the product defining $Z_\Gamma(s)$ (and hence of the sum defining its logarithm) for $\Re(s) > 1$ is well-known, and since in (4) and (5) all terms are replaced by their absolute value when s is replaced by its real part, the various interchanges in the order of summation are automatically justified. The validity of the last line of the proof also follows, since the formula $\sum \mathrm{Tr}(A^\ell)/\ell = -\log\det(1-A)$ is true for any trace class operator A for which $\sum |\mathrm{Tr}(A^\ell)/\ell|$ converges. We can also run the calculation backwards (and hence verify the convergence of the infinite product for $Z_\Gamma(s)$ in the half-plane $\Re(s) > 1$) by showing directly that the sum $M_s := \sum_{\gamma \in \mathsf{Red}} |\chi_s(\gamma)|$ is convergent for $\Re(s) > 1$. Indeed, we have

$$M_s = \sum_{k \geq 3} \frac{c_k^+}{\sqrt{k^2-4}} \left(\frac{k+\sqrt{k^2-4}}{2}\right)^{1-2\Re(s)} + \sum_{k \geq 1} \frac{c_k^-}{\sqrt{k^2+4}} \left(\frac{k+\sqrt{k^2+4}}{2}\right)^{1-2\Re(s)},$$

with $c_k^{\pm} = \#\{\gamma \in \mathsf{Red} \mid \mathrm{Tr}(\gamma) = k,\ \det(\gamma) = \pm 1\}$, and the required convergence follows from the estimate $c_k^{\pm} \ll k^{1+\varepsilon}$ $(\forall \varepsilon > 0)$, which is obtained by straightforward estimates using the divisor function.

§3. Complements. In this section we discuss some further aspects of the proof given in the last two sections.

A. The Selberg zeta function for $SL(2,\mathbb{Z})$. In this subsection we treat the case when the group $\Gamma = GL(2,\mathbb{Z})$ is replaced by its subgroup $\Gamma_1 = SL(2,\mathbb{Z})$, the usual modular group. We denote by $Z(s)$ the Selberg zeta function for Γ_1, which is defined for $\Re(s) > 1$ by the same product expansion as before but with the product running over Γ_1-conjugacy classes of primitive hyperbolic elements of Γ_1. As mentioned in the introduction, the statement of Theorem 1 for Γ_1 is the identity

$$Z(s) = \det(1 - L_s^2)\,. \tag{6}$$

We indicate the changes that have to be made in the proof of §§1–2 in order to prove this.

The calculation (4) is unchanged except that now the summation is over Γ_1-conjugacy classes and the number $k(\gamma)$ must be replaced by $k_1(\gamma)$, the largest integer n such that γ is the nth power of an element in Γ_1. For the first line of (5) we needed that

$$\gamma \in \Gamma \quad \Rightarrow \quad \#\{\gamma' \in \mathsf{Red} \mid \gamma' \underset{\Gamma}{\sim} \gamma\} \;=\; \frac{\ell(\gamma)}{k(\gamma)}\,, \tag{7}$$

which followed from Statement (II). This must now be replaced by

$$\gamma \in \Gamma_1 \quad \Rightarrow \quad \#\{\gamma' \in \mathsf{Red} \mid \gamma' \underset{\Gamma_1}{\sim} \gamma\} \;=\; \frac{\ell(\gamma)}{2k_1(\gamma)}\,, \tag{8}$$

which we will prove in a moment. The first line in (5) then becomes

$$-\log Z(s) = \sum_{\gamma \in \Gamma_1 \cap \mathsf{Red}} \frac{2}{\ell(\gamma)}\,\chi_s(\gamma)\,, \tag{9}$$

and the restriction $\gamma \in \Gamma_1$ implies that in the last line of (5) we sum only over even ℓ.

It remains to prove (8). Write $\gamma = \gamma_0^k$ where $k > 0$ and γ_0 is primitive in Γ, and set $\ell_0 = \ell(\gamma_0)$. Then $k(\gamma) = k$, $\ell(\gamma) = k\ell_0$, and (7) expresses the fact that the Γ-conjugates to γ in Red correspond to the ℓ_0 possible "internal conjugates" of a reduced representative of this conjugacy class. We now distinguish two cases. If $\det(\gamma_0) = +1$, then $k_1(\gamma) = k$ (because $\gamma_0 \in \Gamma_1$ and is clearly primitive there), but ℓ_0 is even and the number of $\gamma' \in \mathsf{Red}$ which are Γ_1-conjugate to γ is $\ell_0/2$, because half of the ℓ_0 "internal" conjugations in our cycle are conjugations by elements of determinant -1 and hence are no longer counted. If on the other hand $\det(\gamma_0) = -1$, then k is even and $k_1(\gamma) = k/2$, because the element γ_0^2 is now primitive in Γ_1, but to make up for it the number of $\gamma' \in \mathsf{Red}$ which are Γ_1-conjugate to γ is now the full number ℓ_0, because there is no longer any distinction between internal conjugacies by elements of determinant $+1$ or -1. (Conjugating by the product of the first r matrices $\left(\begin{smallmatrix} 0 & 1 \\ 1 & n \end{smallmatrix}\right)$ of the cycle of γ_0 is the same as conjugating by the product of the last $\ell_0 - r$ of them, and r and $\ell_0 - r$ have opposite parities.) This establishes (8) in both cases.

B. Identities in the group ring. In this subsection we redo part of the calculation in §1 in a slightly different way in which the elements of Red are built up out of powers of a finite rather than an infinite sum; this also helps us to understand the structure of Red and permits us to make sense of the formal expressions in (1).

Recall that $\sigma = \left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)$ and $\rho = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 1 \end{smallmatrix}\right)$. These elements generate Γ, but of course not at all freely: e.g. one has $(\sigma^{-1}\rho)^2 = (\sigma^{-2}\rho^2)^6 = 1$. On the other hand, the subsemigroup Q of Γ generated by ρ and σ is the *free* semigroup generated by these two elements, i.e. its elements are all words in σ and ρ and all such words distinct. In fact, the set Q is contained in the larger sub-semigroup P of Γ consisting of all matrices with non-negative entries, which is easily seen to be the semigroup generated by the two elements $\kappa = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ and σ, subject to the unique relation $\kappa^2 = 1$. Since $\rho = \sigma\kappa$, every element of P is either a word in ρ and σ or else κ times such a word, so $\mathsf{P} = \mathsf{Q} \cup \kappa\mathsf{Q}$ (disjoint union). This says that $\mathsf{Q} \smallsetminus \{1\} = \sigma\mathsf{P}$, the subset of P consisting of words in κ and σ which begin with a σ, or equivalently of matrices $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ satisfying $c \geq a \geq 0$, $d \geq b \geq 0$. In turn, the subset of Q consisting of words in σ and ρ which end in a ρ is the subset of those elements satisfying the additional inequalities $b \geq a \geq 0$, $d \geq c \geq 0$, i.e. precisely our set Red. This shows again that the elements of Red are uniquely expressible as products of the matrices $\sigma^{n-1}\rho = \left(\begin{smallmatrix} 0 & 1 \\ 1 & n \end{smallmatrix}\right)$ with $n \geq 1$. We define $\ell(\gamma)$ for any $\gamma \in \mathsf{Q}$ as the number of ρ's in the representation of γ as a word in ρ and σ; this agrees with our previous definition on the subset $\mathsf{Red} = \mathsf{Q}\rho$.

Let Q_n be the subset of Q consisting of words in ρ and σ of length n. The recursive description $\mathsf{Q}_0 = \{1\}$ and $\mathsf{Q}_{n+1} = \mathsf{Q}_n\sigma \cup \mathsf{Q}_n\rho$ implies the identity $\sum_{\gamma\in\mathsf{Q}_n}[\gamma] = \big([\sigma]+[\rho]\big)^n$ in the group ring $\mathbb{Z}[\Gamma]$. More generally, if we introduce a variable v and define

$$\mathcal{K}_v = [\sigma] + v\,[\rho] = \left[\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right] + v\left[\begin{smallmatrix} 0 & 1 \\ 1 & 1 \end{smallmatrix}\right] \in \mathbb{Z}[\Gamma][v]\,,$$

then we have $\mathcal{K}_v{}^n = \sum_{\gamma\in\mathsf{Q}_n} v^{\ell(\gamma)}\,[\gamma]$. On the other hand, $\mathsf{Q} = \bigcup_{n=0}^{\infty}\mathsf{Q}_n$, so to deal with all of Q (or Red) we must work with infinite sums of elements of Γ. In particular, let

$$\mathcal{L}_w \;=\; (1 - w\,\sigma)^{-1}\rho \;=\; \sum_{n=1}^{\infty} w^{n-1}\left[\begin{smallmatrix} 0 & 1 \\ 1 & n \end{smallmatrix}\right].$$

This reduces to our previous formal expression $\mathcal{L}$ at $w = 1$, but now makes sense as an element in the ring $\mathbb{Z}[\Gamma][[w]]$ of formal power series in one variable over the group ring $\mathbb{Z}[\Gamma]$, or as an element of $\mathbb{C}[\Gamma]$ if $w \in \mathbb{C}$, $|w| < 1$. Then we have the identities

$$\mathcal{K}_v{}^{n-1}\,[\rho] \;=\; \sum_{\substack{\gamma\in\mathsf{Red}\\ n(\gamma)=n}} v^{\ell(\gamma)-1}\,[\gamma] \qquad \text{and} \qquad \mathcal{L}_w{}^{\ell} = \sum_{\substack{\gamma\in\mathsf{Red}\\ \ell(\gamma)=\ell}} w^{n(\gamma)}\,[\gamma]\,,$$

where $n(\gamma)$ for $\gamma \in \mathsf{Q}$ denotes the length of γ as a word in σ and ρ. Combining, we get

$$\big(1 - w\,\mathcal{K}_v\big)^{-1}[\rho] \;=\; \mathcal{L}_w\,\big(1 - vw\,\mathcal{L}_w\big)^{-1} = \sum_{\gamma\in\mathsf{Red}} v^{\ell(\gamma)-1}\,w^{n(\gamma)-1}\,[\gamma]\,.$$

Integrating with respect to v gives the identity

$$-\log\bigl(1-vw\,\mathcal{L}_w\bigr) \;=\; \sum_{\gamma\in\mathsf{Red}} \frac{v^{\ell(\gamma)}}{\ell(\gamma)}\, w^{n(\gamma)}\,[\gamma]\,,$$

and the content of §1 can now be summarized by saying that we computed $-\log Z_\Gamma(s)$ as the trace of π_s of this sum on $\mathbf{V}$ in the limit $v=w=1$.

C. The Selberg zeta function and the dynamics of the Gauss map. The proof of equation (6) given originally by Mayer is parallel in many ways to the one given above, but was expressed in terms of ideas coming from symbolic dynamics. Specifically, he used the connection between closed geodesics on $\mathcal{H}/\Gamma_1$ and periodic continued fractions to relate the Selberg zeta function to the dynamics of the "continued fraction map" (Gauss map) $F:[0,1)\to[0,1)$ which maps x to the fractional part of $1/x$ (and, say, to 0 if $x=0$). We give a very brief outline of the argument.

To a "dynamical system" $F:X\to X$ and a weight function $h:X\to\mathbb{C}$ one associates for each integer $n\geq 1$ a *partition function*

$$Z_n(F,h)=\sum_{x\in X,\ F^n x=x} h(x)\,h(Fx)\,h(F^2x)\cdots h(F^{n-1}x)$$

(sum over n-periodic points). In our case, $X=[0,1)$, F is the continued fraction map, and we take for $h(x)$ the function $h_s(x)=x^{2s}$ where $s\in\mathbb{C}$ with $\Re(s)>\frac{1}{2}$ (to make the series defining Z_n converge). Using the technique of "transfer operators" and Grothendieck's theory of nuclear operators, Mayer shows that

$$Z_n(F,h_s)=\mathrm{Tr}(L_s^n)-(-1)^n\,\mathrm{Tr}(L_{s+1}^n)\qquad(\forall n\geq 0). \tag{10}$$

On the other hand, the definition of the Selberg zeta function can be written $Z(s)=\prod_{k=0}^{\infty}\zeta_{SR}(s+k)^{-1}$, where $\zeta_{SR}(s)$ (the letters "SR" stand for Smale-Ruelle) is defined as the product over all closed primitive geodesics in $\mathcal{H}/\Gamma_1$ of $(1-e^{-\lambda s})$, λ being the length of the geodesic. The connection between closed geodesics and periodic continued fractions leads to the equation $\zeta_{SR}(s)=\exp\bigl(\sum_{n=1}^{\infty}\frac{1}{n}Z_{2n}(F,h_s)\bigr)$. (Here only even indices occur because the map $x\mapsto x^{-1}-m$ implicit in the definition of F corresponds to a matrix of determinant -1, so that only even iterates of F correspond to the action of elements of Γ_1.) Together with (10) and the determinant formula $\exp\bigl(-\sum_{n=1}^{\infty}\frac{1}{n}\mathrm{Tr}(A^n)\bigr)=\det(1-A)$ this gives $\zeta_{SR}(s)=\dfrac{\det(1-L_{s+1}^2)}{\det(1-L_s^2)}$ and hence finally $Z(s)=\det(1-L_s^2)$.

A similar proof, of course, works also for equation (2), but now using all the $Z_n(F,h_s)$. This version of Mayer's theorem was developed by Efrat [1]. To connect this to the discussion in **A** above, we rewrite (9) slightly as

$$-\log Z(s)=\sum_{\gamma\in\mathsf{Red}}\frac{1+\det(\gamma)}{\ell(\gamma)}\,\chi_s(\gamma)\,,$$

or equivalently as the factorization $Z(s) = Z_+(s) \cdot Z_-(s)$ where $Z_+(s) = \sum_{\mathsf{Red}} \chi_s(\gamma)/\ell(\gamma)$ and $Z_-(s) = \sum_{\mathsf{Red}} \det(\gamma)\chi_s(\gamma)/\ell(\gamma)$. The first factor is $Z_\Gamma(s)$ by the calculation in Sections 1 and 2, so its zeros correspond to even Maass wave forms, while the zeros of the second factor $Z_-(s)$ correspond to the odd wave forms. See §5B for more on this.

Part II. Period functions of Maass wave forms

§4. Various descriptions of the period correspondence. We explained in the introduction how the identity (2) should lead one to expect some sort of correspondence between eigenfunctions of the Laplacian in $\mathcal{H}/\Gamma$ and holomorphic solutions of the three-term functional equation (3). In this section we give several descriptions of this "period correspondence," each of which puts into evidence certain of its properties. There does not seem to be any single description which exhibits all aspects of the correspondence simultaneously.

We first recall some basic facts about Maass wave forms and fix terminology. The Maass wave forms for the modular group $\Gamma_1 = \mathrm{SL}(2,\mathbb{Z})$ are the non-constant Γ_1-invariant eigenfunctions of the hyperbolic Laplacian $\Delta = -y^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)$ which are square-integrable on the modular surface $\mathcal{H}/\Gamma_1$. The space of these forms breaks up under the action of the involution $\iota : z \mapsto -\bar{z}$ into the spaces of even (invariant) and odd (anti-invariant) forms. In particular, the even forms are the eigenfunctions of Δ on $\mathcal{H}/\Gamma$, since Γ is generated by Γ_1 and ι. We will always use the letter u to denote a Maass form and the letter s for its *spectral parameter*, i.e. for the complex number s such that the eigenvalue of u under Δ is $s(1-s)$. (Note that the number $1-s$ is an equally good spectral parameter for u, but to describe the period correspondence $u \leftrightarrow \psi$ we must fix the choice of s since the functional equation (3) depends on s. However, this dependence is very simple because it is known that s always has real part $\frac{1}{2}$ and hence $1-s = \bar{s}$, so that the map $\psi(z) \mapsto \overline{\psi(\bar{z})}$ maps the space of solutions of (3) for one choice of s to the corresponding choice for the other.) The invariance of u under the translation map $T : z \mapsto z+1$ and the conjugation map ι implies that $u(x+iy)$ has a cosine expansion with respect to x, and the square-integrability of u and differential equation $\Delta u = s(1-s)\,u$ imply that this expansion has the form

$$u(x+iy) = \sqrt{y}\,\sum_{n=1}^{\infty} a_n\, K_{s-\frac{1}{2}}(2\pi n y)\,\cos(2\pi n x)\,, \tag{11}$$

where K_ν is a modified Bessel function.

A. Description of the period correspondence via integral transforms. A number of formulas for the period correspondence $u \leftrightarrow \psi$ were given in [2]. A particularly direct one is the integral formula

$$\psi(z) = z\int_0^\infty \frac{t^s\,u(it)}{(z^2+t^2)^{s+1}}\,dt \qquad (\Re(z) > 0)\,. \tag{12}$$

This was obtained after a number of intermediate steps. One of the most striking is that there is an an entire function $g(w)$ which is related to u by

$$g(\pm 2\pi i n) = \tfrac{1}{2}\,(2\pi n)^{-s+1/2}\,a_n \qquad (n = 1,\,2,\,3,\,\ldots) \tag{13}$$

(i.e., g is a "holomorphic interpolation" of the Fourier coefficients of u) and to ψ by

$$g^{(k)}(0) = \frac{1}{\Gamma(2s+k)}\,\psi^{(k)}(1) \qquad (k = 0,\, 1,\, 2,\, \ldots) \tag{14}$$

(so that the Taylor coefficients of g at 0 and ψ at 1 determine each other). The function g in turn is obtained from another intermediate function ϕ which is defined by

$$\phi(w) = w^{1-s} \int_0^\infty \sqrt{wt}\, J_{s-\frac{1}{2}}(wt)\, u(it)\, dt \tag{15}$$

(Hankel transform) and defines ψ by

$$\psi(z) = \int_0^\infty \phi(w)\, w^{2s-1}\, e^{-zw}\, dw \tag{16}$$

(Laplace transform). Substituting (15) into (16) gives (12), while substituting the Fourier expansion (11) into (15) and integrating term by term leads to the formula

$$\phi(w) = w \sum_{n=1}^{\infty} \frac{(2\pi n)^{-s+1/2}\, a_n}{w^2 + (2\pi n)^2}\,.$$

In particular, $\phi(w)$ has simple poles of residue $\frac{1}{2}(2\pi n)^{-s+1/2} a_n$ at $w = \pm 2\pi i n$ and no other poles, so the function $g(w) := (1 - e^{-w})\phi(w)$ is entire and satisfies (13), while on the other hand, once one has proved that $\psi(z)$ satisfies the three-term functional equation (3) one immediately gets

$$\int_0^\infty g(zw)\, w^{2s-1}\, e^{-w}\, dw = z^{-2s}\left[\psi\big(z^{-1}\big) - \psi\big(z^{-1}+1\big)\right] = \psi(z)\,,$$

and (14) follows easily. No single one of these formulas permits one to deduce in a direct way the properties of $\psi(z)$ (i.e., the analytic continuability to $\mathbb{C}' = \mathbb{C} \smallsetminus (-\infty, 0]$ and the functional equation (3)) from the fact that u is a Maass form, and the proof of this in [2] is quite complex. On the other hand, they do give explicit ways to get from u to ψ and back: the forward direction is given by (12), while (13) and (14) determine the Fourier coefficients of u as special values of the power series $g(w) = \sum_k \psi^{(k)}(1) w^k / k!\, \Gamma(2s+k)$.

We refer to [2] and [3] for a more detailed discussion of these ideas and of other related approaches, including one based on a summation formula of Ferrar and another in terms of the Helgason automorphic boundary form of u, which are also important aspects of the story and provide useful perspectives.

B. Description in terms of Fourier expansions. The integral representation (12) makes visible the analyticity of $\psi(z)$ in a neighborhood of the positive real axis, but does not make it clear why ψ satisfies the three-term functional equation. In [3] a different description of ψ was given in which the functional equation becomes obvious and the key point is the continuability of ψ across the positive real axis. The starting point is the following simple algebraic fact.

Lemma. *If $\psi : \mathbb{C} \smallsetminus \mathbb{R} \to \mathbb{C}$ is any function satisfying the three-term functional equation* (3) *then the function $f : \mathcal{H} \to \mathbb{C}$ defined by*

$$f(z) = \psi(z) + e^{-2\pi i s}\, \psi(-z) \tag{17}$$

is 1-periodic (i.e. T-invariant). Conversely, if $f : \mathcal{H} \to \mathbb{C}$ is any 1-periodic function, then the function $\psi : \mathbb{C} \setminus \mathbb{R} \to \mathbb{C}$ defined by

$$\psi(z) = \begin{cases} f(z) - z^{-2s} f(-1/z) & \text{if } \Im(z) > 0 \\ z^{-2s} f(1/z) - f(-z) & \text{if } \Im(z) < 0 \end{cases} \tag{18}$$

satisfies the functional equation (3). *Moreover, if $s \notin \mathbb{Z}$ then the correspondences* (17) *and* (18) *between 1-periodic functions in $\mathcal{H}$ and solutions of* (3) *in $\mathbb{C} \setminus \mathbb{R}$ are inverse maps to each other up to a non-zero scalar factor $1 - e^{-2\pi i s}$.*

Then we have the following very elegant description of the period correspondence.

Theorem 3. *Let u be an even Maass wave form with spectral parameter s and Fourier expansion given by* (11), *and $f : \mathcal{H} \to \mathbb{C}$ the 1-periodic holomorphic function defined by*

$$f(z) = \sum_{n=1}^{\infty} n^{s-\frac{1}{2}} a_n e^{2\pi i n z} \qquad (z \in \mathcal{H}) . \tag{19}$$

Then the function ψ defined by (18) *extends holomorphically from $\mathbb{C} \setminus \mathbb{R}$ to $\mathbb{C}'$ and is bounded in the right half-plane. Conversely, if s is a complex number with $\Re(s) > 0$, $\psi : \mathbb{C}' \to \mathbb{C}$ a holomorphic solution of* (3) *which is bounded in the right half-plane, $f : \mathcal{H} \to \mathbb{C}$ the 1-periodic holomorphic function defined by* (17), *and $\{a_n\}$ the coefficients defined by the Fourier expansion* (19), *then the function $u : \mathcal{H} \to \mathbb{C}$ defined by the Fourier series* (11) *is an even Maass wave form with spectral parameter s.*

The proof of this theorem, given in [3], relies essentially on the properties of L-series. It is well-known that the L-series $L(\rho) = \sum_n a_n/n^\rho$ of a Maass wave form has a holomorphic extension to all complex values of ρ and satisfies a functional equation under $\rho \mapsto 1 - \rho$, and conversely that these properties of the coefficients a_n imply the Γ-invariance of the function u defined by (11). The L-series can be represented as the Mellin transform of the restrictions to the imaginary axis of either u or f (with different gamma-factors). We can now use the inverse Mellin transform to write the function ψ defined by (18) in the uppper and lower half-planes as integral transforms of $L(\rho)$, and the functional equation of L turns out to be just what is needed in order that these two formulas agree and define a holomorphic function in all of $\mathbb{C}'$. Conversely, if u is defined by (11) and f by (19) for some coefficients a_n (satisfying a growth condition), and if ψ is the function defined by (18), then the Mellin transforms of the restrictions of ψ to the positive and the negative imaginary axes are both linear combinations of $L(\rho)$ and $L(1-\rho)$. Now if ψ extends holomorphically across $\mathbb{R}_+$ and satisfies the growth condition, we can rotate the two paths of integration to $\mathbb{R}_+$, and the equality of these two linear combinations then gives the functional equation of the L-series.

This argument makes clear which properties of u correspond to which properties of ψ: if $\{a_n\}$ is any collection of coefficients (of not too rapid growth), then the function u defined by (11) is a T-invariant eigenfunction of the Laplacian with eigenvalue $s(1-s)$, while the function ψ defined by (19) and (18) is a holomorphic solution of the functional equation (3) in the upper and lower half-planes; this gives a bijection between translation-invariant

even eigenfunctions of Δ and functions ψ on $\mathbb{C} \smallsetminus \mathbb{R}$ satisfying (3), and under this bijection the eigenfunctions which are invariant under $z \mapsto -1/z$ correspond to the functions ψ which extend holomorphically across the positive real axis.

C. Unfolding from the positive real axis. In this subsection we state a result from [3] to the effect that the restriction map from the space of holomorphic solutions of (3) in $\mathbb{C}'$ to the space of analytic solutions of (3) on $\mathbb{R}_+$, which is obviously injective, is in fact bijective under suitable growth conditions. This complements the results of the two preceding subsections: in **A** we described how to get from u to $\psi|\mathbb{R}_+$ via an integral transform and how to get the Fourier coefficients of u from the Taylor expansion of ψ at $1 \in \mathbb{R}_+$, and in **B** we explained how to get the values of ψ off the real axis from u and vice versa.

Theorem 4. *Let s be a complex number with $\Re(s) > 0$. Then any bounded real-analytic solution of the functional equation* (3) *on the positive real axis extends to a holomorphic solution of* (3) *in the whole cut plane which is bounded in the right half-plane.*

The proof of this theorem is by a kind of "bootstrapping": by repeated applications of the functional equation (3) one successively extends ψ to larger and larger neighborhoods of $\mathbb{R}_+ \subset \mathbb{C}'$, while preserving the growth conditions. In fact, the growth conditions can be relaxed, e.g. if $\Re(s) = \frac{1}{2}$ then the assumption $\psi(x) = o(1/x)$ as $x \to 0$ already implies that ψ continues to a holomorphic function in $\mathbb{C}'$ and is bounded in $\Re(z) > 0$, which together with Theorem 3 implies that s is the spectral parameter associated to a Maass wave form. This is especially surprising because it turns out that *any* smooth solution of the functional equation on $\mathbb{R}_+$ is $O(1/x)$ as $x \to 0$ and that these solutions form an uncountable-dimensional vector space for any s, whereas the Maass forms exist only for special values of s and then form a finite-dimensional space.

§5. Complements. In the final section of the paper we give various examples of solutions of the functional equation (3), especially the polynomial solutions for negative integral values of s which give the link to the classical theory of periods of modular forms, and also indicate the changes that must be made when Γ is replaced by its subgroup Γ_1.

A. Examples and equivalent forms of the three-term functional equation. If we relax the growth conditions on the function ψ, then there are many more solutions of the functional equation (3). For example, an infinite class of solutions for any s is given by $\psi(z) = f(z) + z^{-2s} f(1/z)$ for any odd and 1-periodic entire function f. There are also more interesting examples which nearly satisfy the growth conditions of Theorem 3 and which correspond to the zeros of the Selberg zeta function other than the spectral parameters coming from Maass wave forms. These zeros occur at $s = 1$ and at the zeros of $\zeta(2s)$, where ζ is the Riemann zeta function (cf. [7], pp. 48-49). The solution of (3) for $s = 1$ is given by $\psi(z) = 1/z$. The solutions corresponding to the trivial zeros of $\zeta(2s)$ at $s = -1, -2, \ldots$ will be discussed in the next section. The solutions corresponding to the non-trivial zeros arise as follows. For $\Re(s) > 1$ define

$$\psi_s(z) = \zeta(2s)\big(1 + z^{-2s}\big) + 2 \sum_{m,n \geq 1} (mz + n)^{-2s},$$

a kind of "half-Eisenstein-series." The series converges absolutely and it is easy to check that it satisfies the functional equation (3). On the other hand, the shifted function $h_s(z) = \psi_s(z+1)$ is *not* a fixed point of the Mayer operator L_s; instead, as one checks in a straightforward way, one has $(L_s h_s)(z) = h_s(z) - \zeta(2s)$. It follows that the (easily obtained) analytic continuation of h_s gives a fixed point of (the analytic continuation of) L_s at the zeros of $\zeta(2s)$.

We also mention two equivalent forms of the period functional equation, as a sample of the algebraic character of the theory. The first is the equation

$$\psi(z) = (z+1)^{-2s}\big[\,\psi\big(\frac{z}{z+1}\big) + \psi\big(\frac{1}{z+1}\big)\,\big]\,.$$

Written in the language of the group algebra $\mathbb{Z}[\Gamma]$, this says that $\pi_s(\mathcal{K})\psi = \psi$, where $\mathcal{K}$ is the element $\mathcal{K}_1 = [\sigma] + [\rho]$ of §3**B** and is related to the Mayer element $\mathcal{L}$ by the equation $\mathcal{L}(1-\mathcal{L})^{-1} = (1-\mathcal{K})^{-1}[\rho]$. The second says that ψ is fixed by the operator $\pi_s\big(\sum_{n\geq 0}[\rho^n\sigma]\big)$. Written out, this is the infinitely-many-term functional equation

$$\psi(z) = \sum_{n=1}^{\infty} \frac{1}{(F_n\,z + F_{n+1})^{2s}}\,\psi\left(\frac{F_{n-2}\,z + F_{n-1}}{F_n\,z + F_{n+1}}\right)$$

where $\{F_n\}$ are the Fibonacci numbers. Note that this series, unlike the one defining the Mayer operator L_s, is rapidly convergent if $\Re(s) > 0$ and $\Re(z) > -(1+\sqrt{5})/2$.

B. Even and odd Maass wave forms. We now consider the modular group Γ_1 instead of Γ. As mentioned at the beginning of §4, the Maass wave forms for Γ_1 break up into two kinds, the even ones (which are invariant under the map $u(z) \mapsto u(-\bar{z})$ and hence under all of Γ) and the odd ones (for which $u(z) = -u(-\bar{z})$). The spectral parameters corresponding to both kinds of Maass forms are zeros of the Selberg zeta function $Z(s)$ of Γ_1, with the ones corresponding to even forms being zeros of $Z_\Gamma(s)$. On the other hand, as we saw in Part I, $Z_\Gamma(s)$ is the determinant of the operator $1 - L_s$, while $Z(s)$ is the determinant of $1 - L_s^2 = (1-L_s)(1+L_s)$. The odd Maass forms should therefore correspond to the solutions in $\mathbf{V}$ of $L_s h = -h$ and hence, after the same shift $\psi(z) = h(z-1)$ as in the even case, to the solutions of the *odd three-term functional equation*

$$\psi(z) = \psi\big(1+z\big) - z^{-2s}\,\psi\big(1+\frac{1}{z}\big)\,, \tag{20}$$

instead of the even functional equation (3). This is in fact true and, as one would expect, the description and properties of this "odd period correspondence" are very similar to those in the even case. The Fourier cosine expansion (11) is naturally replaced by the corresponding sine series. The integral transform (12), which must obviously be modified since $u(iy)$ is now identically zero, is replaced by

$$\psi(z) = \int_0^\infty \frac{t^s\,u_x(it)}{(z^2+t^2)^s}\,dt \qquad (\Re(z) > 0)\,,$$

where $u_x = \dfrac{\partial u}{\partial x}$ $(z = x + iy)$. The algebraic correspondence described in the Lemma in §4**B** is true with appropriate sign changes (change the sign of the second term in (17) and of both terms in the second line of (18)), and Theorem 3 then holds *mutatis mutandum.*

Examples of non-Maass solutions of the odd functional equation are the function $1-z^{-2s}$ (or more generally $f(z)-z^{-2s}f(1/z)$ with f even and 1-periodic) for all s and $\psi(z)=\log z$ for $s=0$. The example $\psi_s(z)$ discussed in Subsection **A** has no odd analogue. (The analogous fact about Selberg zeta functions is that all the zeros of $Z_-(s)=Z(s)/Z_\Gamma(s)$ correspond to the odd spectral parameters, whereas the zeros of $Z_\Gamma(s)$ correspond both to even Maass forms and to zeros of the Riemann zeta function.) The two alternate forms of the even functional equation given at the end of**A** have the obvious odd analogues (replace $\mathcal{K}_{s,1}$ by $\mathcal{K}_{s,-1}$ and $\sum_{n\geq 0}\psi|_{\rho^n\sigma}$ by $\sum_{n\geq 0}(-1)^n\psi|_{\rho^n\sigma}$)).

Finally, one can give a uniform description of the period functions associated to Maass forms, without separating into the even and odd cases. These functions should correspond to the fixed points of L_s^2 on V, and this leads (after the usual shift $\psi(z)=h(z-1)$) to the "master functional equation"

$$\psi(z)=\psi\big(z+1\big)+(z+1)^{-2s}\,\psi\big(\frac{z}{z+1}\big)\,. \tag{21}$$

We will call a solution of (21) a *period function.* Since the involution $\psi(z)\mapsto z^{-2s}\psi(1/z)$ preserves this equation, every period function decomposes uniquely into an even (invariant) and odd (anti-invariant) part, and one checks easily that the even and odd period functions are precisely the solutions of (3) or (20), respectively. The description of the period correspondence given in §4**B** is now modified as follows. Any 1-periodic eigenfunction of Δ with eigenvalue $s(1-s)$ has a Fourier expansion of the form $u(x+iy)=\sqrt{y}\sum_{n\neq 0}a_n\,K_{s-\frac12}(2\pi|n|y)e^{2\pi inx}$. We then define a 1-periodic holomorphic function f on $\mathbb{C}\smallsetminus\mathbb{R}$ by two different Fourier series, using the a_n with $n>0$ in the upper half-plane and the a_n with $n<0$ in the lower half-plane. In each half-plane there is a 1:1 correspondence between the space of 1-periodic functions and the space of solutions of (21) given by the (up to a scalar factor, inverse) transformations

$$f(z)\mapsto\psi(z):=f(z)-z^{-2s}\,f(-1/z)\,,\qquad \psi(z)\mapsto f(z):=\psi(z)+z^{-2s}\,\psi(-1/z)\,.$$

Then, just as before, the invariance of u under $z\mapsto -1/z$ is equivalent (under suitable growth conditions) to the analytic continuability of $\psi(z)$ across the positive real axis.

C. Integral values of s and classical period theory. Let s be a negative integer, which we write in the form $1-k$ with $k\geq 2$. The factor z^{-2s} in the master functional equation (21) (or in its even or odd versions (3) or (20)) now becomes a monomial and we can look for polynomial solutions ψ, which we will then call *period polynomials.* The degree of such a polynomial must be $\leq 2k-2$, so the problem of finding all solutions for a given k is just a matter of finite linear algebra. For $k=2$, 3, 4 and 5 we find that the only polynomial solution is $z^{2k-2}-1$ (which is an odd polynomial but an even period function), but for $k=6$ there are three linearly independent solutions $z^{10}-1$, $z^8-3z^6+3z^4-z^2$, and $4z^9-25z^7+42z^5-25z^3+4z$. This has to do with the fact that for $k=6$ the space S_{2k} of cusp forms of weight $2k$ on the modular group has a non-trivial element for the first time, namely the discriminant function

$$\Delta(z)=e^{2\pi iz}\prod_{n=1}^{\infty}\big(1-e^{2\pi inz}\big)=\sum_{n=1}^{\infty}\tau(n)\,e^{2\pi inz}\,.$$

Associated to this cusp form is its *Eichler integral* $\tilde{\Delta}(z) = \sum_n n^{-11}\tau(n)e^{2\pi inz}$, which is not quite modular (of weight $2-2k=-10$) but instead satisfies $(cz+d)^{10}\tilde{\Delta}\left(\frac{az+b}{cz+d}\right) = \tilde{\Delta}(z) + P_\gamma(z)$ for any $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_1$ with P_γ a polynomial of degree ≤ 10, and the 3-dimensional space of period polynomials is generated by the odd and even parts of the polynomial P_γ for $\gamma = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$, together with the polynomial $z^{10}-1$. In general, if one associates to any cusp form $f(z) = \sum A_n e^{2\pi inz} \in S_{2k}$ its Eichler integral $\tilde{f}(z) = \sum n^{-2k+1}A_n e^{2\pi inz}$, then the difference $\tilde{f}(z) - z^{2k-2}\tilde{f}(-1/z)$ is a polynomial $P = P_f$ of degree $\leq 2k-2$ which satisfies the *period conditions*

$$P(z) + z^{2k-2}P\left(\frac{-1}{z}\right) = P(z) + z^{2k-2}P\left(1-\frac{1}{z}\right) + (z-1)^{2k-2}P\left(\frac{1}{1-z}\right) = 0\,,$$

and the period theory of Eichler, Shimura and Manin tells us that this space has dimension $2\dim S_{2k-2}+1$ and is spanned by $z^{2k-2}-1$ and by the even and odd parts of the polynomials P_f. But an elementary calculation shows that polynomials satisfying the period conditions are precisely the polynomial solutions of (21) with $s = 1-k$ (and further that this space breaks up into the direct sum of its subspaces of odd and even polynomials and that these are precisely the polynomial solutions of (3) and of (20) respectively). This fits in very well with our picture since it is known that $s = 1-k$ is a zero of $Z_\Gamma(s)$ of multiplicity $\delta_k := \dim S_{2k}$ and a zero of $Z(s)$ of multiplicity $2\delta_k + 1$. What's more, one can get directly from cusp forms of weight $2k$ to nearly Γ_1-invariant eigenfunctions of the Laplace operator with eigenvalue $k(1-k)$. For instance, the eigenfunction defined by (11) with $s=-5$ and $a_n = \tau(n)/n^{11/2}$ is not only invariant under the translation T and the reflection ι, but is nearly invariant under $z \mapsto -1/z$, the difference $u(-1/z) - u(z)$ being a polynomial in x, y and $1/y$ with coefficients which are closely related to those of the odd period polynomial $4z^9 - 25z^7 + 42z^5 - 25z^3 + 4z$ above.

References

[1] I. Efrat, *Dynamics of the continued fraction map and the spectral theory of* $SL(2,\mathbf{Z})$, Invent. Math. **114** (1993), 207–218.

[2] J.B. Lewis, *Spaces of Holomorphic Functions equivalent to the even Maass Cusps Forms*, Invent. Math., to appear.

[3] J.B. Lewis and D. Zagier, *Period functions for Maass wave forms*, in preparation.

[4] D. Mayer, *The Thermodynamic formalism Approach to Selberg's Zeta function for* $PSL(2,\mathbf{Z})$, Bull. AMS **25** (1991), 55–60.

[5] D. Mayer, *Continued fractions and related transformations*, in "Ergodic Theory, Symbolic Dynamics and Hyperbolic Spaces," T. Bedford, M. Keane, C. Series (Eds.), Oxford University Press, Oxford 1991, pp. 175–222.

[6] A. Terras, "Harmonic Analysis on Symmetric Spaces and Applications," Vol. I, Springer, New York 1985.

[7] A.B. Venkov, "Spectral Theory of Automorphic Functions", Kluwer, Dordrecht 1990.

[8] D. Zagier, "Zetafunktionen und Quadratische Körper", Springer-Verlag, Berlin 1981.