Selberg's zeta function for PSL(2,$\mathbb{Z}$) via the thermodynamic formalism for the continued fraction map

Dieter H. Mayer

Max–Planck–Institut für Mathematik
Gottfried–Claren–Str. 26
D–5300 Bonn FRG

MPI/90-86

I. Introduction

There exist basically two different approaches to Selberg's zeta function for cocompact Fuchsian groups: the original one dating back to Selberg [S] and proceeding essentially via the trace formula [V], [E], and a more recent one, based on the thermodynamic formalism [R4] and the dynamical zeta functions of Smale [Sm] and Ruelle [R2] for flows on compact manifolds, which for geodesic flows on surfaces of constant negative curvature (c.n.c.) are closely related to Selberg's zeta function [R3].

The problem we adress in the present paper is an extension of the latter approach to surfaces of c.n.c. with finite area and hence to Selberg's function for cofinite Fuchsian groups.

The standard procedure in the thermodynamic formalism approach to the Smale–Ruelle zeta function for uniformly hyperbolic flows on compact manifolds [F] is to start with a Markov partition [Si], [B1] and to construct symbolic dynamics [B1] for the flow: thereby the flow gets described in terms of a much simpler one, built essentially from its Poincaré map and the recurrence times with respect to the Markov partition, where the Poincaré map is finally described by a subshift of finite type [B2], [B3]. The Smale–Ruelle function appears then as some kind of generating function for partition functions of a lattice spin system of classical statistical mechanics [R3]. This allows the transfer operator method to be applied [R1], [R4], [M2], [M4] so that for analytic systems the Smale–Ruelle function can finally be expressed in terms of Fredholm determinants of such transfer operators.

For geodesic flows on compact surfaces of c.n.c. the above approach can be made rather explicit thanks to the work of Bowen and Series [BS] on the boundary expansions at infinity for such flows. These are piecewise analytic, expanding Markov maps with well studied transfer operators [M4] which Pollicott used also recently in his approach to the Selberg function for cocompact Fuchsian groups [Po2].

In trying to extend his results to general cofinite Fuchsian groups one faces the problem that for groups with parabolic elements the Bowen–Series boundary maps are not anymore expanding [BS] and hence their transfer operators not of trace class [M4] (which was an essential point in the whole approach).

For the modular group PSL $(2,\mathbb{Z})$ and the modular surface $M_m = \mathbb{H}/PSL(2,\mathbb{Z})$ ($\mathbb{H}$ the upper halfplane model of hyperbolic 2–space), the solution for the above problem follows from results of Series [Se] respectively Adler and Flatto [AF] extending earlier work of Artin [A]. These authors have shown that the nonexpanding Bowen–Series map can be replaced for PSL $(2,\mathbb{Z})$ by an induced map on the unit interval of the real line which is again expanding. Their real astonishing result indeed is that this map can be chosen the well known continued fraction map $T_G x = \frac{1}{x} \bmod 1$ whose relation to the geodesic flow on M_m was recognized already by Artin [A] in the thirties. From the work in [BS] one should expect something similar to happen for general cofinite Fuchsian groups and the geodesic flows on the corresponding surfaces, even if explicit expressions as for PSL $(2,\mathbb{Z})$ are not to expect in general.

For PSL $(2,\mathbb{Z})$ hence the thermodynamic formalism can be applied and leads to a representantion of its Selberg function as a product of two Fredholm determinants of transfer operators of the continued fraction map. The explicit form of these operators allows a detailed investigation of their analyticity properties leading to a new proof of the analytic properties of the Selberg function. The above representation leads also to a new formulation of Riemann's hypothesis in terms of the above transfer operators.

II. Transfer operators and Ruelle zeta functions for the Gauss map

The Gauss or continued fraction map $T_G : I \longrightarrow I$, $I = [0,1]$ is defined as

$$T_G x = \begin{cases} x^{-1} \bmod 1 & x \neq 0 \\ 0 & x = 0 . \end{cases} \tag{1}$$

It belongs to the class of piecewise analytic, expanding Markov maps with a well developed theory for their transfer operators [M4]. The local branches $T_n x = x^{-1} - n$ of T_G map the intervals $I_n = \left[\frac{1}{n+1} , \frac{1}{n} \right]$ onto I for every $n \in \mathbb{N}$ and have inverses $\psi_n(x) = \frac{1}{x+n}$ which extend to the disc $D = \left\{ z \in \mathbb{C} : |z-1| < \frac{3}{2} \right\}$ and map D holomorphically into some compactum in D [M1].

Denote then by $Z_n(T_G, A)$ the partition functions [R3] for T_G

$$Z_n(T_G, A) = \sum_{x \in \mathrm{Fix}\, T_G^n} \exp \sum_{k=0}^{n-1} A(T_G^k x) \tag{2}$$

with $\mathrm{Fix}\, T_G^n$ the n–periodic points of T_G in I and $A : I \longrightarrow \mathbb{C}$ some function such that the infinite sums in (2) converge for any $n \in \mathbb{N}$. The elements of $\mathrm{Fix}\, T_G^n$ can be characterized through their continued fraction expansions being n–periodic: $x \in \mathrm{Fix}\, T_G^n \Leftrightarrow x = [\, \overline{m_1, \ldots, m_n} \,]$ for some $m_i \in \mathbb{N}$. The only function $A : I \longrightarrow \mathbb{C}$ of interest for the following is

$$A_s(x) = -s \log | T_G'(x) | = s \log x^2 \tag{3}$$

and its analytic extension to D , where s is some complex parameter known in statistical mechanics as "inverse temperature". The special role the function A_1 plays already in the ergodic theory of T_G is well known [R4]. The partition functions $Z_n(T_G, A_s)$ are well defined for all s with $\mathrm{Re}\, s > \frac{1}{2}$ and can be expressed in terms of transfer operators $\mathcal{L}_s$ acting on the Banach space $A_\infty(D)$ of holomorphic functions on D which are continuous on $\bar{D}$. They are defined as [M1, M3]

$$\mathcal{L}_s f(z) = \sum_{n=1}^{\infty} \exp A_s(\psi_n(z)) \; f \circ \psi_n(z) = \sum_{n=1}^{\infty} \left(\frac{1}{z+n}\right)^{2s} f\left(\frac{1}{z+n}\right). \tag{4}$$

In [M3] we proved

Proposition 1 The operators $\mathcal{L}_s : A_\infty(D) \longrightarrow A_\infty(D)$ are for $\mathrm{Re}\, s > \frac{1}{2}$ nuclear operators of order zero [G] and fulfill the trace formulas

$$Z_n(T_G, A_s) = \mathrm{trace}\; \mathcal{L}_s^n - \mathrm{trace}\,(-\mathcal{L}_{s+1})^n .$$

Applying these trace formulas to Ruelle's zeta functions [R2] $\zeta_k(z, A_s)$ for T_G with

$$\zeta_k(z, A_s) = \exp \sum_{n=1}^{\infty} \frac{z^n}{n} Z_{nk}(T_G, A_s), \; k \in \mathbb{N}, \tag{5}$$

we find for $|z| < e^{-kP(A_s)}$ with

$$P(A_s) = \lim_{n\to\infty} \frac{1}{n} \log Z_n(T_G, A_s) \tag{6}$$

the topological pressure of A_s under T_G :

Corollary 1 The Ruelle zeta functions $\zeta_k(z, A_s)$ can be written for $\mathrm{Re}\, s > \frac{1}{2}$ as $\zeta_k(z, A_s) = \frac{\det(1-z(-\mathcal{L}_{s+1}^{\;k}))}{\det(1-z\;\mathcal{L}_s^{\;k})}$, and extend there as meromorphic functions into the entire complex z–plane. For fixed $z \in \mathbb{C}$ the functions are meromorphic in the half plane $\mathrm{Re}\, s > \frac{1}{2}$.

Of special interest for us are the functions $\zeta_k(1, A_s)$ which we denote by ζ_k :

$$\zeta_k(s) := \zeta_k(1, A_s) . \tag{7}$$

Their analyticity properties in s follow from Corollary 1 and

Proposition 2: [M3] The map $s \longrightarrow \mathcal{L}_s$ extends as a meromorphic function into the entire complex s–plane whose values are nuclear operators of order zero in $A_\infty(D)$. It has simple poles at the points $s_k = \frac{1-k}{2}$, $k = 0,1,2,\ldots$ with residues the rank 1 operators $N_k : A_\infty(D) \longrightarrow A_\infty(D)$ with $N_k f(z) = \frac{1}{2}\frac{1}{k!} f^{(k)}(0)$.

From this follows immediately [M3]:

Corollary 2: The Fredholm determinants $\det(1 \pm \mathcal{L}_s^k)$ extend as meromorphic functions into the entire s–plane with (possibly removable) singularities at $s_k = \frac{1-k}{2}$, $k = 0,1,\ldots$.

The proof uses Grothendieck's Fredholm theory for nuclear operators [G].

III. The Selberg zeta function for PSL (2,Z)

For a general flow $\phi_t : M \longrightarrow M$ the dynamical Smale–Ruelle zeta function is defined through the length spectrum $L(\phi_t)$ characterising ϕ_t's periodic orbits and their prime periods $l(\gamma)$ [Sm], [R2]:

$$\zeta_{SR}(s) = \prod_{\gamma} (1 - e^{-s\, l(\gamma)})^{-1} \tag{8}$$

For uniformly hyperbolic flows [P] the above product is known to converge for all s with $\mathrm{Re}\, s > h_{top}(\phi_t)$ (= topological entropy of ϕ_t). For ϕ_t the geodesic flow on a surface of c.n.c. the function ζ_{SR} determines the Selberg zeta function $Z(s)$ for the corresponding Fuchsian group Γ through the relation [R3]

$$Z(s) = \prod_{\gamma} \prod_{k=0}^{\infty} (1 - e^{-(s+k) l(\gamma)} = \prod_{k=0}^{\infty} \zeta_{SR}(s+k)^{-1} \tag{9}$$

For $\Gamma = PSL(2,\mathbf{Z})$ respectively the corresponding modular surface M_m the geodesic flow and its length spectrum are closely related to a special flow ψ_t built over the natural extension $\tilde{T}_G : I \times I \times \mathbf{Z}_2 \longrightarrow I \times I \times \mathbf{Z}_2$ of the map $T_G : I \to I$ [Se], [AF], [Po1] with

$$\tilde{T}_G(x,y,\varepsilon) = (T_G x, \frac{1}{y+[1/x]}, -\varepsilon)\ , \ \varepsilon = \pm 1\ . \tag{10}$$

Hence $L(\phi_t)$ coincides with $L(\psi_t)$ which can be simply described in terms of the periodic points of $\tilde{T}_G$ [Po1], [P]

$$L(\psi_t) = \left\{ \sum_{k=0}^{2r-1} A_1(T_G{}^k x) : [(x,y,\varepsilon)] \in \mathrm{Fix}_p(\tilde{T}_G{}^{2r}) \, , \, r \in N \right\} \tag{11}$$

where $\mathrm{Fix}_p(\tilde{T}_G{}^{2r})$ denotes the equivalence classes of periodic points of $\tilde{T}_G$ of prime period $2r$, where two points are equivalent if they belong to the same orbit under $\tilde{T}_G$. The function A_1 was defined in (3).

The Smale–Ruelle function ζ_{SR} for the special flow ψ_t with length spectrum (11) can be written as [P]

$$\zeta_{SR}(s) = \exp \sum_{n=1}^{\infty} \frac{z^n}{n} Z_n(\tilde{T}_G, \bar{A}_s) \tag{12}$$

with $\bar{A}_s : I \times I \times \mathbb{Z}_2 \longrightarrow \mathbb{C}$ defined in terms of A_s in (3) as $\bar{A}_s(x,y,\varepsilon) = A_s(x)$. Since $\mathrm{Fix}\, \tilde{T}_G{}^n = \phi$ for n odd we find from definition (2) of the partition functions Z_n :

$$Z_n(\tilde{T}_G, \bar{A}_s) = \begin{cases} 0 & n \text{ odd} \\ 2\, Z_n(T_G, A_s) & n \text{ even} \end{cases} \tag{13}$$

and hence

$$\zeta_{SR}(s) = \zeta_2(1, A_s) = \zeta_2(s) \tag{14}$$

with ζ_2 as defined in (7).

Inserting finally (14) into relation (9) and using Corollary 1 for the case $z = 1$ we get

Proposition 3 The Selberg zeta function $Z(s)$ for $PSL(2,\mathbb{Z})$ has in the half plane $\mathrm{Re}\, s > \frac{1}{2}$ the representation

$$Z(s) = \det(1-\mathscr{L}_s)\det(1+\mathscr{L}_s) = \det(1-\mathscr{L}_s^2)$$

where $\mathscr{L}_s$ is the transfer operator for T_G defined in (4), and is holomorphic there.

The above representation allows now a simple meromorphic extension into the entire complex s–plane: we only have to apply Corollary 2 which gives meromorphic extensions for the corresponding Fredholm determinants.

Corollary 3 The Selberg zeta function for $PSL(2,\mathbb{Z})$ is meromorphic in the entire s–plane and has there the representation $Z(s) = \det(1-\mathscr{L}_s)\det(1+\mathscr{L}_s)$.

It would be interesting to extend Patterson's recent approach to determine the divisor of $Z(s)$ for cocompact groups through the transfer operator also to the present case [Pa].

Let us finally combine the representation found for $Z(s)$ in Corollary 3 with classical results derived from Selberg's trace formula for $PSL(2,\mathbb{Z})$ [V]. The nontrivial zeros s_n of $Z(s)$ corresponding obviously to s values $s \neq s_k = \frac{1-k}{2}$, $k = 0,1,\ldots$, for which $\mathscr{L}_s$ has $\lambda = +1$ or $\lambda = -1$ as an eigenvalue, fall into two classes: either s_n is 1/2 times a nontrivial zero of Riemann's zeta function $\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$, or s_n corresponds via the formula $s_n = \frac{1}{2} + i\, r_n$ to the discrete eigenvalue $\varrho_n = \frac{1}{4} + r_n^{\,2}$ of the Laplace–Beltrami operator

$-\Delta_{M_m}$ on M_m. The value $s_0 = 1$ is obviously such a non–trivial zero of $Z(s)$ since the operator $\mathscr{L}_1$, which is nothing else than the Perron–Frobenius operator of T_G, has $\lambda = 1$ as a simple eigenvalue [M1]. It corresponds to the "small" eigenvalue $\rho = 0$ of $-\Delta_{M_m}$. It is then tempting to conjecture that the factorization
$Z(s) = \det(1-\mathscr{L}_s)\det(1+\mathscr{L}_s)$ is analogous to the one found for cocompact groups in [Sa], [Vo]: in this case the factor $\det(1-\mathscr{L}_s)$ would describe the eigenvalues of $-\Delta_{M_m}$ and $\det(1+\mathscr{L}_s)$ the non trivial zero's of $\zeta(s)$. If this is true the Riemann Hypothesis on $\zeta(s)$ could be expressed in terms of $\mathscr{L}_s$: RH $\Leftrightarrow$ $\mathscr{L}_s$ has eigenvalue $\lambda = -1$ only on the line $\mathrm{Re}\, s = \frac{1}{4}$.

One could then also speculate if the operator $\mathscr{L}_s$ is the one Hilbert proposed to look for to prove the RH [Su]. Certainly, he primarily had in mind a selfadjoint operator in some Hilbert space, and indeed, the transfer operator $\mathscr{L}_s$ can be considered for $s \neq s_k$ the analytic extension of selfadjoint operators on the real axis $s > \frac{1}{2}$ [M3].

Literature

[AF] R. Adler, L. Flatto: Cross section maps for the geodesic flow on the modular surface. Contemp. Math. 26, 9 – 24, AMS, Providence, R.I. (1984).

[A] E. Artin: Ein mechanisches System mit quasi–ergodischen Bahnen. Collected Papers, Addison–Wesley, Reading, Mass. (1965) 499 – 504.

[B1] R. Bowen: Symbolic dynamics for hyperbolic flows. Am. J. Math. 95 (1972) 429 – 459.

[B2] R. Bowen: "Equilibrium states and the ergodic theory of Anosov diffeomorphisms". Lect. Notes in Math. 470, Springer Verlag, Berlin 1975.

[B3] R. Bowen: On Axion–A diffeomorphisms. CBMS Reg. Conf. 35, AMS, Providence, R.I. 1978.

[BS] R. Bowen, C. Series: Markov maps associated to Fuchsian groups. IHES Publ. Math. 50 (1979), 153–170.

[E] J. Elstrodt: Die Selberg'sche Spurformel für kompakte Riemann'sche Flächen. Jber. d. Dt. Math. Verein, 83 (1981), 45 – 77.

[F] D. Fried: The zeta functions of Ruelle and Selberg I. Ann. Sci. Ecole Norm. Sup. (4), 19 (1986) 491 – 517.

[G] A. Grothendieck: La theorie de Fredholm. Bull. Soc. Math. France 84 (1956) 319 – 384.

[M1] D. Mayer: On a zeta function related to the continued fraction transformation. Bull. Soc. Math. France 104 (1976) 195 – 203.

[M2] D. Mayer: "The Ruelle–Araki transfer operator in classical statistical mechanics". Lecture Notes in Physics 123, Springer Verlag, Berlin 1980.

[M3] D. Mayer: On the thermodynamic formalism for the Gauss map. Commun. Math. Phys. 130 (1990) 311 – 333.

[M4] D. Mayer: Continued fractions and related transformations. In Proc. Topical Meeting on Hyperbolic Geometry and Ergodic Theory. ICTP Trieste, ed. C. Series, Cambridge Univ. Press (1990).

[P] W. Parry: An analogue of the prime number theorem for closed orbits of shifts of finite type and their suspensions. Isr. J. Math. 45 (1983) 41 – 52.

[Pa] S. Patterson: On Ruelle's zeta function. Mathematica Gottingensis, Heft 36 (1989).

[Po1] M. Pollicott: Distribution of closed geodesics on the modular surface and quadratic irrationals. Bull. Soc. Math. France 114 (1986) 431 – 446.

[Po2] M. Pollicott: Some applications of the thermodynamic formalism to manifolds of constant negative curvature. To appear in Adv. in Math.

[R1] D. Ruelle: Statistical mechanics of a one dimensional lattice gas. Commun. Math. Phys. 9 (1968) 267 – 278.

[R2] D. Ruelle: Zeta functions for expanding maps and Anosov flows. Invent. Math. 34 (1976) 231 – 242.

[R3] D. Ruelle: Zeta functions and statistical mechanics. Asterisque 40 (1976) 167 – 176.

[R4] D. Ruelle: "Thermodynamic Formalism". Addison–Wesley, Reading, Mass. 1978.

[Sa] P. Sarnak: Determinants of Laplacians. Commun. Math. Phys. 110 (1987). 113 – 120.

[S] A. Selberg: Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series. J. Ind. math. Soc. 20 (1956) 47 – 87.

[Se] C. Series: The modular surface and continued fractions. J. Lond. Math. Soc. (2) 31 (1985) 69 – 80.

[Si] Y. Sinai: Construction of Markov partitions. Funk. Anal. Appl. 2 (1968) 70 – 80.

[Sm] S. Smale: Differentiable dynamical systems. Bull. AMS 73 (1967) 747 – 817.

[Sn] T. Sunada: Fundamental groups and Laplacians, in "Geometry and Analysis on Manifolds". Ed. T. Sunada, Lecture Notes in Math. 1339, Springer Verlag, Berlin 1988.

[V] A. B. Venkov: Spectral theory of automorphic functions, the Selberg zeta function, and some problems of analytic number theory and mathematical physics. Russ. Math. Surveys 34 (3) (1979) 79 – 153.

[Vo] A. Voros: Spectral functions, special functions and the Selberg zeta function. Commun. Math. Phys. 110, (1987) 439 – 465.