Clay Mathematics Proceedings
Volume 8, 2009

Lectures on Modular Symbols

Yuri I. Manin

Abstract. In these lecture notes, written for the Clay Mathematics Institute Summer School "Arithmetic Geometry", Göttingen 2006, I review some classical and more recent results about modular symbols for $SL(2)$, including arithmetic motivations and applications, an iterated version of modular symbols, and relations with the "non–commutative boundary" of the modular tower for elliptic curves.

1. Introduction: arithmetic functions and Dirichlet series

1.1. Arithmetic functions. Many basic questions of number theory involve the behavior of *arithmetic functions*, i.e. sequences of integers $\{a_n \,|\, n \geq 1\}$ defined in terms of divisors of n, or numbers of solutions of a congruence modulo n, etc. After having chosen such a function, one might ask for example:

(i) Is $\{a_n \,|\, n \geq 1\}$ multiplicative, that is, does $a_{mn} = a_m a_n$ for $(m, n) = 1$?

(ii) What is the asymptotic behavior of $\sum_{n \leq N} a_n$ as $N \to \infty$?

(iii) Can one give a "formula" for a_n if initially it was introduced only by a computational prescription, such as $a_n :=$ *the number of representations of* n *as a sum of four squares*?

A very universal machinery for studying such questions consists in introducing *a generating series* for a_n depending on a complex parameter, and studying the analytic and algebraic properties of this series.

Two classes of series that are used most often are the Fourier series

$$f(z) := \sum_{n=1}^{\infty} a_n e^{2\pi i n z} \tag{1.1}$$

and the Dirichlet series

$$L_f(s) = \sum_{n=1}^{\infty} a_n n^{-s}. \tag{1.2}$$

In full generality, they must be considered as formal series; however, if a_n does not grow too fast, e.g. is bounded by a polynomial in n, then (1.1) converges in the

2000 *Mathematics Subject Classification.* Primary 11F67, Secondary 11F11, 11F75, 11M41.

upper half-plane $H := \{z \in \mathbf{C} \,|\, \mathrm{Im}\, z > 0\}$, whereas (1.2) converges in some right half plane $\mathrm{Re}\, s > D$.

1.2. Mellin transform and modularity. Some of the properties of $\{a_n\}$ are directly encoded in the generating Dirichlet series. For example, multiplicativity of $\{a_n\}$ translates into the existence of an Euler product over primes p:

$$L_f(s) = \prod_p L_{f,p}(s), \qquad L_{f,p}(s) := \sum_{n=1}^{\infty} a_{p^n} p^{-ns}. \tag{1.3}$$

Hence the Dirichlet series for the logarithmic derivative of such a function carries information about the values of a_n restricted to powers of primes. This idea leads to famous "explicit formulas" expressing partial sums of a_{p^n}'s via poles of the logarithmic derivative of $L_f(s)$ i.e. essentially zeroes of $L_f(s)$. Applied to the simplest multiplicative sequence $a_n = 1$ for all n, this formalism produces the classical relationship between primes and zeroes of Riemann's zeta.

It turns out, however, that to establish the necessary analytic properties of $L_f(s)$ such as the analytic continuation in s and a functional equation, and generally even the existence of an Euler product, one should focus first upon the Fourier series $f(z)$. The main reason for this is that interesting functions $f(z)$ more often than not possess, besides the obvious periodicity under $z \mapsto z+1$, additional symmetries, for example, a simple behavior with respect to the substitution $z \mapsto -z^{-1}$. This is the case for $f(z) = \sum_{n\geq 1} e^{2\pi i n^2 z}$ (or the more symmetric $\sum_{n\in\mathbf{Z}} e^{2\pi i n^2 z}$) corresponding to $L_f(s) = \zeta(2s)$.

The transformations $z \mapsto z+1$ and $z \mapsto -z^{-1}$ together generate the full modular group $PSL(2,\mathbf{Z})$ of fractional linear transformations of H, and Fourier series of various *modular forms* with respect to this group and its subgroups of finite index generate a vast supply of most interesting arithmetic functions.

The basic relation between $f(z)$ and $L_f(s)$ allowing one to translate analytic properties of $f(z)$ into those of $L_f(s)$ is the integral *Mellin transform*

$$\Lambda_f(s) := \int_0^{i\infty} f(z) \left(\frac{z}{i}\right)^s \frac{dz}{z}. \tag{1.4}$$

Here the s–th power in the integrand is interpreted as the branch of the exponential function which takes real values for real s and imaginary z. Convergence at $i\infty$ is usually automatic whereas convergence at 0 is justified by a functional equation (possibly after disposing of a controlled singularity).

Whenever we can integrate termwise using (1.1) (for large $\mathrm{Re}\, s$), an easy calculation shows that

$$\Lambda_f(s) = (2\pi)^{-s}\Gamma(s) L_f(s). \tag{1.5}$$

A functional equation for $f(z)$ with respect to $z \mapsto -z^{-1}$ (or more generally, $z \mapsto -(Nz)^{-1}$ for some N) then leads formally to a functional equation of Riemann type connecting $\Lambda_f(s)$ with $\Lambda_f(1-s)$ or $\Lambda_f(D-s)$ for an appropriate D defining *the critical strip* $0 \leq \mathrm{Re}\, s \leq D$ for $L_f(s)$.

This is a very classical story, which acquired its final shape in the work of Hecke in the 1920's and 30's. More modern insights concern the role of Γ–factors as Euler factors at *arithmetic infinity*, and most important, the universality of this picture and the existence of its vast generalizations crystallized in the *Taniyama–Weil conjecture* and the so-called *Langlands program.* This involves, in particular, consideration of much more general arithmetic groups than $PSL(2)$ as modular groups.

We will not discuss this vast development in these lectures and focus upon the classical modular group and related modular symbols. For some generalizations, see [**AB90**], [**AR79**].

2. Classical modular symbols and Shimura integrals

2.1. Modular symbols as integrals. Since we are interested in Mellin transforms of the form (1.4) where $f(z)$ has an appropriate modular behavior with respect to a subgroup of $PSL(2,\mathbf{Z})$, we must keep track of similar integrals taken over $PSL(2,\mathbf{Z})$–images of the upper semi–axis as well. The latter are geodesics connecting two *cusps* in the partial compactification $\overline{H} := H \cup \mathbf{P}^1(\mathbf{Q})$.

Roughly speaking, the *classical modular symbols* are linear functionals (spanned by)

$$\{\alpha,\beta\}\ :\ f \mapsto \int_\alpha^\beta f(z) z^{s-1} dz, \qquad \alpha,\beta \in \mathbf{P}^1(\mathbf{Q})$$

on appropriate spaces of 1–forms $f(z)z^{s-1}dz$. To be more precise, we must recall the following definitions.

The group of real matrices with positive determinant $GL^+(2,\mathbf{R})$ acts on H by fractional linear transformations $z \mapsto [g]z$. Let $j(g,z) := cz + d$ where (c,d) is the lower row of g. Then we have, for any function f on H and homogeneous polynomial $P(X,Y)$ of degree $k-2$,

$$\begin{aligned} g^*[f(z)\,P(z,1)\,dz] &:= f([g]z)\,P([g]z,1)\,d([g]z) \\ &= f([g]z)\,(j(g,z))^{-k} P(az+b, cz+d) \det g\, dz \end{aligned} \tag{2.1}$$

where (a,b) is the upper row of g. From the definition it is clear that the diagonal matrices act identically so that we have in fact an action of $PGL^+(2,\mathbf{R})$.

This action induces for any integer $k \geq 2$ the weight k action of $GL^+(2,\mathbf{R})$ on functions on H. In the literature one finds two different normalizations of such an action. They differ by a determinantal twist and therefore coincide on $SL(2,\mathbf{R})$ and the modular group. For example, in [**Mer94**] and [**Man06**] the action

$$f\,|[g]_k(z) := f([g]z)\,j(g,z)^{-k}\,(\det g)^{k/2} \tag{2.2}$$

is used.

A holomorphic function $f(z)$ on H is a modular form of weight k for a group $\Gamma \subset SL(2,\mathbf{R})$ if $f|[\gamma]_k(z) = f(z)$ for all $\gamma \in \Gamma$ and $f(z)$ is finite at cusps.

Such a form is called a cusp form if it vanishes at cusps.

Let $S_k(\Gamma)$ be the space of cusp forms of weight k. Denote by $Sh_k(\Gamma)$ the space of 1–forms on the complex upper half plane H of the form $f(z)\,P(z,1)\,dz$ where $f \in S_k(\Gamma)$, and $P = P(X,Y)$ runs over homogeneous polynomials of degree $k-2$ in two variables. Thus, the space $Sh_k(\Gamma)$ is spanned by 1–forms *of cusp modular type with integral Mellin arguments in the critical strip* in the terminology of [**Man06**], Def. 2.1.1, and 3.3 below.