Reduction theory and periods of modular forms

Lecture given at the last Manin Seminar on Feburary 21, 2023

PART 0: Introductory Remarks

Today we are paying tribute to Yuri Ivanovich Manin. He was a wonderful person and for me also a wonderful friend and collaborator.

I got to know Yuri in 1987 during two months I spent in Moscow. He was extremely warm right from the beginning and to my great surprise invited me to one of the two 50th birthday parties he gave in his amazing apartment full of innumerable books in many languages. A few years later, and to my great joy, he came to the Max Planck Institute and became my colleague and over the years, with his wife Xenia Glebovna Semenova, also close friend.

I do not plan to say anything today about those years or about the two papers that we wrote together, both of which meant a great deal to me and taught me a lot about doing and thinking about mathematics, often in ways that were often very different from those I knew. Instead, I want to give a purely mathematical lecture about a circle of ideas that originated from work of his and that we often discussed together, although to my regret we never did any joint work in that direction. (However, his first two papers with Matilde, who will speak after me, were on a related subject and were also closely connected with some earlier work of mine and John Lewis.)

This circle idea of ideas originated from Manin's two highly influential papers of 1972 and 1973 on modular symbols and on periods of modular forms of higher weight, both closely related to the Eichler-Shimura cohomology theory that had been developed in the previous decade. In the first part of this lecture I will review these two papers and mention some later results, of myself and others, to which they led. The second part will be concerned with the inverse problem of determining a modular form from its periods. I posed this question to myself more than 25 years ago and found a very nice solution for the case of modular forms on the full modular group, and I had planned to submit this paper to a volume dedicated to Manin on his 60th birthday. But then I discovered experimentally a surprising conjecture in reduction theory, generalizing in an unexpected way the classical theory of continued fractions that had played a key role in both of Manin's papers, that would solve the inversion problem for modular forms on arbitrary Fuchsian groups, and since I did not want to present anything half-baked for such an important occasion, I decided to keep the paper "on ice" until I could prove my conjecture. The same thing happened, to my shame, for the later volumes for his 70th and 80th birthday volumes. But the conjecture still resisted solution and is in fact still not proved even now, although there are partial results. So I will tell this story today, as a kind of very belated present to Yuri.

PART I: Historical Overview

1. Manin's 1972 paper (Izvestia Akademii Nauk SSSR) "Cusps and zeta functions of modular curves"

$E=$ elliptic curve/$\mathbb{Q}$, conductor N, Neron differential ω.

Weil's 1969 conjecture says that there is a map $\varphi: X_0(N) \to E$ (Weil parametrization) such that the pull-back $\varphi^*\omega$ is a multiple of $f(z)\,dz$, where f is the Hecke eigenform of weight 2 corresponding to E ($\leftrightarrow$ "Taniyama-Weil conjecture").

Then $L(E/\mathbb{Q},1) \sim L(f,1) \sim \int_0^\infty f(z)\,dz$. Consider more generally $\{a,b\} = \int_a^b$ (or NOT the integral, but just the abstract class as a relative homology class in $(X_0(N), \text{cusps})$).

Basic idea: introduce the space of formal integer combinations of symbols $\{a,b\}$ with

$$\{a,a\}=0, \quad \{a,b\}+\{b,a\}=0, \quad \{a,b\}+\{b,c\}+\{c,a\}=0, \quad \{\gamma a,\gamma b\}=\{a,b\}\ (\gamma\in\Gamma)$$

and use this as a model of $S_2(\Gamma_0(N))$. (Essentially the same idea was proposed independently by Birch at about the same time.) This was a very important theoretical idea and, with its higher-weight generalizations, is also at the basis of most numerical computations with modular forms done today.

2. Manin's 1973 paper (Matematicheskii Sbornik) "Periods of cusp forms and p-adic Hecke series"

The p-adic part gave rise to a huge literature, starting with papers by Manin–Vishik and Manin–Panchishkin and later also by Hida, Mazur, Coleman and many others, but I'll only talk about the higher-weight part. Here there are two main discoveries:

(i) the algebraic nature of periods of modular forms,
(ii) explicit formulas coming from the action of Hecke operators.

Following Manin, I will illustrate both for the modular form $\Delta(z)$ of weight 12.

Recall first that a modular form of weight k on the full modular group $\Gamma_1 = \mathrm{SL}(2,\mathbb{Z})$, or on any discrete group $\Gamma \subset \mathrm{SL}(2,\mathbb{R})$ of finite covolume, is a holomorphic function $f(z)$ in the upper half-plane $\mathfrak{H} = \{z\in\mathbb{C} \mid \Im(z)>0\}$ that is invariant with respect to the action $f(z) \mapsto (f|_k\gamma)(z) := (cz+d)^{-k} f\big(\frac{aX+b}{cX+d}\big)$ of $\gamma = \left(\begin{smallmatrix} a & b\\ c & d\end{smallmatrix}\right) \in \Gamma$ and satisfies a suitable growth condition. Since Γ_1 is generated by the two elements $T = \left(\begin{smallmatrix} 1 & 1\\ 0 & 1\end{smallmatrix}\right)$ and $S = \left(\begin{smallmatrix} 0 & -1\\ 1 & 0\end{smallmatrix}\right)$, this is equivalent to the two conditions $f(z+1)=f(z)$, meaning that $f(z)$ has a *Fourier expansion* $f(z) = \sum a_n q^n$ ($q = e^{2\pi i z}$), and $f(-1/z) = z^k f(z)$. The most famous example is Ramanujan's discriminant function

$$\Delta(z) \;=\; q\prod_{n=1}^\infty (1-q^n)^{24} \;=\; q - 24q^2 + 252q^3 - \cdots - 6048q^6 + \cdots \;=:\; \sum_{n=1}^\infty \tau(n)\,q^n$$

which, as Ramanujan discovered experimentally in 1916 and Mordell proved one year later, satisfies the multiplicativity property $\tau(mn) = \tau(m)\tau(n)$ for $(m,n)=1$, e.g.

$$\tau(6) \;=\; -6048 \;=\; -24\cdot 252 \;=\; \tau(2)\,\tau(3)\,.$$

Modular forms with this property are called *Hecke eigenforms* and play a central role in the whole theory and in Manin's work.

Manin's results (i) and (ii) for the special case $f = \Delta$, $k = 12$ are then as follows:

(i) Set $r_n(f) = \int_0^\infty f(z)\, z^n\, dz$ $(n = 0, \dots, k-2)$. Then for $f = \Delta$, $k = 12$ we have

$$\big(r_0(\Delta) : r_2(\Delta) : r_4(\Delta)\big) = \big(\tfrac{360}{691} : -\tfrac{4}{9} : \tfrac{1}{7}\big)\,, \quad \big(r_1(\Delta) : r_3(\Delta) : r_5(\Delta)\big) = \big(\tfrac{48}{5} : -5 : 4\big)\,.$$

In other words, the $r_n(f)$ for n even/odd are algebraic (here rational) multiples of two basic periods $\omega_\pm = \omega_\pm(f)$. This result has inspired countless others, notably Deligne's conjecture on special values of L-functions as multiples of periods of algebraic forms, with special cases for $\mathrm{Sym}^n(f)$ involving monomials in the same numbers ω_+ and ω_-.

(ii) Combining the modular symbols idea with the theory of Hecke operators, Manin proved the now famous formula (in which $\sigma_{11}(n) := \sum_{d|n} d^{11}$)

$$\sigma_{11}(n) - \tau(n) \;=\; \frac{691}{18} \sum_{\Delta\Delta' + \delta\delta' = n} \Delta^2 \delta^2 (\Delta^2 - \delta^2)^3\,, \tag{1}$$

where the sum is over all integral solutions of the equation with $0 < \delta < \Delta$ and *either* $0 < \delta' < \Delta'$ *or* $\delta' = 0$ and $0 < \delta \le^* \frac{1}{2}\Delta$ (* = count endpoint as 1/2). This expresses the nth Fourier coefficient of a modular form in terms of the representations of n by an indefinite quadratic form and can be seen as the start of a theory of holomorphic theta series attached to indefinite quadratic forms that has slowly been being developed during the last 20 years by Zwegers, myself, and others.

These results can be stated very nicely in terms of the language of *period polynomials.* For $f \in S_k(\Gamma_1)$ (cusp forms) we define the period polynomial $P_f(X) \in \mathbb{C}[X]$ of f by

$$P_f(X) \;=\; \int_0^\infty f(z)\,(z - X)^{k-2}\, dz \;=\; \sum_{n=0}^{k-2} (-1)^n \binom{k-2}{n} r_n(f)\, X^{k-n}\,.$$

Then Manin's result (i) says that the even and odd parts $P_f^\pm(X)$ of this polynomial, up to multiplicative constants $\omega_\pm(f)$, have algebraic (and even rational for $f = \Delta$) coefficients whenever f is a Hecke eigenform, e.g.,

$$P_\Delta(X) \;=\; \omega_+\big(\tfrac{36}{691} X^{10} - X^2(X^2-1)^3 - \tfrac{36}{691}\big) \;+\; \omega_-\, X(X^2-1)^2(X^2-4)(4X^2-1)\,.$$

The polynomial P_f is the special case $P_f = P_{f,S}$ of a map $\Gamma \ni \gamma \mapsto P_{f,\gamma} \in V_k$ given by

$$P_{f,\gamma}(X) \;=\; \int_{\gamma^{-1}(\infty)}^\infty f(z)\,(z - X)^{k-2}\, dz\,.$$

Here V_k is the space of polynomials in X of degree $\le k-2$, with the group action given by $P(X) \mapsto (P|_{2-k}g)(X) = (cX+d)^{k-2} P\big(\frac{aX+b}{cX+d}\big)$ for $g = \big(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\big) \in \mathrm{SL}(2,\mathbb{R})$, and the map $\gamma \mapsto P_{f,\gamma}$ is a cocycle for this action, meaning that $P_{f,\gamma\gamma'} = P_{f,\gamma}|_{2-k}\gamma' + P_{f,\gamma'}$ for all γ, $\gamma' \in \Gamma$. In particular, $P_{f,\gamma}$ for all γ are determined by $P_{f,S} = P_f$ and $P_{f,T} = 0$, since S and T generate Γ_1. The theory of period polynomials was developed by Eichler, Shimura, and Manin, and for $\Gamma = \Gamma_1$ we know that the maps $S_k(\Gamma) \to H^1(\Gamma, V_k)$ sending a modular form to the even or odd part of its period polynomial are both injective.

That $\gamma \mapsto P_{f,\gamma}$ is a cocycle can be checked by direct calculation or by writing it as a coboundary in a larger space of functions. If f is a modular form of weight k on Γ_1 (or any other lattice Γ), then its *Eichler integral* is by definition any holomorphic function F in $\mathfrak{H}$ with $d^{k-1}F/dz^{k-1} = f$. Explicit representations if $f(z) = \sum_{n=1}^{\infty} a_n e^{2\pi i n z}$ is a cusp form are given by

$$F(z) \;=\; \sum_{n=1}^{\infty} \frac{a_n}{(2\pi i n)^{k-1}}\, e^{2\pi i n z} \;=\; \frac{1}{(k-1)!} \int_z^{i\infty} f(\tau)\,(\tau - z)^{k-2}\, d\tau\,. \tag{2}$$

An easy calculation shows that $F|_{2-k}\gamma$ is given by the same integral as in (2) but with the upper limit of integration replaced by $\gamma^{-1}\infty$. This gives $\bigl(F(z) \,-\, F|_{2-k}\gamma\bigr)(z) = P_{f,\gamma}(z)$, from which the cocycle property follows immediately.

The result (ii) was also established using period polynomials. For $n \in \mathbb{N}$, the *nth Hecke operator* T_n acting on the space of modular forms sends $f \in M_k(\Gamma)$ to the sum of $f|_k M$ with M ranging over the Γ_1-orbits of 2×2 integral matrices of determinant n, and Manin showed that this action could be lifted to period polynomials by $P_{f|T_n} = P_f|\widetilde{T}_n$, where $P|\widetilde{T}_n$ is defined as the sum of the polynomials $P|_{2-k}\bigl(\begin{smallmatrix} \Delta & -\delta \\ \delta' & \Delta' \end{smallmatrix}\bigr)$ with $\Delta, \Delta', \delta, \delta'$ satisfying the conditions in (1). This "period version" of Hecke theory had many later variants (Mazur 1972, Merel 1991, Choie-Zagier 1993). One of them could be refined in a way that led to a new and elementary proof of the Eichler-Selberg trace formula for $\mathrm{Tr}(T_n, M_k(\Gamma_1))$, as I sketched in a lecture in Japan in 1993 and completed with Alexandru Popa in 2020.

I end this first part of the lecture with a rather surprising consequence of the theory of period polynomials that I found some 20 years ago: For every real number x, the sum

$$\sum_{\substack{a,b,c\in\mathbb{Z},\ a<0 \\ b^2-4ac=5}} \max(0, ax^2 + bx + c) \tag{3}$$

converges, and its value is the constant function 2 ! For example, if $x = 1/\pi$ then there are only triples (a, b, c) with $-100000 < a < 0$ and $ax^2 + bx + c > 0$, namely,

$$(-1, 1, 1),\ (1, -1, 1),\ (-5, 5, -1),\ (-11, 7, -1),\ (-409, 259, -4),\ (-541, 345, -55),$$

and the corresponding six values

$$1.21699,\ 0.58037,\ 0.08494,\ 0.11364,\ 0.00190,\ 0.00215$$

of $ax^2 + bx + c$ already add up to 1.99998. If we replace "5" in (3) by any positive integer D, then the same result is true with the constant "2" replaced by a constant $a_D \in \mathbb{Z}$ depending only on D and essentially equal to $\zeta_{\mathbb{Q}(\sqrt{D})}(2)$, and if we replace the quantity $\max(0, ax^2 + bx + c)$ by its cube, then the result still holds, with a_D replaced by a constant b_D essentially equal to $\zeta_{\mathbb{Q}(\sqrt{D})}(4)$, but if we put the 5th power instead in (2) then the right-hand side becomes $c_D + d_D\Phi(x)$ where $\Phi(x)$ is a function that is differentiable but not C^∞ and that essentially coincides with the extension to $\mathbb{R}$ of the Eichler integral of Ramanujan's discriminant function Δ !

PART 2: The Inversion Problem

Given the injectivity of $f \mapsto P_f$ (or $f \mapsto P_f^{\pm}$), it is natural to pose the question of reconstructing f explicitly from P_f (or P_f^+ or P_f^-). This question has two versions: giving a formula for $f(z)$ in terms of P_f for $z \in \mathfrak{H}$ and giving a formula for the nth Fourier coefficient a_n of f in terms of P_f for $n \in \mathbb{N}$. Here I'll describe only the former.

The solution, like both of Manin's papers, is based on the theory of continued fractions. Given a real number x, its "minus" continued fraction is given by $x = n_0 - 1/(n_1 - 1/\cdots)$ with the $n \in \mathbb{Z}$ defined inductively by $x_0 = x$, $x_i \le n_i < x_i + 1$, and $x_{i+1} = 1/(n_i - x_i)$. One then has $x_i = \gamma_i(x)$ with $\gamma_0 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ and $\gamma_{i+1}^{-1} = \gamma_i^{-1}\left(\begin{smallmatrix} n_i & -1 \\ 1 & 0 \end{smallmatrix}\right)$. On the other hand, the Fourier integral expression of the Eichler integral F of f in (2) converges also for z real, so one can talk about the values $F(x)$ for $x \in \mathbb{R}$, and of course the equalities $F|T = F$ and $F|S = F - P_f$ still hold there. These equalities imply that

$$P_f|\gamma_{i+1} \;=\; (F - F|T^{n_i}S)|\gamma_{i+1} \;=\; F|\gamma_{i+1} - F|\gamma_i$$

and hence (after verifying the convergence)

$$F(x) \;=\; -\sum_{i=1}^{\infty}(P_f|\gamma_i)(x) \qquad (x \in \mathbb{R})$$

One now uses the following surprising lemma, in which the key idea is to *forget* that the matrices γ_i's are naturally ordered!

Lemma. For every $x \in \mathbb{R} \smallsetminus \mathbb{Q}$, the set of matrices $\{\gamma_1, \gamma_2, \dots\}$ defined by the continued fraction of x coincides with the set of $\gamma \in \Gamma_1$ satisfying $0 \le \gamma(\infty) < 1 \le \gamma(x)$.

Combining this lemma with the easy formula $f(z) \doteq \int_{-\infty}^{\infty} F(t)\,(z-t)^{-k}\,dt$ expressing the cusp form f in terms of the extension to $\mathbb{R}$ of its Eichler integral, we get the formula

$$\boxed{f(z) \;\doteq\; \sum_{\substack{\gamma \in \Gamma_1 \\ 0 \le \gamma(\infty) < 1}} \bigl(Q_f\big|_k\gamma\bigr)(z) \quad \text{with} \quad Q_f(z) := \int_1^{\infty} \frac{P_f(t)}{(t-z)^k}\,dt \;\in\; \mathbb{C}\Bigl[\frac{1}{z-1}\Bigr]}$$

giving $f(z)$ as an infinite sum of rational functions determined completely by P_f.

One can obtain other representations of f in terms of P_f by changing the choice of continued fraction algorithm. For instance, with the "nearest-integer" algorithm (x_i and γ_i defined as above but now with $x_i - \frac{1}{2} \le n_i < x_i + \frac{1}{2}$), the analogue of the lemma above says that the set of γ_i ($i \ge 1$) for a given number $x \in \mathbb{R}$ consists of all $\gamma \in \Gamma_1$ with

$$\bigl(\gamma(x), \gamma(\infty)\bigr) \;\in\; \bigl[2, \infty\bigr) \times \bigl[\tfrac{1-\sqrt{5}}{2}, \tfrac{3-\sqrt{5}}{2}\bigr] \;\cup\; \bigl(-\infty, -2\bigr] \times \bigl[\tfrac{-3+\sqrt{5}}{2}, \tfrac{-1+\sqrt{5}}{2}\bigr]\,. \qquad (4)$$

To understand the situation better, I considered new continued fraction algorithms defined by

$$x \mapsto \begin{cases} x+1 \quad (= Tx) & \text{if } x < a, \\ -1/x \quad (= Sx) & \text{if } a < x < b, \\ x-1 \quad (= T^{-1}x) & \text{if } x > b \end{cases}$$

(and some choice at the end points), where a and b are real numbers satisfying $b - a > 1$ and $ab < -1$. Then, based on extensive numerical experiments, I conjectured that for these continued fraction algorithms there is always a similar statement to the one shown in (4), but with the right-hand side replaced by some finite union of rectangles with horizontal and vertical sides. This is illustrated for the example $(a,b) = \left(-\frac{2}{3}, \frac{7}{10}\right)$ by the following pictures, which were kindly made for me by Anke Pohl:

The first picture shows a domain ("attractor") in $\mathbb{R} \times \mathbb{R}$ bounded by finitely many horizontal and vertical lines:

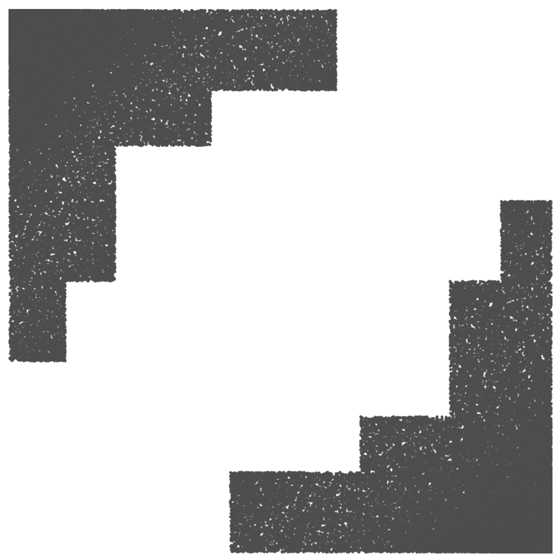

Figure 1: The attractor

The second picture shows this region divided into three subregions corresponding to $\gamma(x)$ belonging to $(-\infty, a)$, (a, b) or (b, ∞), colored red, blue and green respectively for convenience of visualization:

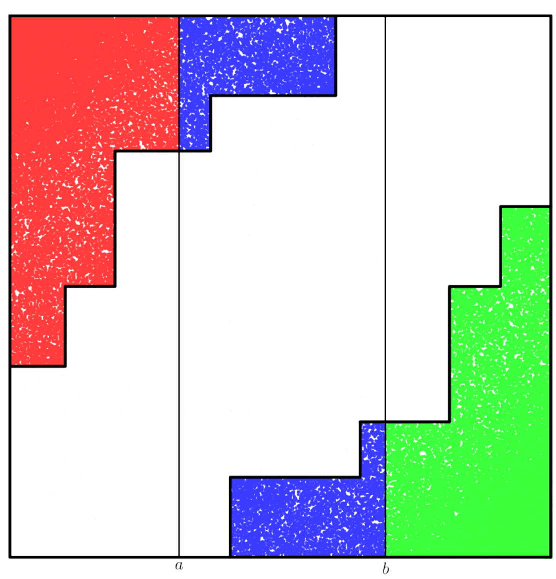

Figure 2: "Before"

The third picture shows the images of these three domains under the diagonal maps $(x,y) \to (Tx, Ty)$, $(x,y) \to (Sx, Sy)$, and $(x,y) \to (T^{-1}x, T^{-1}y)$, respectively:

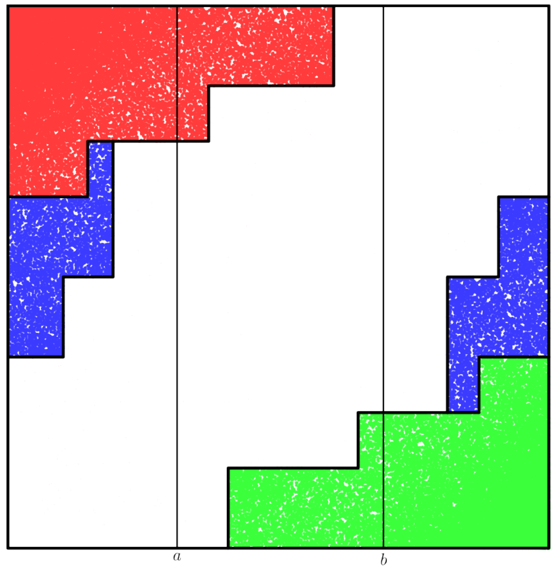

Figure 3: "After"

Comparing these last two pictures, we see that they give two different partitions of the original attractor into three sets, corresponding to the generators T, S and T^{-1}, and this is precisely the property that we wish to achieve. Finally, we remark that all of these pictures actually live on the 2-torus $\mathbb{P}^1(\mathbb{R}) \times \mathbb{P}^1(\mathbb{R})$, so that for instance the "After" picture is perhaps better imagined as repeated periodically on an unfolded torus:

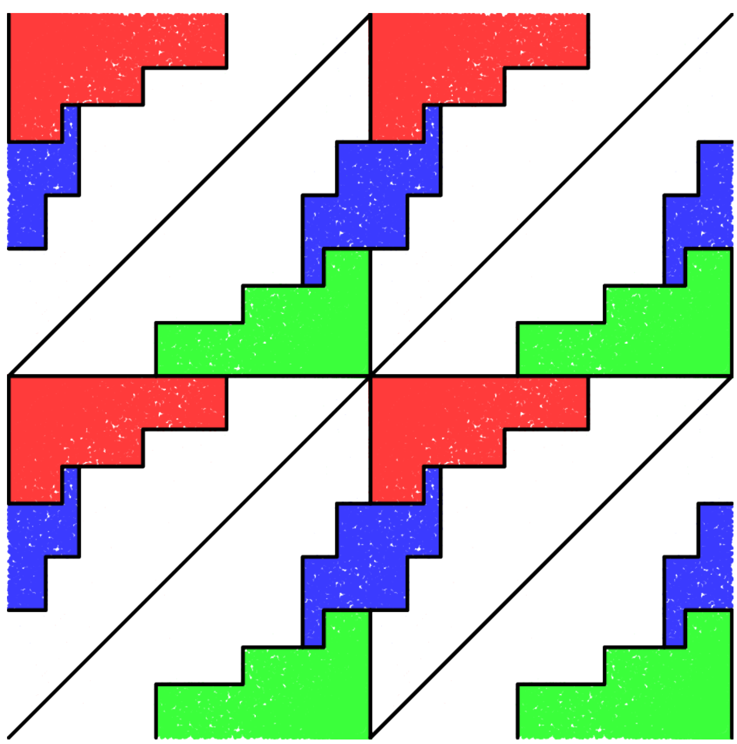

Figure 4: Repeated periodically

I conjectured that the same behavior holds for any $(a, b) \in \mathbb{R}$ satisfying $b - a > 1$ and $ab < -1$. This was proved later by Svetlana Katok and Ilie Ugarcovici at the MPIM.

However, as I mentioned at the beginning of this lecture, what I really wanted was not to have infinitely many different representations of a modular form on $\mathrm{SL}(2,\mathbb{Z})$ in terms of its associated period cocycle, but at least one such representation for modular forms on any nice Fuchsian group Γ, say the congruence subgroup

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_1 \;\middle|\; c \equiv 0 \pmod{N} \right\}$$

with $N \in \mathbb{N}$, since this property would imply the existence of an explicit formula for reconstructing a cusp form on Γ from its period cocycle. Specifically, the conjecture says that the analogue of the lemma above holds for some "continued fraction algorithm" defined by $x \mapsto \alpha_m(x)$ for all $x \in I_m$. Here $(I_1, \ldots, I_M)$ is a partition of $\mathbb{R}$ into finitely many disjoint intervals I_m labelled by suitable elements $\alpha_m \in \Gamma$, one of the key parts of the conjecture being to define what "suitable" means. (I have a number of different candidate definitions and believe that in a fair level of generality they are all equivalent to one another.) This statement is still unproven, as I said at the beginning, but results of Anke Pohl in her thesis combined with recent results of her and Paul Wabnitz (arXiv, Sept. 2022) imply that there is indeed at least one "suitable" algorithm for $\Gamma = \Gamma_0(N)$ for many values of N, including all primes. If a conjecture formulated by Anke's student Nicolas Herzog in his 2011 Bachelor's thesis and verified by him up to $N = 5000$ is true, then their result would in fact apply to all N which are either odd or have at most three prime factors.