Continued Fractions and Modular Forms

Ilan Vardi
IHES, Bures-sur-Yvette

April 3, 2000

[summary by Cyril Banderier]

Abstract

This incursion into the realm of elementary and probabilistic number theory of continued fractions, via modular forms, allows us to study the alternating sum of coefficients of a continued fraction, thus solving the longstanding open problem of their limit law.

1. Introduction

For the readers of these proceedings,[1] it is not a secret anymore that the continued fraction expansion of p/q, the Jacobi symbol $\left(\frac{p}{q}\right)$, or the gcd of two integers (p,q), or even Gauss' lattice reduction algorithm cover phenomena of similar computational complexity. However for continued fractions, two distinct cases have to be considered: the continuous and the discrete case. The discrete case deals with continued fraction expansions of rational numbers whereas the continuous case deals with continued fractions of real (irrational) numbers.

For the continuous model, given the apparatus of ergodic theory, many basic results on continued fractions fall as application of more general theorems (see Chapter 9 in Paul Lévy's book [9]). Ergodic theory, which was guessed by Maxwell and formulated by Boltzmann, concerns itself principally with quantifying how points in a continuous space evolve under iteration of a transformation. The ergodic theorem (due to Birkhoff in 1931) states that for almost all initial points x_0 of the continuous space E with measure λ,

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{j=1}^{n} f\left(T^j(x_0)\right) = \int_E f(y)\, d\lambda(y).$$

For the discrete model, there is some kind of effect that precludes the use of ergodic theory. At least, results from the continuous model may serve as a heuristic for guessing corresponding facts about the discrete world.

What one would need in order to make this heuristic rigorous is a kind of "ergodic theory with an error term." This is to some extent afforded by the introduction of Ruelle operators (see the works by Brigitte Vallée in 1995) and of modular forms (see the works by Ilan Vardi in 1987–1993) in the discrete model.

The main object of this lecture is the alternating sum of coefficients of a continued fraction. A motivation for studying this is the evaluation of Legendre symbol $\left(\frac{d}{c}\right)$, which essentially expresses whether d is a perfect square modulo c. This symbol can be evaluated using the Euclidean reduction

[1] For newcomers, I highly recommend the reading of summaries of previous talks by I. Vardi [15] and B. Vallée [12]!

$\left(\frac{d}{c}\right) = \left(\frac{d \bmod c}{c}\right)$ (where c, d are odd) and the quadratic reciprocity law

$$\left(\frac{c}{d}\right) = (-1)^{(c-1)(d-1)/4} \left(\frac{d}{c}\right).$$

In fact, it was shown by Rademacher [10] that

$$\left(\frac{d}{c}\right) = (-1)^{(3-a-d+c\sum(-1)^i a_i)/4},$$

where $\frac{d}{c} = [0, a_1, \ldots, a_r]$ (brackets stand for the expansion in continued fraction) and $0 < a, d < c$, $ad \equiv 1 \bmod c$ with c and r odd. Note that $d^{-1} \bmod c$ can itself be computed using the Euclidean algorithm.

There is also a geometrical motivation: the alternating sum expresses the number of times that a geodesic winds around the cusp of a modular surface.

In the continuous case, Guivarc'h and Le Jan [5] established that the average alternating sum converges to a Cauchy distribution with characteristic function $\exp\left(-\pi|t|/(2\ln 2)\right)$. For the discrete case, the stumbling block is that even the expected asymptotics estimated $\sigma(d,c) \sim \frac{12}{\pi^2} \ln c \ln\ln c$ is unproved ($\sigma(d,c)$ stands for the sum of the coefficient of the continued fraction of $\frac{d}{c}$). The problem remained open until Vardi found another approach (see [14]) via Dedekind sums.

2. **Dedekind Sums**

The Dedekind sum appears to be have been mistakenly defined and instead should have been defined as the alternating sum of continued fraction coefficients. Historically, the Dedekind sum is defined for relatively prime integers d and c as

$$s(d,c) = \sum_{h=1}^{c-1} ((hd/c)) \, ((h/c)) \,,$$

with the notation $((x)) = \begin{cases} 0, & \text{if } x \text{ is an integer,} \\ x - \lfloor x \rfloor - 1/2, & \text{otherwise.} \end{cases}$

This sum was introduced by Dedekind in 1876 while editing Riemann's collected works. He used this sum to express the functional equation of the Dedekind η function

$$\eta(z) = e^{\pi i z/12} \prod_{n=1}^{\infty} (1 - e^{2\pi i n z})$$

which, he proved, satisfies

$$\ln \eta\left(\frac{az+b}{cz+d}\right) = \begin{cases} \ln\eta(z) + \frac{1}{2}\ln(cz+d) + \frac{\pi i}{12}\left(-3 + \frac{a+d}{c} - 12 s(d,c)\right) & \text{for } c > 0, \\ \ln\eta(z) + \frac{\pi i b}{12}\,, & \text{c} = 0, \end{cases} \tag{1}$$

where $\Im(z) > 0$, $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, and a, b, c, d are integers satisfying $ad - bc = 1$. Note that

$$\ln \eta(z) = \frac{\pi i z}{12} - \sum_{m,n=1}^{\infty} \frac{e^{2\pi i m n z}}{m}\,,$$

so $\ln \eta$ is holomorphic for $\Im(z) > 0$. Using the functional equation (1) Dedekind proved a fundamental identity for Dedekind sums, namely the reciprocity law

$$s(c,d) = \frac{c}{d} + \frac{d}{c} + \frac{1}{cd} - s(d,c)\,.$$

Note that the definition of the Dedekind sum gives that $s(d,c) = s(d \bmod c, c)$ and the reciprocity law relates the value of $s(d,c)$ to $s(c,d)$. It follows that $s(d,c)$ can be computed by using the Euclidean algorithm, so it should be expressible in terms of the continued fraction expansion of d/c. In fact, this is the statement of a result found independently by three authors in 1977:

Theorem 1 (Barkan [1], Hickerson [7], Knuth [8]). *If $[0, a_1, a_2, \ldots, a_r]$ is the regular continued fraction expansion of d/c with r odd (with $d < c$ and $0 < a < c$ such that $ad \equiv 1 \bmod c$) then*

$$s(d,c) = \frac{1}{12}\left(-3 + \frac{a+d}{c} - \sum_{i=1}^{r}(-1)^i a_i\right).$$

It remains to find the distribution of the values of the Dedeking sums when c and d range over large intervals. Vardi did so by using Paul Lévy's theorem (details in Section 4) and for this, needed to justify several approximations. To this aim, let us recall a few facts about modular forms and Kloosterman sums, since these objects appeared to be the key to the asymptotic analysis.

3. Modular Forms

The group $SL(2,\mathbb{Z})$ acts on the upper half complex plane $\mathbf{H}$ by $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) z = \frac{az+b}{cz+d}$. One now considers subgroups G of $SL(2,\mathbb{Z})$ containing every matrices of $SL(2,\mathbb{Z})$ congruent to the identity matrix in $SL(2,\mathbb{Z}/N\mathbb{Z})$. Every such group G has a fundamental domain: an open set $D \subset \mathbf{H}$ such that for all $z \in \mathbf{H}$ there is at most one $g \in G$ with $g(z) \in D$ and at least one $g \in G$ with $g(z) \in \overline{D}$.

Definition. A *modular form* of weight k is a holomorphic function on $\mathbf{H}$ satisfying:

1) Modularity condition: $f(gz) = (cz+d)^k f(z)$ for $g \in G$,
2) Meromorphy condition: $f(z)$ is bounded in the cusps (i.e., parts of D going off to infinity).

Definition. A *non-holomorphic modular form* of *weight* r and *multiplier system* χ is a function $f(z)$ on $\mathbf{H}$ satisfying:

$$1')\ f(gz) = \chi(g)\left(\frac{cz+d}{|cz+d|}\right)^r f(z) \quad (\text{for } g \in G), \quad \text{and} \quad 2')\ \iint_D |f(x+iy)|^2 \frac{dxdy}{y^2} < \infty\,.$$

Condition 2') shows that the non-holomorphic modular forms form a Hilbert space $L^2(D,\chi,r)$ under the Petersson inner product

$$\langle f,g\rangle = \iint_D f(z)\overline{g(z)}y^r \frac{dxdy}{y^2}\,.$$

The Kloosterman sum (introduced by Kloosterman in 1927 in a refinement of the Hardy–Littlewood circle method) is defined by

$$S(m,n,c) = \sum e^{2\pi i(ma+nd)/c}\,,$$

where the sum ranges over $d < c$ for $ad \equiv 1 \bmod c$ and $\gcd(d,c) = 1$.

In 1948, André Weil proved the estimate $S(m,n,c) = O(c^{1/2+\epsilon})$. Asymptotics of sums of Kloosterman sums is a vivid subject, e.g., this kloostermania recently succeeded [3] to prove that there are infinitely many numbers of the form x^2+y^4. Sums of Kloosterman sums exhibit strong cancellations that can be estimated by making use of modular forms.

Generalized Kloosterman sums (for some subgroup G of $SL(2,\mathbb{Z})$) are defined by

$$S(m,n,c,\chi,G) = \sum \overline{\chi(g)}\, e^{2\pi i((m-\alpha)a+(n-\alpha)d)/(qc)}\,,$$

where the sums ranges over $g = \left(\begin{smallmatrix} a & c \\ c & d \end{smallmatrix}\right) \in G$ with $0 < a < qc$ and $0 < d < qc$. In the sum, q is the smallest integer such that $\left(\begin{smallmatrix} 1 & q \\ 0 & 1 \end{smallmatrix}\right) \in G$ and α is defined by $e^{-2\pi i\alpha} = \chi\left(\begin{smallmatrix} 1 & q \\ 0 & 1 \end{smallmatrix}\right)$ with $0 \leq \alpha < 1$.

The Goldfeld and Sarnak formulation (see [4]) of the Kuznetsov's trace formula gives

$$\sum_{c<N} S(m,n,c,\chi,G) = \sum_{1/2<s_j<1} \tau_j(m,n,\chi,G)\,\frac{N^{2s_j-1}}{2s_j} + O(N^{\beta/3+\varepsilon})\,, \tag{2}$$

where β is the best constant that can be put in the estimate $S(m,n,c,\chi,G) = O(c^{\beta+\varepsilon})$, and the sum is over exceptional eigenvalues s_j (defined hereafter) of the operator $\Delta_r = y^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) - iry\frac{\partial}{\partial x}$. Since Δ_r has a self adjoint extension to $L^2(D,\chi,r)$, its spectrum is discrete and real: There is a sequence of eigenvalues going to infinity, with only a finite set of negative eigenvalues which correspond to holomorphic modular forms if r is an even integer. The non-negative eigenvalues are simple except the case $\lambda = 1/4$, which could have multiplicity 2.

According to Selberg's notation, one writes an eigenvalue as $\lambda = s(1-s)$, with $\Re(s) \geq 1/2$. It follows that there is a finite number of *exceptional* eigenvalues for which $\lambda < 1/4$. An exceptional eigenvalue corresponds to $s > 1/2$, while the other eigenvalues have $\Re(s) = 1/2$ (note the analogy with the Riemann hypothesis).

For a given non-holomorphic Poincaré series P_m (see [2]), the Petersson product $\langle P_m, u_j\rangle$ gives the m-th Fourier coefficient $\rho_j(m)$ of the eigenfunction u_j (which is a modular form) associated to the eigenvalue $s_j(1-s_j)$. This allows to make explicit the τ_j's of the formula (2):

$$\tau_j(m,n,\chi,G) = \frac{q^2\,\overline{\rho_j(m)}\rho_j(n)\left(\pi^2(m-\alpha)(n-\alpha)/q^2\right)^{1-s_j}\Gamma(s_j+r/2)\Gamma(2s_j-1)}{(-i)^r\,\pi\Gamma(s_j-r/2)}$$

Finally, the following theorem provides the link between a sum of generalized Kloosterman sums and a sum whose asymptotics allows the application of Lévy's theorem:

Theorem 2 (Vardi [13])**.**

$$e^{\pi ir/2}\sum_{\substack{0<d<c\\ \gcd(d,c)=1}} e^{2\pi irs(d,c)} = S\left(1,1,c,\chi_r,SL(2,\mathbb{Z})\right)\,,$$

where $\chi_r = e^{2\pi ir\left(\frac{a+d}{c}-3-12s(d,c)\right)}$.

4. **Limiting Distribution**

One says that an arithmetic function $f(n)$ has a limiting distribution $F(x)$ if

$$\lim_{N\to\infty}\frac{1}{N}\,|\{n<N:\quad f(n)<x\}\,| = F(x)\,.$$

In other words, one takes a histogram of values of the function $f(n)$ and looks at its shape. One method of showing that an arithmetic function has a limiting distribution is due to Paul Lévy (see examples in [11]).

Theorem 3 (Paul Lévy)**.** *If there exists a function $g(t)$ continuous in 0 such that*

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n<N} e^{itf(n)} = g(t)\,,$$

then $f(n)$ has a limiting distribution $F(x)$ satisfying $g(t) = \int_{-\infty}^{\infty} e^{itx}\,dF(x)$.

This is simply the probabilist's terminology for the Fourier transform; g is the *characteristic function* of the distribution.

In order to prove the limiting distribution result for Dedekind sums (and thus for alternating sums of continued fraction coefficients) one applies Lévy's theorem to $s(d,c)/\ln c$. What we want to prove is the estimate

$$\lim_{N\to\infty} \frac{\sum\limits_{\substack{0<d<c<N\\ \gcd(d,c)=1}} e^{its(d,c)/\ln c}}{|\{0<d<c<N:\ \gcd(d,c)=1\}|} = e^{-|t|/(2\pi)}, \tag{3}$$

where the right-hand side corresponds to the characteristic function of the Cauchy distribution

$$\int_{-\infty}^{\infty} \frac{\alpha\, e^{ity}}{\alpha^2+y^2}\, dy = e^{-\alpha|t|},$$

with $\alpha = 1/(2\pi)$. The well-known estimate [6]

$$|\{0<d<c<N:\ \gcd(d,c)=1\}| = \frac{3N^2}{\pi^2} + O(N\ln N)$$

shows that the sought estimate (3) can be rewritten as

$$\sum_{\substack{0<d<c<N\\ \gcd(d,c)=1}} e^{its(d,c)/\ln c} \sim e^{-|t|/(2\pi)}\, \frac{3N^2}{\pi^2}.$$

Proving such a formula presents a number of technical difficulties. For example, one would like to remove the absolute values on the right hand side, and the bothersome $1/\ln c$ term in the exponential. The first point is solved since $s(c-d,c) = -s(d,c)$ so that the left-hand side is independent of the sign of t. Consequently, one may restrict attention to $t>0$. The second point is solved noting that the ln function does not vary very much, and that for most values of $c<N$, $\ln c$ is almost equal to $\ln N$. An estimate obtained by the continued fraction formula for Dedekind sums and the subsequent upper bound of $S(d/c) \leq (\ln N)^{3/2+\varepsilon}$ for almost all $d<c<N$, shows that

$$\sum_{\substack{0<d<c<N\\ \gcd(d,c)=1}} e^{its(d,c)/\ln c} = \sum_{\substack{0<d<c<N\\ \gcd(d,c)=1}} e^{its(d,c)/\ln N} + O\left(N^2(\ln N)^{-1/5+\varepsilon}\right).$$

See [14] for details. The problem is therefore reduced to showing that

$$\sum_{\substack{0<d<c<N\\ \gcd(d,c)=1}} e^{its(d,c)/\ln N} \sim e^{-t/(2\pi)}\, \frac{3N^2}{\pi^2}, \quad t>0. \tag{4}$$

Summing the relation of Theorem 2 leads to

$$\sum_{\substack{0<d<c<N\\ \gcd(d,c)=1}} e^{2\pi i r s(d,c)} = e^{-i\pi r/2} \sum_{0<c<N} S\left(1,1,c,\chi_r, SL(2,\mathbb{Z})\right),$$

so it suffices to obtain asymptotics of the last right-hand side when $r = t/(2\pi\ln N)$. This is given by the specialization of Kuznetsov's trace formula (2) with $\beta = 1$ (the trivial bound) which yields

$$\frac{(1/4)^r}{\pi A_r(1-r/2)} N^{2-r} + O(N^{4/3+\varepsilon}) = \frac{1}{\pi A_0}\, e^{-t/(2\pi)} N^2 + O(N^2/\ln N)$$

where $A_r = \iint_D y^r |\eta(x+iy)|^{4r} \frac{dxdy}{y^2}$ is the Petersson norm of $y^{r/2}\eta^{2r}(x+iy)$ (the eigenfunction of the only exceptional eigenvalue $\frac{r}{2}(1-\frac{r}{2})$).

As one easily computes $A_0 = \pi/3$, Equation (4) is satisfied.

5. Conclusion

This talk is based on [2], for more details, refer to Ilan Vardi's articles, available from

`http://www.ihes.fr/~ilan/publications.html`.

The alternating sum of coefficients of a continued fraction seems to be the first example where one needs not only upper bounds for sums of Kloosterman sums, but also their precise asymptotics.

The following fact is noteworthy. Euclidean algorithms are fundamental in several branches of science while counting amongst the oldest known algorithms. It is another testimony of the "unreasonable effectiveness of mathematics" (a phrase due to Eugene Wigner [16]) that they reveal their finest secrets only with our recent knowledges of dynamical systems and of analytical number theory. Long live applied mathematics!

References

[1] Barkan (Philippe). – Sur les sommes de Dedekind et les fractions continues finies. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences. Séries A et B, Paris*, vol. 284, n° 16, 1977, pp. A923–A926.

[2] Flajolet (Philippe), Vallée (Brigitte), and Vardi (Ilan). – Continued fractions from Euclid to the present day. *In preparation.*

[3] Friedlander (John) and Iwaniec (Henryk). – The polynomial X^2+Y^4 captures its primes. *Annals of Mathematics. Second Series*, vol. 148, n° 3, 1998, pp. 945–1040.

[4] Goldfeld (D.) and Sarnak (P.). – Sums of Kloosterman sums. *Inventiones Mathematicae*, vol. 71, n° 2, 1983, pp. 243–250.

[5] Guivarc'h (Y.) and Le Jan (Y.). – Asymptotic winding of the geodesic flow on modular surfaces and continued fractions. *Annales Scientifiques de l'École Normale Supérieure. Quatrième Série*, vol. 26, n° 1, 1993, pp. 23–50.

[6] Hardy (G. H.) and Wright (E. M.). – *An introduction to the theory of numbers.* – The Clarendon Press Oxford University Press, New York, 1979, fifth edition, xvi+426p.

[7] Hickerson (Dean). – Continued fractions and density results for Dedekind sums. *Journal für die reine und angewandte Mathematik*, vol. 290, 1977, pp. 113–116.

[8] Knuth (Donald E.). – Notes on generalized Dedekind sums. *Acta Arithmetica*, vol. 33, n° 4, 1977, pp. 297–325.

[9] Lévy (P.). – *Théorie de l'addition des variables aléatoires.* – Gauthier–Villars, 1937.

[10] Rademacher (Hans). – Generalization of the reciprocity formula for Dedekind sums. *Duke Mathematical Journal*, vol. 21, 1954, pp. 391–397.

[11] Tenenbaum (Gérald). – *Introduction to analytic and probabilistic number theory.* – Cambridge University Press, Cambridge, 1995, xvi+448p. Translated from the second French edition (1995) by C. B. Thomas.

[12] Vallée (Brigitte). – Analyse en moyenne d'une classe d'algorithmes d'Euclide. In *Algorithms seminar 1998–1999, INRIA Research Report 3830.* – 1999.

[13] Vardi (Ilan). – A relation between Dedekind sums and Kloosterman sums. *Duke Mathematical Journal*, vol. 55, n° 1, 1987, pp. 189–197.

[14] Vardi (Ilan). – Dedekind sums have a limiting distribution. *International Mathematics Research Notices*, n° 1, 1993, pp. 1–12.

[15] Vardi (Ilan). – Les fractions continues d'Euclide jusqu'au présent. In *Algorithms seminar 1998–1999, INRIA Research Report 3830.* – 1999.

[16] Wigner (Eugene P.). – The unreasonable effectiveness of mathematics in the natural sciences. *Communications in Pure and Applied Mathematics*, vol. 13, n° I, 1960. – Wigner's article can also be found at the address http://www.txwesleyan.edu/aegis/aegistwo/Unreasonable.html.