An Analysis of the gaussian algorithm for lattice reduction

Hervé Daude, Philippe Flajolet, Brigitte Vallée

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

An Analysis of the Gaussian Algorithm for Lattice Reduction

Hervé DAUDÉ
Philippe FLAJOLET - Brigitte VALLÉE

N° 2243
Avril 1994

PROGRAMME 2

Calcul symbolique,
programmation
et génie logiciel

An Analysis of the Gaussian Algorithm for Lattice Reduction

Hervé DAUDÉ, Philippe FLAJOLET, Brigitte VALLÉE

Abstract. *The Gaussian algorithm for lattice reduction in dimension 2 (under both the standard version and the centered version) is analysed. It is found that, when applied to random inputs, the complexity is asymptotically constant, the probability distribution decays geometrically, and the dynamics is characterized by a conditional invariant measure. The proofs make use of connections between lattice reduction, continued fractions, continuants, and functional operators. Detailed numerical data are also presented.*

Une analyse de l'algorithme de Gauss pour la réduction des réseaux

Résumé. L'algorithme de Gauss pour la réduction des réseaux en dimension 2 est analysé sous ses deux formes, standard et centrée. Il s'avère que sur des données aléatoires, la complexité de l'algorithme est asymptotiquement constante, la distribution de probabilité des coûts décroit géométriquement, et la dynamique de l'algorithme est caractérisée par une distribution invariante conditionnelle. Les preuves font appel aux rapports existant entre fractions continues, continuants, et certains opérateurs fonctionnels associés. Des résultats numériques détaillés sont également présentés.

Paper presented at the *Algorithmic Number Theory Symposium (ANTS'94)*, Cornell University, May 1994. Proceedings are to be published in the *Lecture Notes in Computer Science* series, Springer-Verlag, L. Adleman editor.

AN ANALYSIS OF THE GAUSSIAN ALGORITHM FOR LATTICE REDUCTION

Hervé Daudé[1], Philippe Flajolet[2], and Brigitte Vallée[3]

[1] Département de Mathématiques, Université de Provence,
Case 96, 3 Place Victor Hugo F-13331 Marseille Cedex 3 (France)
[daude@gyptis.univ-mrs.fr].
[2] INRIA-Rocquencourt, F-78153 Le Chesnay (France),
[Philippe.Flajolet@inria.fr].
[3] Département d'Informatique,Université de Caen, F-14032 Caen (France),
[Brigitte.Vallee@univ-caen.fr].

Abstract. The Gaussian algorithm for lattice reduction in dimension 2 (under both the standard version and the centered version) is analysed. It is found that, when applied to random inputs, the complexity is asymptotically constant, the probability distribution decays geometrically, and the dynamics is characterized by a conditional invariant measure. The proofs make use of connections between lattice reduction, continued fractions, continuants, and functional operators. Detailed numerical data are also presented.

1 Introduction

The lattice reduction problem consists in finding a short basis of a lattice of Euclidean space given a (usually skew) basis. This reduction problem is well-known to be central to many areas of approximation and optimization with deep consequences in computational number theory, cryptography, and symbolic computation.

In dimension $d = 1$, lattice reduction may be viewed as a mere avatar of the Euclidean GCD algorithm and of continued fraction expansions. Lattice reduction *per se* really started with Gauss who gave an algorithm that solves the problem exactly using what resembles a lifting of the Euclidean algorithm to 2-dimensional lattices. In recent times, an important discovery was made by Lenstra, Lenstra and Lovász in 1982 [4, 12, 25]; their algorithm, called the LLL algorithm, is able to find reduced bases in all dimensions $d \geq 3$. The LLL algorithm itself proceeds by stages based on the Gaussian algorithm as the main reduction step.

The Euclidean algorithm and the continued fraction algorithm are by now reasonably well understood as regards complexity questions. Knuth's book [10] provides a detailed account till 1981. From results of Lamé, Dupré, Heilbronn, Dixon, Wirsing, Babenko, and Hensley, the following facts are known. The worst case complexity of the Euclidean algorithm is $\mathcal{O}(\log N)$ when applied to integers at most N (Lamé and Dupré); the average case on random inputs is also logarithmic (Dixon); the distribution of the number of iterations obeys in the asymptotic limit a normal law with a variance that is logarithmic (a recent result of Hensley).

There are some deep connections between these properties and an invariant measure for the continued fraction transformation whose existence was first conjectured by Gauss and proved in this century by Lévy and Kuzmin. Most of these results are obtained by means of functional operators related to continued fractions and continuants of which extensive use will be made here. We refer in particular to the works of Wirsing [24], Babenko [1], Mayer [16, 17, 14, 15], and Hensley [8].

This paper provides a detailed analysis of the Gaussian algorithm, both in the average case and in probability. Like its one–dimensional counterpart, the algorithm is known to be of worst–case logarithmic complexity, a result due to Lagarias [11], with best possible bounds being provided by Vallée [21] and Kaib-Schnorr [9]. The probabilistic behaviour of the Gaussian algorithm turns out to be appreciably different however. The main results of the paper are as follows.

— The average–case complexity of the Gaussian algorithm (measured in the number of iterations performed) is asymptotically constant, and thus essentially independent on the size of the input vectors.
— The distribution of the number of iterations is closely approximated by a geometric law.
— The dynamics of the algorithm is governed by a conditional invariant measure that constitutes the analogue of the invariant measure first observed by Gauss for continued fractions.

Precise characterizations of the behaviour of the algorithm are given here. In particular the geometric rate of decrease of the distribution of costs and the conditional invariant measure are expressed simply in terms of spectral properties of an operator that generalizes the operator associated with Euclid's algorithm.

In this paper, we mostly focus on the analysis of what we call the "standard" version of the Gaussian reduction algorithm, which generalizes the standard Euclidean algorithm. Another often encountered version, called here the "centered" version, is analogous to the centered Euclidean algorithm and is amenable to a similar treatment as we briefly explain at the end of the paper.

Our analytic results are naturally expressed as multiple infinite sums involving the continuants of continued fraction theory. As such sums tend to be rather slowly convergent, some attention is also paid to obtaining precise estimates by means of simple convergence acceleration techniques. For instance, we establish that the average case complexity of the "inner part" of the algorithm is asymptotic to the constant $\mu = 1.35113\,15744\ldots$.

On average, the Gaussian algorithm is thus of complexity $\mathcal{O}(1)$, which is of an order different from the worst-case. The case of dimension $d = 2$ therefore departs significantly from its 1–dimensional analogue, and it would be of interest to determine to which extent such a phenomenon propagates to higher dimensions. Our analytic knowledge of the LLL algorithm in higher dimensions is of course less advanced, but Daudé and Vallée [6] already succeeded in proving that the LLL algorithm, when applied to d–dimensional lattices, has an average–case complexity that is bounded from above by a constant K_d, where $K_d = \mathcal{O}(d^2 \log d)$. The present work thus fits as a component of a more global

Fig. 1. A lattice and two of its bases represented by the parallelogram they span. The first basis is skew, the second one is reduced.

enterprise whose aim is to understand theoretically why the LLL algorithm performs in practice much better than worst-case bounds predict, and to quantify precisely the probabilistic behaviour of lattice reduction in higher dimensions.

2 Lattice reduction in dimension 2

*Lattices and bases.*This paper addresses specifically the reduction of 2-dimensional lattices. A *lattice* of rank 2 in the complex plane $\mathbb{C}$ is the set $\mathcal{L}$ of elements of $\mathbb{C}$ ("vectors") defined by

$$\mathcal{L} = \mathbb{Z}u \oplus \mathbb{Z}v = \{\lambda u + \mu v \mid \lambda, \mu \in \mathbb{Z}\},$$

where (u, v), called a *basis*, is a pair of $\mathbb{R}$-linearly independent elements of $\mathbb{C}$.

A lattice is generated by infinitely many bases that are related to each other by integer matrices of determinant ± 1. Amongst all the bases of $\mathcal{L}$, some, called *minimal*, enjoy the property of being formed with a shortest vector u of the lattice and another vector v which is shortest in the set of all vectors independent of u. Minimality is the specialization to dimension 2 of the general notion of reduced basis in arbitrary dimensions. A minimal basis (u, v), when it is in addition *acute*, is characterized by the two simultaneous conditions:

$$(I_1): \ |\frac{v}{u}| \geq 1 \qquad \text{and} \qquad (I_2): \ 0 \leq \Re(\frac{v}{u}) \leq \frac{1}{2}. \tag{1}$$

The angle between the two vectors of a minimal acute basis thus lies between $\frac{\pi}{3}$ and $\frac{\pi}{2}$.

The Gaussian reduction schema. A reduction algorithm takes an arbitrary basis of a lattice and determines another basis that is minimal. The Gaussian algorithm is a reduction algorithm whose principle consists in satisfying the two simultaneous conditions of (1). Condition I_1 is satisfied by exchanges between vectors, then condition I_2 is satisfied by an integral translation of the longer vector v parallel to the shorter vector u. The schema underlying this reduction process is then the following.

Input: an acute basis (u, v) of $\mathcal{L}$.
Output: a minimal acute basis (u, v) of $\mathcal{L}$.

repeat

(i). If $|u| > |v|$, then exchange u and v so as to satisfy condition I_1;
(ii). Translate v parallel to u: $v := v - m\,u$ for some $m \in \mathbb{N}$, so as to satisfy condition I_2; if (u, v) is not acute then change v to $-v$;

until $|v| \geq |u|$.

The complex framework. Many structural characteristics of lattices and bases are invariant under linear transformations —similarity transformations in geometric terms— of the form $S_\lambda : z \mapsto \lambda z$ with $\lambda \in \mathbb{C}$. An instance is the characterization of minimal acute bases that only depends on the ratio v/u. It is thus natural to consider lattices and bases taken up to equivalence under linear transformation (similarity). For such similarity invariant properties, it is sufficient to restrict attention to lattices generated by a basis of the form $(1, z)$. In that case, the property for a basis to be minimal and acute corresponds to the fact that z belongs to the so-called *fundamental domain* $\mathcal{F} = \left\{ z \;\middle|\; |z| \geq 1 \text{ and } 0 \leq \Re(z) \leq \frac{1}{2} \right\}$. Such a domain is familiar from the theory of modular forms [20] or the reduction theory of quadratic forms [19].

The Gaussian algorithm precisely has the property that its execution trace is invariant under lattice similarity. Let $(u_0, v_0), \ldots, (u_k, v_k)$ be the sequence of bases constructed by the Gaussian algorithm. We associate to it the sequence $(1, z_0), \ldots, (1, z_k)$ where $z_j = v_j/u_j$. The geometric transformation effected by the each step of the algorithm consists of an exchange $(u, v) \mapsto (v, u)$, a translation $v \mapsto v - m\,u$, and a possible sign change $v \mapsto \varepsilon v$ with $\varepsilon = \pm 1$. In the complex framework, this corresponds to an inversion $S : z \mapsto 1/z$, followed by a translation $z \mapsto T^{-m} z$ with $T(z) = z + 1$, and by a possible sign change $z \mapsto J_\varepsilon z$ where $J_\varepsilon(z) = \varepsilon z$. In this context, the Gaussian algorithm aims at realizing directly the conditions by bringing $z = v/u$ in the strip $\widetilde{\mathcal{B}} = \{0 \leq \Re(z) \leq \frac{1}{2}\}$.

The Gaussian reduction schema that we have just described involves a sign-changing operation $(v \mapsto -v)$. We introduce below a variant of the algorithm —the "standard" algorithm— that has the advantage of avoiding this operation whose presence complicates the analysis. We propose to return to the original Gaussian algorithm in Section 7.

The standard algorithm. The next sections are devoted to the analysis of a variant of the Gaussian algorithm that is directed towards bringing z inside the strip $\mathcal{B} = \{0 \leq \Re(z) \leq 1\}$. In order to do so, it suffices to consider a transformation U formed with an inversion S and a translation T^{-m} aimed at bringing z into $\mathcal{B}$. It is readily realized that this is achieved by the transformation

$$U(z) = \frac{1}{z} - \left\lfloor \Re\left(\frac{1}{z}\right) \right\rfloor,$$

with $\lfloor u \rfloor$ the integer part of u. This transformation U is an extension to the complex domain of the operation defining standard continued fraction expansions.

In the rest of the paper, we assume that the Gaussian algorithm is applied to complex numbers z such that $\Im(z) \neq 0$, which corresponds to nondegenerate

lattices. One also operates with bases that are acute, so that z belongs to the half-plane $\Re(z) \geq 0$. For reasons explained below, see Eq. (2), it suffices to consider the situation where the reduction algorithm takes as input complex numbers from the disk $\mathcal{D}$ of diameter $[0, 1]$. The transformation U is then iterated till exit from that disk. This defines an algorithm called the *standard Gaussian algorithm* (SGA) because of its close connection with standard continued fractions and the standard Euclidean algorithm.

Algorithm $SGA(z$: complex) [Standard Gaussian Algorithm]
Input: $z \in \mathcal{D}$ (the disk of diameter $[0, 1]$)
while $(z \in \mathcal{D})$ **do** $z := U(z)$;
Output: $z \in \mathcal{B} \setminus \mathcal{D}$. (the strip $\mathcal{B}$ is defined by $0 \leq \Re(z) \leq 1$)

For this algorithm, upon exit from the main iteration loop, it is no longer true that z belongs to the fundamental domain $\mathcal{F}$. However, z then lies in the union of six simple transforms of $\mathcal{F}$, namely

$$\mathcal{B} \setminus \mathcal{D} = \mathcal{F} \cup S\mathcal{F} \cup SJ\mathcal{F} \cup ST\mathcal{F} \cup TJ\mathcal{F} \cup STJ\mathcal{F}. \tag{2}$$

Thus simply adding a 6–way test produces an algorithm whose output is an element of $\mathcal{F}$. In addition, the analysis of the full reduction algorithm obtained in this way is then only a trivial variant of the analysis of the core algorithm SGA.

Probabilistic models. The question addressed here is the estimation of the number L of iterations performed by the standard algorithm. The model considered is in essence equivalent to applying the reduction algorithm to random bases, where similar bases are identified.

The *continuous model* is defined by the fact that the inputs are taken uniformly over the definition domain $\mathcal{D}$. The eventual goal is to analyse the behaviour of the algorithm under a *discrete model* where inputs are members of $\mathbb{Q}(i)$ of the form $\mathbb{Q}^{(N)}(i) = \{\frac{a}{N} + i\frac{b}{N} \mid b \neq 0\}$, suitably restricted to $\mathcal{D}$. The random variable $L^{(N)}$ then depends on N. However, as N gets large, it converges, both in moments and distribution, to its continuous counterparts, a fact to be proved in Section 5.

Thus, the results to be enounced later for the continuous model —that the average number of iterations is constant and that the probability distribution admits exponential tails — carries over to the more accurate discrete model. In other words, the behaviour of lattice reduction in dimension 2 is essentially insensitive to the size of the input vectors. This is a notable difference with the one–dimensional case of Euclid's algorithm.

3 Continued fractions and lattice reduction

The Gaussian algorithm is closely related to the linear fractional transformations (also called homographies) that are associated to continued fractions, and thus also to the classical continuant polynomials. In this way, a first analysis of the probability distribution and of the average cost of the algorithm can be given.

The fundamental disks. In its complex formulation, the algorithm *SGA* produces a sequence $z_0, z_1, \ldots, z_k$ of transforms of $z_0 \in \mathcal{D}$ obtained by iterating the transformation U. As we saw, each step corresponds to a particular transformation

$$z_{j+1} = -m_j + \frac{1}{z_j} \qquad \text{or} \qquad z_j = \frac{1}{m_j + z_{j+1}}. \tag{3}$$

While z_j is in $\mathcal{D}$, $1/z_j$ lies in the exterior of $\mathcal{D}$ and it satisfies $\Re(1/z_j) > 1$, so that we have the condition $m_j \geq 1$. Thus, from (3), there results that an execution of the Gaussian algorithm on input z_0 translates into a terminating "continued fraction" expansion

$$z_0 = \cfrac{1}{m_1 + \cfrac{1}{m_2 + \cfrac{1}{\ddots \cfrac{}{m_k + z_k}}}}, \tag{4}$$

where the expansion is stopped as soon as z_k lies in $\mathcal{B} \setminus \mathcal{D}$. The number of iterations, L, then assumes the value k. All the m_j are at least 1.

This leads to introducing the set $\mathcal{H}_k$ of linear fractional transformations of *depth* k (for $k \geq 1$) defined as the collection of all $h(z)$ of the form

$$h_{\mathbf{n}}(z) = h_{n_1, n_2, \ldots, n_k}(z) = \cfrac{1}{n_1 + \cfrac{1}{n_2 + \cfrac{1}{\ddots \cfrac{}{n_k + z}}}}, \tag{5}$$

where the $n_j \in \mathbb{N} = \{1, 2, \ldots\}$.

From the preceding discussion, we thus have the equivalence

$$z_k = U^k(z_0) \qquad \Longleftrightarrow \qquad \exists \mathbf{n} \in \mathbb{N}^k \ \ (z_0 = h_{\mathbf{n}}(z_k)).$$

The event $\{L \geq k+1\}$ coincides with the set of complex z such that all the $U^j(z)$, for $j = 0, \ldots, k$, lie in $\mathcal{D}$. Thus, defining $\mathcal{D}_k = U^{(-k)}(\mathcal{D})$ with $\mathcal{D}_0 = \mathcal{D}$, we have $\{L \geq k+1\} \equiv \mathcal{D}_k$. By definition, these domains form an infinite descending chain, $\mathcal{D}_0 \supset \mathcal{D}_1 \supset \mathcal{D}_2 \supset \cdots$. We also have that each $\mathcal{D}_k$ is the disjoint union of transforms of $\mathcal{D}$ by the transformations of $\mathcal{H}_k$ of (5),

$$\mathcal{D}_k = U^{(-k)}(\mathcal{D}) = \bigcup_{\mathbf{n} \in \mathbb{N}^k} h_{\mathbf{n}}(\mathcal{D}).$$

From elementary properties of geometrical inversion, $h_{\mathbf{n}}(\mathcal{D})$ is the disk of diameter $[h_{\mathbf{n}}(0), h_{\mathbf{n}}(1)]$. Within the theory of continued fractions, the interval $[h_{\mathbf{n}}(0), h_{\mathbf{n}}(1)]$ is known as a fundamental interval. A rendering of the domains, also called the fundamental disks, is given in Figure 2.

These considerations imply that, under the uniform probabilistic model of use, the probability ϖ_k that the algorithm performs at least $k+1$ iterations is

$$\varpi_k = \frac{\|\mathcal{D}_k\|}{\|\mathcal{D}\|} = \frac{4}{\pi} \sum_{\mathbf{n} \in \mathbb{N}^k} \|h_{\mathbf{n}}(\mathcal{D})\|, \tag{6}$$

where $\|A\|$ denotes the area of a domain $\mathcal{A}$ of the plane.

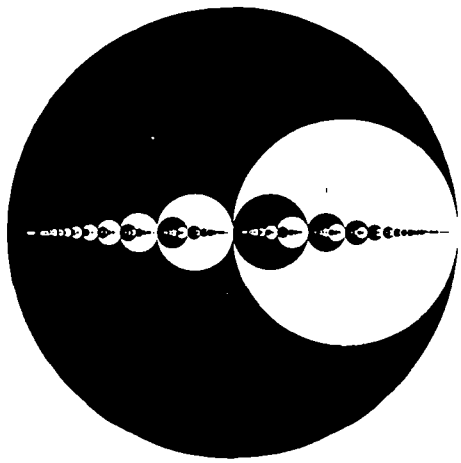

Fig. 2. The domains $\mathcal{D}_0 \setminus \mathcal{D}_1$, $\mathcal{D}_1 \setminus \mathcal{D}_2$, $\mathcal{D}_2 \setminus \mathcal{D}_3$, $\mathcal{D}_3 \setminus \mathcal{D}_4$, $\mathcal{D}_4 \setminus \mathcal{D}_5$ represented alternatively in black and white. (The largest disk is $\mathcal{D}_0 \equiv \mathcal{D}$ which is the disk of diameter $[0, 1]$.)

Continuants. Homographies of $\mathcal{H}_k$ are naturally associated with continued fractions of depth k themselves expressible in terms of *continuants*, see for instance the books by Knuth [10, p. 340] or by Rockett and Szüsz [18]. The continuant polynomials are defined by

$$Q_n(x_1, x_2, \ldots, x_n) = x_n Q_{n-1}(x_1, \ldots, x_{n-1}) + Q_{n-2}(x_1, \ldots, x_n),$$

with $Q_0 = 1, Q_1(x_1) = x_1$. Classically, a function $h_{\mathbf{n}} \in \mathcal{H}_k$ with $\mathbf{n} = (n_1, \ldots, n_k)$ admits the expression

$$h(z) = \frac{P_k + zP_{k-1}}{Q_k + zQ_{k-1}}, \tag{7}$$

where

$$\begin{array}{ll} Q_k = Q_k(n_1, \ldots, n_k), & Q_{k-1} = Q_{k-1}(n_1, \ldots, n_{k-1}), \\ P_k = Q_{k-1}(n_2, \cdots, n_k), & P_{k-1} = Q_{k-2}(n_2, \ldots, n_{k-1}). \end{array} \tag{8}$$

As is well–known the continuant polynomial $Q_n(x_1, \ldots, x_n)$ is also the sum of all monomials that obtain by crossing out pairs $x_i x_{i+1}$ of consecutive variables in the product $x_1 x_2 \cdots x_n$. Continuants thus satisfy the symmetry property $Q_k(x_1, \ldots, x_k) = Q_k(x_k, \ldots, x_1)$ and the determinant identity $Q_k P_{k-1} - Q_{k-1} P_k = (-1)^{k-1}$.

Probabilistic analysis. The previous considerations permit to express the probability distribution of the Gaussian algorithm in terms of continuants.

Theorem 1. *The probability ϖ_k that algorithm SGA performs more than k iterations on a random input $z \in \mathcal{D}$ is expressible as*

$$\varpi_k \equiv \Pr\{L \geq k + 1\} = \sum_{n_1, \ldots, n_k} \frac{1}{Q_k^2 (Q_k + Q_{k-1})^2},$$

where $Q_k = Q_k(n_1, \ldots, n_k)$, $Q_{k-1} = Q_{k-1}(n_1, \ldots, n_{k-1})$, *and the sum is over all integers* $n_j \geq 1$.

The following table displays the probability distribution of *SGA* computed by Theorem 1 and the numerical methods of Section 5 against the result of 10^8 simulations of the algorithm.

k	$\Pr\{L \geq k\}$	Simulations
1	0.28986	0.28984361
2	0.04848	0.04847104
3	0.01027	0.01027170
4	0.00200	0.00200478
5	0.00040	0.00040299
6	0.00008	0.00008031
7	0.00002	0.00001569
Expectation:	1.35113	1.351094

Average-case analysis. Elementary number–theoretic considerations permit to express the expected cost of the Gaussian algorithm under a form no longer involving continuants. We have:

Theorem 2. *The mean number of iterations of algorithm SGA applied to a random* $z \in \mathcal{D}$ *is*

$$\mathrm{E}\{L\} = \frac{5}{4} + \frac{180}{\pi^4} \sum_{d \geq 1} \frac{1}{d^2} \sum_{d<c<2d} \frac{1}{c^2}.$$

Dynamic analysis. We have already mentioned the importance of the invariant measure of Gauss that has density $\frac{1}{\log 2} \frac{1}{1+x}$, in the 1–dimensional case. No such invariant measure can exist here as the reduction algorithm terminates. However, a rôle quite similar to the invariant measure of Gauss is played by a function that describes the distribution of successive transforms of the input as the reduction algorithms proceeds.

Initially, the input distribution is uniform in the disk $\mathcal{D}$, so that to z_0 is associated the constant density function over $\mathcal{D}$. Assume now that the algorithm performs at least $k+1$ iterations. Then the kth iterate is well defined and is an element of $\mathcal{D}$. A natural question is to determine its distribution inside $\mathcal{D}$. The corresponding *conditional* density function $F_k(z)$ must be proportional to $\lim_{\rho \to 0} \frac{1}{\pi\rho^2} \Pr\{z_k \in D(z,\rho)\}$, where $D(z,\rho)$ is the disk of center z and radius ρ. The proportionality factor must be taken so as to ensure that the integral of the density over $\mathcal{D}$ equals 1, so that the legitimate definition of the density function is

$$F_k(z) = \lim_{\rho \to 0} \frac{1}{\pi\rho^2} \frac{\Pr\{z_k \in D(z,\rho)\}}{\Pr\{z_k \in \mathcal{D}\}}.$$

We shall call F_k the *dynamic density* (of order k) of the algorithm.

Theorem 3. *The dynamic density* F_k *is given by*

$$F_k(z) = \frac{1}{\varpi_k} \sum_{\mathbf{n} \in \mathbb{N}^k} \frac{1}{|Q_{k-1}z + Q_k|^4}, \qquad \textit{where} \qquad \varpi_k = \Pr\{L \geq k+1\}.$$

Furthermore, a functional relation holds for *real* x,

$$\frac{\varpi_{k+1}}{\varpi_k} F_{k+1}(x) = \sum_{m\geq 1} \frac{1}{(m+x)^4} F_k(\frac{1}{m+x}).$$

Thus, assuming that F_k admits a limit F_∞ and ϖ_{k+1}/ϖ_k converges to some constant λ, the quantities λ and F_∞ must be an eigenvalue and a corresponding eigenvector of the operator defined by the right hand side. This sharply motivates the introduction of the operator $\mathcal{G}$ in the next section, where we shall also establish the assumptions regarding F_k and ϖ_k.

4 The $\mathcal{G}$ operator

The complete analysis of the probability distribution and of the dynamics of the Gaussian algorithm depends on the introduction of an operator $\mathcal{G}_s$, formally defined by

$$\mathcal{G}_s[f](t) = \sum_{m\geq 1} \frac{1}{(m+t)^s} f(\frac{1}{m+t}), \tag{9}$$

and more specifically on the instance $s = 4$ that we simply denote by $\mathcal{G} \equiv \mathcal{G}_4$. (Continued fractions and the Euclidean algorithm correspond to the case $s = 2$.) Let V denote the open disk of center 1 and radius $\frac{3}{2}$. For all s with $\Re(s) > 1$, the operator $\mathcal{G}_s$ acts on the space $A_\infty(V)$ of functions f that are holomorphic in V and continuous on the closure $\overline{V}$ of V. The set $A_\infty(V)$ endowed with the sup-norm $\|f\| = \sup_{t\in\overline{V}} |f(t)|$ is a Banach space.

Such operators permit to "invert" the continued fraction operator U, and at the same time their functional analysis properties (related to the Perron–Frobenius theory) have useful consequences for the Gaussian algorithm. There is a close relationship between the iterates of $\mathcal{G}_s$ and continuants.

Lemma 4. *The iterates of $\mathcal{G}_s$ generate the continuants of depth k in the following sense:*

$$\mathcal{G}_s^k[f](t) = \sum_{n_1\dots n_k} \frac{1}{(Q_{k-1}t+Q_k)^s} f(\frac{P_{k-1}t+P_k}{Q_{k-1}t+Q_k}); \quad \mathcal{G}_s^k[f](0) = \sum_{n_1\dots n_k} \frac{1}{Q_k^s} f(\frac{Q_{k-1}}{Q_k}). \tag{10}$$

The quantities involved in Theorems 1, 3 precisely admit such expressions:

$$\varpi_k = \Pr\{L \geq k+1\} = \mathcal{G}^k[u](0) \qquad \text{where} \qquad u(t) = \frac{1}{(1+t)^2}$$
$$F_k(z) = \frac{1}{\varpi_k}\mathcal{G}^k[v_z(t)](0) \qquad \text{where} \qquad v_z(t) = \frac{1}{(1+tz)^2(1+t\bar{z})^2}.$$

Spectral properties of the $\mathcal{G}_s$ operators have been investigated in detail by Mayer and we globally refer to [17] and references therein. For s such that

$\Re(s) > 1$, the operators $\mathcal{G}_s$ are nuclear of order 0. In other words, they have a discrete spectrum and admit a spectral decomposition:

$$\mathcal{G}_s[f](t) = \sum_{i=1}^{\infty} \lambda_i e_i^*[f] e_i(t), \tag{11}$$

where the λ_i are the eigenvalues, $\{e_i\}$ is a basis of eigenfunctions, and the coefficients $\{e_i^*\}$ are the dual basis of $\{e_i\}$; in addition the λ_i are ρ-summable for all real $\rho > 0$: $\sum_i |\lambda_i|^\rho < +\infty$. (These quantities implicitly depend on s.)

For real $s > 1$ (we need the case $s = 4$), the operator $\mathcal{G}_s$ is in addition a Perron–Frobenius operator [17]: it has a unique positive dominant eigenvalue λ_1, the corresponding eigenfunction $e_1(t)$ is strictly positive on $\overline{V} \cap \mathbb{R}$, and $e_1^*[f]$ is strictly positive if f is itself positive on $\overline{V} \cap \mathbb{R}$. In particular, if λ_2 denotes the second eigenvalue (in order of absolute values), and if f is positive on $\overline{V} \cup \mathbb{R}$, one has

$$\| \frac{1}{\lambda_1^k} \mathcal{G}_s^k[f] - P[f] \| \leq \|f\| \cdot \left| \frac{\lambda_2}{\lambda_1} \right|^k, \tag{12}$$

where $P[f]$ denotes projection on the dominant eigensubspace: $P[f](t) = \lambda_1 e_1^*[f]\, e_1(t)$.

These considerations (with $s = 4$) apply to the continuant form of the probability distribution and to the (conditional) invariant measure of the Gaussian algorithm.

Theorem 5. *There exist real numbers c_j and λ_j with $\lambda_1 > |\lambda_2| > |\lambda_3| > \cdots$, such that*

$$\Pr\{L \geq k+1\} = \sum_{j=1}^{\infty} c_j \lambda_j^k.$$

In particular, with $c_1 \approx 1.3$, $\lambda_1 \approx 0.1993$, and $\lambda_2 \in [-\frac{1}{20}, -\frac{1}{10}]$, one has asymptotically:

$$\Pr\{L \geq k+1\} = c_1 \lambda_1^k \left[1 + \mathcal{O}((\frac{\lambda_2}{\lambda_1})^k) \right].$$

This theorem is in accordance with observation of the numerical data following Theorem 1, as the probabilities decay roughly like $(\frac{1}{5})^n$.

Theorem 6. *The dynamic density $F_k(z)$ converges geometrically to a (conditional) invariant density F_∞:*

$$F_\infty(x+iy) = \alpha \int_0^1 (1-w^2) \left[e_1(x+iyw) + e_1(x-iyw) \right] dw.$$

There, e_1 is the eigenfunction of $\mathcal{G}$ corresponding to the dominant eigenvalue λ_1, and α the normalization constant determined by $\iint F_\infty \, dx\, dy = 1$. In particular, on the real axis, the invariant density F_∞ is proportional to the eigenfunction e_1.

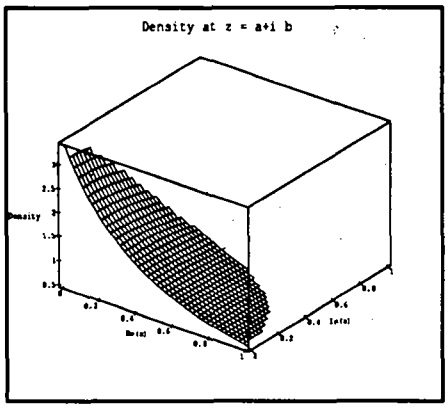

Fig. 3. The conditional invariant density F_∞.

5 The discrete model

The analysis of the standard algorithm under the discrete model where inputs are taken from the discrete set

$$\mathbb{Q}^{\langle N\rangle}(i) = \{\frac{a}{N} + i\frac{b}{N} \mid b \neq 0\}, \tag{13}$$

is solved by a combination of two arguments: (*i*) by "Gauss's principle" a circle of radius x contains $\pi N^2 x^2 + \mathcal{O}(xN + 1)$ lattice points of $\mathbb{Q}^{\langle N\rangle}(i)$; (*ii*) from worst-case bounds, the reduction algorithm performs number of iterations at most $\mathcal{O}(\log N)$.

Theorem 7. *Let $L^{\langle N\rangle}$ be the number of iterations of the standard Gaussian algorithm applied to random inputs from $\mathbb{Q}^{\langle N\rangle} \cap \mathcal{D}$. The random variable $L^{\langle N\rangle}$ converges in moments and in distribution to the random variable L associated with the continuous model. In particular, the mean value satisfies*

$$\mu^{\langle N\rangle} \equiv \mathrm{E}\{L^{\langle N\rangle}\} = \mu + \mathcal{O}(\frac{\log N}{N}).$$

6 Numerical estimates

We have already cited some numerical estimates for the mean and the probability distribution of the Gaussian algorithm, as well as the approximate value $\lambda_1 \approx 0.1993$. Most of the expressions involve slowly converging sums. The purpose of this section is to give indications on series transformations that permit to evaluate some of these quantities to great accuracy, as well as on ways in which precise bounds can be proved on λ_1 using trace formulae.

A real number α is said to be *polynomial time computable* if there exists an integer r such that an approximation of α to accuracy 10^{-d} can be computed

in time $\mathcal{O}(d^r)$. We let $\mathbf{P}$ denote the class of such numbers. A major problem is to find which of the constants of this paper are polynomial time computable. Effective numerical procedures usually result from proofs of membership in $\mathbf{P}$.

The expected cost. The expected number of iterations of the Gaussian algorithm admits an expression as a sum which, once truncated till terms of order m, results in an error of $\mathcal{O}(\frac{1}{m^2})$. This sum can be expressed instead as a definite integral involving the dilogarithm function, $\mathrm{Li}_2(z) = \sum_{n=1}^{\infty} \frac{z^n}{n^2}$, itself amenable to series representations (involving Bernoulli numbers) that exhibit geometric convergence.

Theorem 8. (*i*). *The mean number of iterations of the Gaussian algorithm SGA admits the integral representation*

$$\mu = -\frac{3}{4} + \frac{180}{\pi^4} \int_0^{\infty} \frac{\mathrm{Li}_2(e^{-t}) - \mathrm{Li}_2(e^{-2t})}{1 - e^{-t}}\, t\, dt. \tag{14}$$

(*ii*). *The number μ lies in the class $\mathbf{P}$ of polynomial time computable numbers:* $\mu = 1.35113\,15744\,91659\,00179\,38680\,05256\,46466\,84404\,78970\,85087 \pm 10^{-50}$.

Probability distribution. The probability distribution of the Gaussian algorithm can be expressed in terms of complicated series involving the zeta function, the resulting expressions being useful for small values of k.

Theorem 9. *The probability distribution of the number of iterations of the Gaussian algorithm has initial values:* $\Pr\{L \geq 1\} = 1$, $\Pr\{L \geq 2\} = \frac{\pi^2}{3} - 3$,

$$\Pr\{L \geq 3\} = -5 + \frac{2\pi^2}{3} - 2\zeta(3) + 2\sum_{n=0}^{\infty} (-1)^n (n+1)\zeta(n+4)(\zeta(n+2) - 1).$$

In general, each $\varpi_k = \Pr\{L \geq k+1\}$ is in the class $\mathbf{P}$.

The proof is based on the fact that summations of analytic functions of several complex variables at integer points,

$$\varphi = \sum_{n_1,\ldots,n_k=1}^{\infty} F(\frac{1}{n_1}, \frac{1}{n_2}, \ldots \frac{1}{n_k}),$$

can be represented as multiple sums of zeta functions at the integers. (See for instance [23] for the univariate case.) The following values have been determined in this way to great accuracy:

$$\begin{aligned} \varpi_1 &= 0.28986\,81336\,96452\,87294 \\ \varpi_2 &= 0.04848\,08014\,49463\,63270 \\ \varpi_3 &= 0.01027\,81647\,79066\,59643. \end{aligned}$$

Eigenvalues. The last numerical task is to estimate the dominant and subdominant eigenvalues, λ_1 and λ_2 that determine the rate of geometric decay of the probabilities ϖ_k. A first class of bounds is obtained by a nonlinear optimization problem based on Wirsing's approach [24] and the specific "test functions" pairs:

$$\begin{cases} \psi_a(t) = \dfrac{1}{(1+at)(1+(a+1)t)(1+(a+2)t)(1+(a+3)t)} \\ \phi_a(t) = \mathcal{G}[\psi_a](t) = \dfrac{1}{3}\dfrac{1}{(z+a+1)(z+a+2)(z+a+3)}. \end{cases} \tag{15}$$

With $a = 0.487$, the ration $\phi_a(t)/\psi_a(t)$ lies in $[0.170, 0.205]$ for $t \in [0,1]$. By iterating $\mathcal{G}$, a first bound

$$0.170 < \lambda_1 < 0.205$$

is then obtained. The more refined estimates mentioned in Theorem 5 are derived from adapting trace formulae originally due to Babenko and Mayer.

Theorem 10. (*i*). *The trace of the operator* $\mathcal{G}^k$ *satisfies*

$$\mathrm{Tr}\,\mathcal{G}^k \equiv \sum_i \lambda_i^k = \sum_{\mathbf{n}} \frac{\tau(\mathbf{n})^4}{1-(-1)^k\tau(\mathbf{n})^2}, \quad \tau(\mathbf{n}) = \frac{(Q_k + P_{k-1}) - \sqrt{(Q_k+P_{k-1})^2 - 4}}{2}. \tag{16}$$

(*ii*). *Each* $\mathrm{Tr}\,\mathcal{G}^k$ *is computable in polynomial time. In particular* $\mathrm{Tr}\,\mathcal{G}$ *admits the explicit form*

$$\mathrm{Tr}\,\mathcal{G} = \frac{7}{2} - \frac{2}{\sqrt{5}} - \frac{7}{\sqrt{2}} + \frac{1}{2}\sum_{n=2}^{\infty}(-1)^n \frac{n-1}{n+1}\binom{2n}{n}[\zeta(2n) - 1 - \frac{1}{2^{2n}}].$$

7 The centered Gaussian algorithm

The centered Gaussian algorithm constitutes the classical implementation of the general reduction schema described in Section 2. The aim of the algorithm is to bring z in the strip $\widetilde{\mathcal{B}} = \{z \mid 0 \le \Re(z) \le \frac{1}{2}\}$ by means of the transformation $z := \widetilde{U}(z)$ where

$$\widetilde{U}(z) = \varepsilon(\frac{1}{z})(\frac{1}{z} - \lfloor \Re(\frac{1}{z}) \rceil),$$

with $\lfloor u \rceil$ the integer nearest to u, and $\varepsilon(u)$ the sign of $\Re(u) - \lfloor \Re(u) \rceil$.

As was done with the standard algorithm, it is sufficient to restrict consideration to the *core algorithm*

Algorithm $CGA(z : \text{complex})$ [Centered Gaussian Algorithm]
Input: $z \in \widetilde{\mathcal{D}}$ (the disk of diameter $[0, \frac{1}{2}]$)
  while $(z \in \widetilde{\mathcal{D}})$ **do** $z := \widetilde{U}(z)$;
Output: $z \in \widetilde{\mathcal{B}} \setminus \widetilde{\mathcal{D}}$ ($\widetilde{\mathcal{B}}$ is the strip $0 \le \Re(z) \le \frac{1}{2}$)

An analysis of the average case has been given in [22]. Methods of this paper lead to a complete analysis. The theory develops with linear fractional transformations related to centered continued fraction expansions,

$$h_{\mathbf{m},\boldsymbol{\varepsilon}}(z) = \cfrac{1}{m_1 + \cfrac{\varepsilon_1}{m_2 + \cfrac{\varepsilon_2}{\ddots \cfrac{}{m_k + \varepsilon_k z}}}}, \tag{17}$$

and the corresponding continuants $\widetilde{Q}_k$. In (17), the pairs (m_j, ε_j) satisfy the basic condition:

$$\text{if } \varepsilon > 0 \text{ then } m \geq 2; \quad \text{if } \varepsilon < 0 \text{ then } m \geq 3. \tag{18}$$

The corresponding functional operator is then

$$\widetilde{\mathcal{G}}_s[f](t) = \sum_{m,\varepsilon} \left(\frac{\varepsilon}{m+t}\right)^s f\left(\frac{\varepsilon}{m+t}\right),$$

where the pair (m, ε) satisfies the conditions (18). The operator can be proved to enjoy properties similar to those of $\mathcal{G}_s$ (discrete spectrum, nuclearity; Perron-Frobenius properties for even integral s). This is summarized by the following statement:

Theorem 11. (*i*). *The mean number of iterations of the centered Gaussian algorithm CGA satisfies*

$$\widetilde{\mu} \equiv \mathrm{E}\{\widetilde{L}\} = \frac{360}{\pi^4} \sum_{d=1}^{\infty} \frac{1}{d^2} \sum_{c=\lceil d\phi^{-2} \rceil}^{\lfloor d\phi^{-1} \rfloor} \frac{1}{c^2}, \qquad \textit{with } \phi = (1+\sqrt{5})/2.$$

(*ii*). *The probability distribution decays exponentially:*

$$\widetilde{\varpi}_k = \mathrm{E}\{\widetilde{L} \geq k+1\} \sim \widetilde{c} \cdot \widetilde{\lambda}^k \qquad \textit{with} \qquad \widetilde{\lambda} \approx 0.077\,.$$

(*iii*). *The algorithm admits a limit conditional invariant density.*

It is a perhaps surprising fact that, despite its "nonanalytical" character, the average-case constant $\widetilde{\mu}$ can be computed in polynomial time. For $\theta \in (0,1)$ with convergent sequence $\{\frac{p_n}{q_n}\}$, one has:

$$\sum_{d=1}^{\infty} \sum_{c=1}^{\lfloor d\theta \rfloor} \frac{1}{c^2 d^2} = \sum_{n=0}^{\infty} \frac{(-1)^n}{(p_n + p_{n+1})^2 (q_n + q_{n+1})^2} \int_0^1 \int_0^1 \frac{\log x \log y}{\left(1 - x^{\frac{p_n}{p_n+p_{n+1}}} y^{\frac{q_n}{q_n+q_{n+1}}}\right)\left(1 - x^{\frac{p_{n+1}}{p_n+p_{n+1}}} y^{\frac{q_{n+1}}{q_n+q_{n+1}}}\right)} dx\, dy,$$

a formula that results from identities of Mahler and Borwein–Borwein [13, 3]. From this, we find

$$\widetilde{\mu} = 1.08922\,14740\,95380\ldots$$

Acknowledgements. The work of Philippe Flajolet was supported by the ESPRIT Basic Research Action No. 7141 (ALCOM II).

References

1. BABENKO, K. I. On a problem of Gauss. *Soviet Mathematical Doklady 19*, 1 (1978), 136–140.
2. BERNDT, B. C. *Ramanujan's Notebooks, Part I.* Springer Verlag, 1985.
3. BORWEIN, J. M., AND BORWEIN, P. B. Strange series and high precision fraud. *American Mathematical Monthly 99*, 7 (Aug. 1992), 622–640.
4. COHEN, H. *A Course in Computational Algebraic Number Theory.* No. 138 in Graduate Texts in Mathematics. Springer–Verlag, 1993.
5. DAUDÉ, H. *Des fractions continues à la réduction des réseaux: analyse en moyenne.* PhD thesis, Université de Caen, 1993.
6. DAUDÉ, H., AND VALLÉE, B. An upper bound on the average number of iterations of the LLL algorithm. *Theoretical Computer Science 123*, 1 (1994), 95–115.
7. EDWARDS, H. M. *Riemann's Zeta Function.* Academic Press, 1974.
8. HENSLEY, D. The number of steps in the Euclidean algorithm. Preprint, 1993.
9. KAIB, M., AND SCHNORR, C. P. A sharp worst–case analysis of the Gaussian lattice basis reduction algorithm for any norm. Preprint, 1992.
10. KNUTH, D. E. *The Art of Computer Programming*, 2nd ed., vol. 2: Seminumerical Algorithms. Addison-Wesley, 1981.
11. LAGARIAS, J. C. Worst–case complexity bounds for algorithms in the theory of integral quadratic forms. *Journal of Algorithms 1*, 2 (1980), 142–186.
12. LENSTRA, A. K., LENSTRA, H. W., AND LOVÁSZ, L. Factoring polynomials with rational coefficients. *Mathematische Annalen 261* (1982), 513–534.
13. MAHLER, K. Arithmetische Eigenschaften der Lösungen einer Klasse von Funktionalgleichungen. *Mathematische Annalen 101* (1929), 342–366.
14. MAYER, D., AND ROEPSTORFF, G. On the relaxation time of Gauss's continued fraction map. I. The Hilbert space approach. *Journal of Statistical Physics 47*, 1/2 (Apr. 1987), 149–171.
15. MAYER, D., AND ROEPSTORFF, G. On the relaxation time of Gauss's continued fraction map. II. The Banach space approach (transfer operator approach). *Journal of Statistical Physics 50*, 1/2 (Jan. 1988), 331–344.
16. MAYER, D. H. On a ζ function related to the continued fraction transformation. *Bulletin de la Société Mathématique de France 104* (1976), 195–203.
17. MAYER, D. H. Continued fractions and related transformations. In *Ergodic Theory, Symbolic Dynamics and Hyperbolic Spaces*, M. K. Tim Bedford and C. Series, Eds. Oxford University Press, 1991, pp. 175–222.
18. ROCKETT, A., AND SZÜSZ, P. *Continued Fractions.* World Scientific, Singapore, 1992.
19. SCHARLAU, W., AND OPOLKA, H. *From Fermat to Minkowski, Lectures on the Theory of Numbers and its Historical Developments.* Undergraduate Texts in Mathematics. Springer-Verlag, 1984.
20. SERRE, J.-P. *A Course in Arithmetic.* Graduate Texts in Mathematics. Springer Verlag, 1973.
21. VALLÉE, B. Gauss' algorithm revisited. *Journal of Algorithms 12* (1991), 556–572.
22. VALLÉE, B., AND FLAJOLET, P. Gauss' reduction algorithm: An average case analysis. In *Proceedings of the 31st Symposium on Foundations of Computer Science* (Oct. 1990), IEEE Computer Society Press, pp. 830–839.
23. VARDI, I. *Computational Recreations in Mathematica.* Addison Wesley, 1991.
24. WIRSING, E. On the theorem of Gauss–Kusmin–Lévy and a Frobenius–type theorem for function spaces. *Acta Arithmetica 24* (1974), 507–528.
25. ZIPPEL, R. *Effective Polynomial Computations.* Kluwer Academic Publishers, Boston, 1993.

Unité de Recherche INRIA Rocquencourt
Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 LE CHESNAY Cedex (France)

Unité de Recherche INRIA Lorraine Technopôle de Nancy-Brabois - Campus Scientifique
615, rue du Jardin Botanique - B.P. 101 - 54602 VILLERS LES NANCY Cedex (France)
Unité de Recherche INRIA Rennes IRISA, Campus Universitaire de Beaulieu 35042 RENNES Cedex (France)
Unité de Recherche INRIA Rhône-Alpes 46, avenue Félix Viallet - 38031 GRENOBLE Cedex (France)
Unité de Recherche INRIA Sophia Antipolis 2004, route des Lucioles - B.P. 93 - 06902 SOPHIA ANTIPOLIS Cedex (France)

EDITEUR
INRIA - Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 LE CHESNAY Cedex (France)

ISSN 0249 - 6399