The lattice reduction algorithm of Gauss : an average case analysis

Brigitte Vallée, Philippe Flajolet

UNITÉ DE RECHERCHE
INRIA-ROCQUENCOURT

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P. 105
78153 Le Chesnay Cedex
France
Tél.:(1) 39 63 55 11

Rapports de Recherche

N° 1277

Programme 1
Programmation, Calcul Symbolique
et Intelligence Artificielle

THE LATTICE REDUCTION ALGORITHM OF GAUSS : AN AVERAGE CASE ANALYSIS

Brigitte VALLEE
Philippe FLAJOLET

Août 1990

RR-1277

The Lattice Reduction Algorithm of Gauss: An Average Case Analysis*

Brigitte Vallée
Département de Mathématiques
Université de Caen
F-14032 Caen (France)

Philippe Flajolet
INRIA Rocquencourt
F-78150 Le Chesnay (France)

August 16, 1990

Abstract. *The lattice reduction algorithm of Gauss is shown to have an average case complexity which is asymptotic to a constant.*

L'algorithme de réduction des réseaux de Gauss: une analyse en moyenne

Résumé: On établit que l'algorithme de réduction de Gauss présente une complexité moyenne qui est asymptotiquement constante.

*Extended abstract presented at the 31st IEEE Symposium on Foundations of Computer Science, St. Louis, Missouri, October 22–24, 1990.

The Lattice Reduction Algorithm of Gauss: An Average Case Analysis*

Brigitte Vallée
Département de Mathématiques
Université de Caen
F-14032 Caen (France)

Philippe Flajolet
INRIA Rocquencourt
F-78150 Le Chesnay (France)

Abstract. *The lattice reduction algorithm of Gauss is shown to have an average case complexity which is asymptotic to a constant.*

Introduction. The "reduction" algorithm of Gauss plays an important rôle in several areas of computational number theory, principally in matters related to the reduction of integer lattice bases. It is also intimately connected with extensions to complex numbers of the Euclidean gcd algorithms and continued fraction expansions.

Continued Fractions. Every rational or real number has a continued fraction expansion. For instance, the number $193/71 \approx 2.71830$ leads to

$$\frac{193}{71} = 2 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{4 + \frac{1}{2}}}}}}, \tag{1}$$

which is obtained by the rule

$$\mathrm{CF}(x) = \lfloor x \rfloor + \cfrac{1}{\mathrm{CF}(\cfrac{1}{x - \lfloor x \rfloor})} \tag{2}$$

with $\mathrm{CF}(x) = x$ if x is an integer. For algorithms' analysts, such expansions are of interest as they relate to the standard Euclidean GCD algorithm: The number of stages in the computation of $\gcd(p, q)$ is precisely equal to the number of stages in the continued fraction expansion of the rational p/q.

Knuth [7, Sec. 4.5.3] has a nice exposition of this theory. The complexity of the standard GCD algorithm applied to numbers p, q at most N is known in the worst case (WC) as well as in the average case (AC):

$$\text{(WC):}\quad \log_\phi N + O(1), \tag{3}$$

$$\text{(AC):}\quad \frac{12 \log 2}{\pi^2} \log N + O(1), \tag{4}$$

with $\phi = (1 + \sqrt{5})/2$ being the golden ratio. These results are due respectively to Lamé [8] and Heilbronn–Dixon [5, 1].

There are other continued fractions, called "centered continued fractions", that are related to the centered Euclidean algorithm. For instance

$$\frac{193}{71} = 3 - \cfrac{1}{4 - \cfrac{1}{2 + \cfrac{1}{4 + \frac{1}{2}}}}, \tag{5}$$

presents better convergence properties than the standard fraction (1) since it uses only 4 stages instead of 6. That "centered" fraction is obtained by the modified process (where one uses the nearest-integer function $\lceil x \rfloor$ to represent $\lceil x \rfloor = \lceil x - \frac{1}{2} \rceil$):

$$\mathrm{CCF}(x) = \lceil x \rfloor + \cfrac{\epsilon(x)}{\mathrm{CCF}(\cfrac{\epsilon(x)}{x - \lceil x \rfloor})}$$

$$\text{with}\quad \epsilon(x) = \mathrm{sign}(x - \lceil x \rfloor), \tag{6}$$

and again $\mathrm{CCF}(x) = x$ if x is an integer. There exist corresponding results for the complexity of this process, due to Dupré [2] for the worst case and Knuth [7, Ex. 4.5.3.30-31] for the average case. One obtains:

$$\text{(WC):}\quad \log_{1+\sqrt{2}} N + O(1), \tag{7}$$

$$\text{(AC):}\quad \frac{12 \log \phi}{\pi^2} \log N + O(1). \tag{8}$$

For complex rationals or general complex numbers, i.e. numbers in $\mathbf{Q}(i)$ or $\mathbf{C}$, continued fraction expansions that, in a way, generalize (1) or (5) have been extensively studied by Gauss. For instance, we have

$$\frac{35470}{99661} + \frac{315}{99661} i = \cfrac{1}{3 - \cfrac{1}{5 + \cfrac{1}{\frac{1}{3} + \frac{7}{5} i}}}. \tag{9}$$

*Extended abstract presented at the 31st IEEE Symposium on Foundations of Computer Science, St. Louis, Missouri, October 22–24, 1990.

Every (nonreal) complex number admits such an expansion that *terminates*. It is called a *Gaussian fraction* and it has the property that all "quotients" but the last one are integers while the last quotient is a complex number which is furthermore constrained to be close to the imaginary axis, i.e., to belong to a suitably defined "fundamental domain" $\mathcal{F}$. More precisely, the expansion is obtained by a rule similar to rule (6),

$$\mathrm{GCCF}(z) = \lceil \mathrm{Re}(z) \rfloor + \frac{\epsilon(z)}{\mathrm{GCCF}(\frac{\epsilon(z)}{z - \lceil \mathrm{Re}(z) \rfloor})}$$

$$\text{with } \epsilon(z) = \mathrm{sign}(\mathrm{Re}(z) - \lceil \mathrm{Re}(z) \rfloor), \qquad (10)$$

together with the termination condition:

$$\mathrm{GCCF}(z) = z \qquad \text{if } z \in \mathcal{F} \text{ where}$$
$$\mathcal{F} = \{ z \in \mathbf{C} \mid 0 \leq \mathrm{Re}(z) \leq 1/2 \text{ and } |z| \geq 1 \}. \qquad (10')$$

Lattice Reduction. Gaussian continued fractions like (9) are also of special interest in relation to the *reduction of lattice bases* in dimension 2, see for instance [9]. The 2-dimensional lattice $L = L(u, v)$ generated by the basis of complex numbers $(u, v) \in \mathbf{C}^2$ is defined as

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

Two different bases (u, v) and (u', v') of the same lattice are connected by a unimodular transformation with integer coefficients,

$$\begin{pmatrix} u' \\ v' \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix},$$

with $ad - bc = \pm 1$. Given a lattice, an important operation consists in reducing it. Informally, the reduction of a lattice, defined by one of its bases (u, v), consists in finding a "good" basis (u^*, v^*) in the sense that it is nearly orthogonal[1].

Viewed in terms of their ratios $z = v/u$ and $z^* = v^*/u^*$, reducing a lattice means that we start from an arbitrary z and try to derive a z^* that is "near" the imaginary axis. (Formally, this means $z^* \in \mathcal{F}$.)

In the reduction perspective, the tools of the reduction trade that correspond to standard linear operations on bases, namely

$$(u, v) \mapsto (v, u), \quad (u, v) \mapsto (u, u + v), \quad (u, v) \mapsto (u, -v),$$

are the homographic transformations,

$$S(z) := \frac{1}{z}; \qquad T(z) := z + 1; \qquad J(z) := -z.$$

We then see that the process of reducing a lattice $L(u, v)$ and the process of expanding into a Gaussian continued fraction the complex number $z = v/u$, with $(u, v) \in \mathbf{C}^2$, are two aspects of one and the same thing.

[1] Out of such a basis *inter alia* closest points and the Voronoi diagram can be determined easily, see Fig. 1.

Higher dimensions. The interest of Gauss' reduction process is also largely due to the fact that it enters as a basic component of lattice reduction algorithms in higher dimensions, most notably the Lenstra–Lenstra–Lovasz's algorithm, nicknamed LLL [6, 12]. These reduction algorithms have far reaching implications in numerous areas, like polynomial factorization, cryptography and the like. The present paper could be seen as just a very first step in the direction of the average case analysis of this rich class of semi-numerical algorithms. (We briefly discuss some conjectures that naturally suggest themselves at the end of the paper.)

Linear Transformations. Our subject is closely related to the classical study of the so-called *modular group* [3, 10, 11].

The three transformations S, T, J are known to generate the group $\mathbf{U}$ of all unimodular transformations

$$z \longrightarrow \frac{az + b}{cz + d}$$

where a, b, c, d are integers that satisfy $ad - bc = \pm 1$. The continued fraction algorithm can be used to effectively decompose a transformation of $\mathbf{U}$ in terms of the generators S, T, J. As we shall see in Section 2, the Gauss reduction algorithm can serve as an alternative algorithm in order to compute such a decomposition.

Here is what now awaits the reader.

This paper proposes to study the *average case* complexity of the complex continued fraction algorithm and thus of lattice reduction as well. The results differ somewhat from the real variable case summarized by Eq. (4) or (8). The worst case of Gauss' reduction algorithm applied to numbers of $\mathbf{Q}(i)$ whose size is bounded by N was determined by Vallée [13] and is of the form

$$\log_{1+\sqrt{2}} N + O(1), \qquad (11)$$

which is the same bound as in the centered Euclidean Algorithm. However, in sharp contrast with this situation, the average case for such numbers is *asymptotically constant*. Actually, considering a suitable variant of the algorithm, we are able to prove that the average case complexity of the "core" of the algorithm is of the form $(16/\pi)\beta + o(1)$, where the constant β admits the nice closed form,

$$\beta = \frac{\pi}{4\zeta(4)} \sum_{m \geq 1} \frac{1}{m^2} \sum_{n = \lceil m\phi^{-2} \rceil}^{\lfloor m\phi^{-1} \rfloor} \frac{1}{n^2}, \qquad (12)$$

with ϕ being the golden ratio, and $\zeta(4) = \pi^4/90$. Numerically, $\beta \approx .2138681$, so that the average case complexity is small, being close to 1.09. In addition, we prove that the probability distribution of the algorithm's cost has an exponential tail. For instance, the

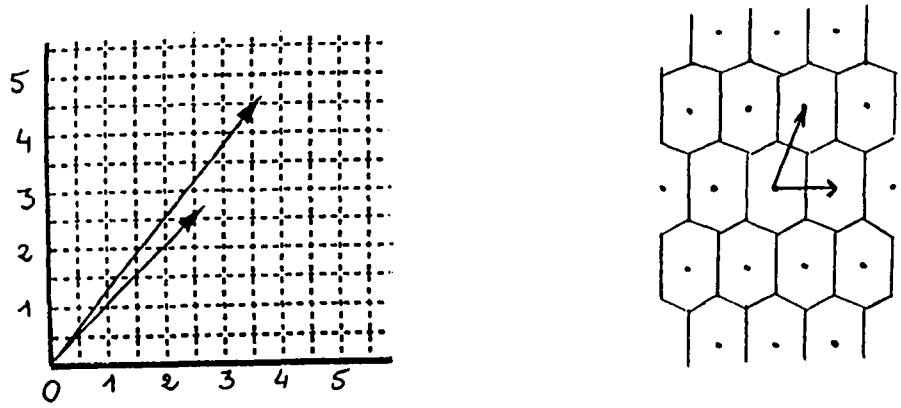

Figure 1. The pair of complex numbers $(u, v) = 5/2 + i\sqrt{7}, 4 + 2i\sqrt{7})$ (represented on the left) illustrates a "skew" representation of a lattice L with long and narrow meshes. Applying the reduction algorithm of Gauss to $z = v/u$ leads to $z^* = v^*/u^*$ with a new basis $(u^*, v^*) = (1, (1 + i\sqrt{7})/2)$ (represented on the right) that gives a better presentation on the same lattice L: on this basis—which is such that z^* lies in the fundamental domain $\mathcal{F}$—closest points and the Voronoi cell decomposition become apparent.

number of iterations hardly ever (i.e., with probability $\ll 10^{-4}$) exceeds 5.

The Gauss reduction algorithm, is reviewed in Section 1. In order to make the algorithm amenable to analysis, we first carry out a normalizing task. We show that the algorithm decomposes into some preparation steps and a "core" (called Algorithm In-Gauss) that itself consists of a succession of elementary geometric transformations of quite a "regular" form. It is from this core that most of the complexity of the algorithm arises in the worst case, and this is also the place where the interesting operations take place.

Section 2 analyses this suitably regularized version of the algorithm and it constitutes the kernel of our analysis. The major point is a characterisation of the reduction transformations in number-theoretic terms which is combined with elementary geometric arguments. The average-case complexity is obtained in terms of the β constant. Also as correlates of this analysis we obtain through Theorem 2 an exponential tail result—large deviations are thus extremely unlikely—as well as quantitative estimates that relate the discrete lattice-points model to our continuous probabilistic model (Theorem 3).

Section 3 brings the analysis to its final conclusion by taking care of all the steps in the standard algorithm (Theorem 4). The developments there are a continuation of the techniques of Section 2, only a little more intricate, and they lead to a full analysis.

Our analysis makes occasional use of elementary properties of continued fractions and of linear fractional transformations. On these rich topics, we refer the reader to the classical treateses of Hardy and Wright [4], Ford [3], Schoenberg [10], and Serre [11].

1 The basic algorithms

Gauss's reduction generalizes the centered continued fraction algorithm, as we saw with Eq (10,10'). In this section, we examine with a geometrical point of view the iterative version of the algorithm.

The algorithm operates on the right half plane $\mathcal{R}$ of non-real complex numbers with a non-negative real part,

$$\mathcal{R} = \{z \in \mathbf{C} \mid \mathrm{Im}(z) \neq 0 \text{ and } \mathrm{Re}(z) \geq 0\}$$

and starts with a complex z that lies inside the half disk $\mathcal{C} = \{z \in \mathcal{R} \mid |z| \leq 1\}$. The Gauss algorithm uses S, T and J to bring z inside the closed strip $\mathcal{B}$ of numbers of $\mathcal{R}$ with a real part between 0 and 1/2, and simultaneously outside the open half disk $\mathcal{C}^\circ$. The algorithm always terminates and it stops when z has been brought into the domain $\mathcal{F}$ defined in (10'),i.e., $\mathcal{F} = \mathcal{B} \setminus \mathcal{C}^\circ$.

The collection of all $\rho(\mathcal{F})$, $\rho \in \mathbf{U}$ forms a *tessellation* of the plane in the following sense. First, given $\rho, \sigma \in \mathbf{U}$, two distinct domains $\rho(\mathcal{F})$ and $\sigma(\mathcal{F})$ are *quasi-disjoint* in the sense that their intersection, if

non–empty, is wholly contained in their frontier (and thus, in particular, has measure 0!). Second, the collection of the $\rho(\mathcal{F})$ covers the complex plane.

Thus the Gauss reduction maps $\mathcal{C}$ into $\mathcal{F}$. Our probabilistic model for the analysis is the simplest possible with a *uniform* model over the legal inputs to the algorithm: In this continuous model, the probability that point z belongs to a domain $\Omega \subset \mathcal{C}$ is equal to the ratio of the areas $|\Omega|/|\mathcal{C}|$. At the end of the next section, we shall show that analysis under this model coincides asymptotically with analysis under discrete lattice point models.

Here comes now the iterative definition of the basic Gauss algorithm: After a preliminary step, it performs a sequence of steps which we call Step-Gauss; in each of these steps, one uses successively the homographic transformations S, T, J. More precisely, we define:

$$\text{Step-Gauss}(z) = \epsilon(\frac{1}{z}) \left(\frac{1}{z} - \lceil \text{Re}(\frac{1}{z}) \rfloor \right)$$
$$\text{where } \epsilon(z) = \text{sign}(\text{Re}(z) - \lceil \text{Re}(z) \rfloor), \quad (13)$$

and we take $\text{sign}(0) = 1$. The preliminary step brings z into $\mathcal{C}_1 = \mathcal{C} \cap \mathcal{B}$, with the help of T and J; it is only effective when z belongs to $\mathcal{C}_2 = \mathcal{C} \setminus \mathcal{C}_1$.

Algorithm Gauss(z);
Input: a number z that belongs to $\mathcal{C}$.
Output: a number z that belongs to $\mathcal{F}$.

Pre-Gauss: **If** $\text{Re}(z) > 1/2$ **then** $z := 1 - z$;

While $z \in \mathcal{C}^\circ$ **do** $z :=$ Step-Gauss(z);

It is clear that the Gauss Algorithm has an execution trace which is precisely isomorphic to the Gaussian fraction process defined by the rule (10). If n is the number of calls of Step-Gauss, we denote by z_{-1} the input, by z_0 the value of z after the preliminary step, and by z_i $(1 \le i \le n)$ the value of z at the end of the i-th call of Step-Gauss. Each of these calls uses a pair (m_i, ϵ_i): m_i is the positive integer found in the translation step; it is defined by $m_i = \lceil \text{Re}(1/z_{i-1}) \rfloor$, while $\epsilon_i = \pm 1$, is defined by the relation $\epsilon_i = \text{sign}(\text{Re}(1/z_{i-1}) - m_i)$.

So each elementary transition can be written, for $1 \le i \le n$, as

$$z_i = J_{\epsilon_i} T^{-m_i} S(z_{i-1}) \text{ or } z_{i-1} = \frac{1}{m_i + \epsilon_i z_i}. \quad (14)$$

(We denote by J_ϵ the transformation defined by $J_\epsilon(z) = \epsilon z$, so that $J_{+1} = I$, the identity transformation, and $J_{-1} = J$.)

In fact, as we are going to show, the execution traces

$$z_{-1}, z_0, z_1, \ldots, z_{n-2}, z_{n-1}, z_n$$

of the Gauss algorithm have a particular structure: With possible exceptions for z_{n-2} and z_{n-1} *only*, and naturally excluding $z_n \in \mathcal{F}$, the z_j with $j \ge 0$ belong to the open disk $\mathcal{D}$ whose diameter is $[0, 1/2]$. The posssible exceptions are related to the occurrence in the algorithm of special (m, ϵ) pairs for which we define the set D,

$$D = \{(m, \epsilon) \mid m \ge 2,\ m + \epsilon \ge 2\}, \quad (15)$$

and its complement

$$\bar{D} = \{(0, +1), (1, -1), (1, +1), (2, -1)\}.$$

The structure of a trace is precisely described by the following proposition.

Proposition 1.- *If z_i is the current value of z before the $(i+1)$-st call of* Step-Gauss *$(0 \le i \le n-1)$, then the following three properties hold true:*

(i) If the point z_i belongs to the open disk $\mathcal{D}$ with diameter $[0, 1/2]$, then the pair $(m_{i+1}, \epsilon_{i+1})$ belongs to the set D.

(ii) Conversely, if the pair $(m_{i+1}, \epsilon_{i+1})$ belongs to the set D, then the point z_i belongs to the closure of the disk $\mathcal{D}$.

(iii) Let ℓ be the least $i \ge 0$ such that z_i does not lie in $\mathcal{D}$. The only possible values for ℓ are $\ell = n-2$, $\ell = n-1$, $\ell = n$, and in each of these three cases, we have respectively,

$$z_\ell = ST^2JS(z_n)$$
$$z_\ell = S(z_n),\ z_\ell = ST(z_n),\ z_\ell = STJ(z_n),\ z_\ell = ST^2J(z_n)$$
$$z_\ell = z_n$$

Proof.-(cf Fig. 2) Remark that the domain $S(\mathcal{D})$ is the strip $\text{Re}(z) > 2$. So the facts *(i)* and *(ii)* are clear. If now z is a point of $\mathcal{B} \setminus \mathcal{D}$, then $S(z)$ is a point of $S(\mathcal{B}) \setminus S(\mathcal{D})$. But, this last domain is the intersection of the strip $0 \le \text{Re}(z) \le 2$ with the outside of the open disk of diameter $[0, 2]$. Thus the domain $S(\mathcal{B}) \setminus S(\mathcal{D})$ is equal to the union of six domains that are transforms of $\mathcal{F}$; so $\mathcal{B} \setminus \mathcal{D}$ is also the union of six domains $\sigma(\mathcal{F})$ which are the transforms of $\mathcal{F}$ by the six elements σ of $\mathbf{U}$ belonging to $\mathbf{S}$

$$\mathbf{S} = \{I, S, STJ, ST, ST^2J, ST^2JS\}. \quad (16)$$

These six domains are *quasi-disjoint.* ∎

In the next section, we shall concern ourselves with the segment of those computations of the Gauss reduction algorithm that operate on the disk $\mathcal{D}$. This represents the "core" of the algorithm since it includes all steps save at most three (the preparation step of PreGauss and the last two steps described by Prop. 1), namely

$$z_0, z_1, \ldots, z_\ell.$$

Equivalently, we may introduce a simple variant of the Gauss algorithm, called In-Gauss and defined below; our problem then transforms into that of analyzing the In-Gauss algorithm.

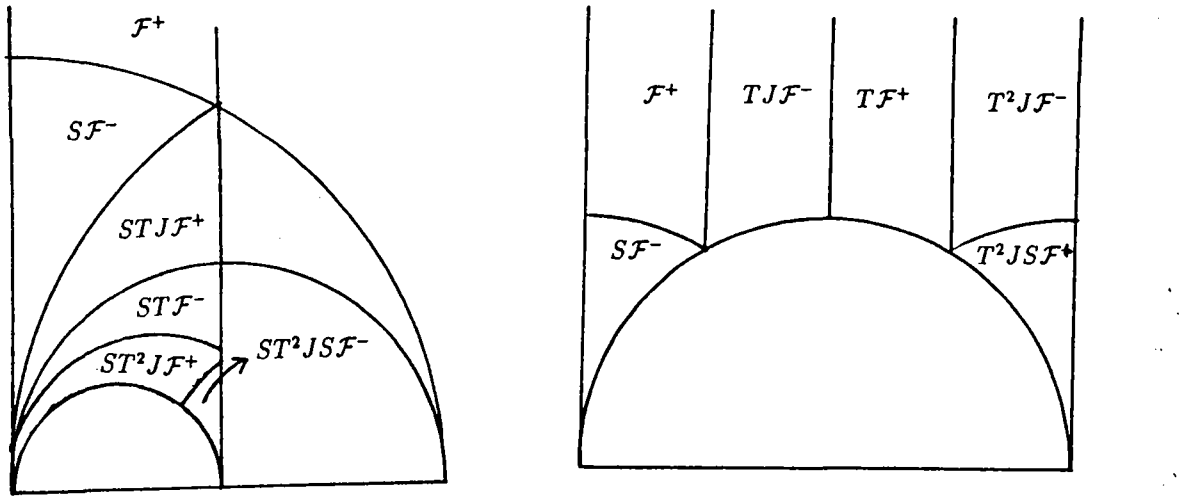

Figure 2. The upper part of domains $\mathcal{B} \setminus \mathcal{D}$ and $S(\mathcal{B}) \setminus S(\mathcal{D})$. ($\mathcal{F}^+$ and $\mathcal{F}^-$ denote the components of $\mathcal{F}$ in the upper and lower half planes.)

Algorithm In-Gauss(z);
Input: a number z that belongs to $\mathcal{D}$.
Output: a number z that belongs to $\mathcal{B} \setminus \mathcal{D}$.

While $z \in \mathcal{D}$ **do** $z :=$ Step-Gauss(z);

With our previous notations, the execution traces of In-Gauss are thus $z_0 z_1 \cdots z_\ell$. We shall let $L(z)$ denote the number of iterations (the value of ℓ) performed by this algorithm when presented with input $z = z_0$. If we take z uniformly over the disk $\mathcal{D}$, L becomes a random variable. The next section is then devoted to studying the expectation $E(L)$ of L.

2 The regularized algorithm

The main theorem of this paper gives an evaluation of the expected number of loops $E(L)$ performed by the regularized algorithm In-Gauss. As we saw already this represents the analysis of the core of the reduction algorithm, since in the general case, almost all the interesting computations of a reduction take place there.

Two major ingredients enter the analysis. First, the linear fractional transformations $\mathbf{G}$ defined by the regularized algorithm are "regular" enough, so that the expected cost $E(L)$ can be expressed as a simple sum indexed by $\mathbf{G}$ (Prop. 2). Second, the transformations of $\mathbf{G}$ can be characterized in arithmetical terms (Prop. 3), so that $E(L)$ becomes expressible in some explicit form.

This analysis is the essential part of the paper. Technical refinements of it later lead to a full analysis of the complete reduction process which is discussed in Section 3.

Theorem 1.- *The number L of iterations of algorithm* In-Gauss *has an expectation $E(L)$ equal to $(16/\pi)\beta$ with*

$$\beta = \frac{\pi}{4\zeta(4)} \sum_{m \geq 1} \frac{1}{m^2} \sum_{n=\lceil m\phi^{-2} \rceil}^{\lfloor m\phi^{-1} \rfloor} \frac{1}{n^2},$$

where ϕ is the golden ratio. Numerically $(16/\pi)\beta \approx$ 1.08922.

First we need some basic definitions. We have introduced the variant In-Gauss in order to have all the pairs (m, ϵ) satisfying conditions (D). We introduce the set $\mathbf{G}$ of transformations defined by this algorithm, and its subsets $\mathbf{G}_\ell$ corresponding to ℓ iterations:

$$\mathbf{G}_\ell = \{\sigma = \prod_{i=1}^{\ell} ST^{m_i} J_{\epsilon_i}\}, \qquad (17)$$

where each pair (m_i, ϵ_i) belongs to the set D, and

$$\mathbf{G} = \bigcup_{\ell \geq 0} \mathbf{G}_\ell \quad \text{and} \quad \mathbf{G}_+ = \bigcup_{\ell > 0} \mathbf{G}_\ell.$$

The transformations of $\mathbf{G}$ are called *regular*. Observe that, when the algorithm transforms z_0 into z_ℓ, it is the relation $z_0 = \sigma(z_\ell)$ that defines $\sigma \in \mathbf{G}$.

Proposition 2.- *With $\mathbf{G}$ the class of regular transformations, the expected number of steps of Algorithm*

In-Gauss satisfies

$$E(L) = \frac{16}{\pi} \sum_{\rho \in \mathbf{G}} |\rho(\mathcal{D})| = 4 \sum_{\rho \in \mathbf{G}} |\rho(0) - \rho(1/2)|^2. \quad (18)$$

Proof.- Algorithm In-Gauss performs ℓ iterations on the input $z_0 \in \mathcal{D}$ if and only if there exists a pair (σ, z_ℓ) of $\mathbf{G}_\ell \times (\mathcal{B} \setminus \mathcal{D})$ such that $z_0 = \sigma(z_\ell)$. Since domains $\sigma(\mathcal{B} \setminus \mathcal{D})$ are quasi-disjoint, we deduce the equality

$$\Pr[L = \ell] = \frac{16}{\pi} \sum_{\sigma \in \mathbf{G}_\ell} |\sigma(\mathcal{B} \setminus \mathcal{D})|.$$

On the other hand, an element σ of $\mathbf{G}_\ell$ is sufficiently "regular" so that it can be split into two factors of $\mathbf{G}$, in ℓ different ways; in symbols:

$$\sigma = \rho\rho' \text{ with } \rho' \in \mathbf{G}_k,\ \rho \in \mathbf{G}_{\ell-k} \text{ and } 1 \le k \le \ell.$$

From these properties we derive a new form for the expectation $E(L)$

$$E(L) = \frac{16}{\pi} \sum_{\rho \in \mathbf{G}} |\rho(\sum_{\rho' \in \mathbf{G}_+} \rho'(\mathcal{B} \setminus \mathcal{D}))|.$$

But, by definition of our algorithm, as summarized by Eq. (17), and, since $\mathcal{B} \setminus \mathcal{D}$ is a quasi-disjoint union of domains $\sigma(\mathcal{F})$, for $\sigma \in \mathbf{S}$, we get

$$\mathcal{D} = \sum_{\rho' \in \mathbf{G}_+} \rho'(\mathcal{B} \setminus \mathcal{D}), \quad (19)$$

(with a quasi-disjoint sum) so that

$$E(L) = \frac{16}{\pi} \sum_{\rho \in \mathbf{G}} |\rho(\mathcal{D})|.$$

Since the domain $\rho(\mathcal{D})$ is the disk of diameter $[\rho(0), \rho(1/2)]$, we obtain finally the statement of the proposition. ∎

The difference $|\rho(0) - \rho(1/2)|$ is a function of the coefficients c and d in the homographic transformation $\rho : z \to (az + b)/(cz + d)$ which evaluates to

$$|\rho(0) - \rho(\frac{1}{2})|^2 = \frac{1}{d^2(2d + c)^2}.$$

So we propose to characterize those pairs of coprime integers (c, d) that are denominator coefficients of a unimodular transformation from $\mathbf{G}$. The fundamental result is the following.

Proposition 3.- *Let $\sigma(z) = (az+b)/(cz+d)$ be a regular transformation, $\sigma \in \mathbf{G}_+$. Such a transformation is uniquely determined by its denominator coefficients (c, d) with $d > 0$. The denominator pairs are characterised by the conditions:*

(i) $d > 1$,

(ii) $\gcd(c, d) = 1$.

(iii) $-d/\phi^2 < c < d/\phi$.

For our developments, we shall resort to a fundamental relation between *Gaussian fractions* and *centered continued fraction expansions*. The proof relies on a few definitions and two lemmas.

Consider a transformation $\rho : z \to (az + b)/(cz + d)$ of $\mathbf{G}_+$. It sends $\mathcal{B} \setminus \mathcal{D}$ into $\mathcal{D}$, and the three "cusps" that it defines:

$$\rho(i\infty) = \frac{a}{c},\ \rho(0) = \frac{b}{d},\ \rho(1/2) = \frac{a + 2b}{c + 2d}$$

are three rational points of the interval $]0, 1/2]$. One can always choose d positive; then b is also positive, and the sign of a and c (they have the same sign) is given by the sign of the last ϵ (i.e., ϵ_ℓ) found in Algorithm In-Gauss. So, when given a pair (c, d), the pair (a, b) satisfies the two conditions:

(i) $ad - bc = \pm 1$,

(ii) a/c and b/d are rational numbers of $]0, 1/2]$.

This means that a/b is the last convergent of c/d in the proper centered continued fraction expansion of c/d and this condition uniquely determines (a, b) from the pair (c, d): (a, b) is the *associate* of (c, d).

We now use the relation between the Gauss algorithm and the centered Euclidean algorithm. In fact, when applying In-Gauss to a complex number z belonging to $\rho(\mathcal{B} \setminus \mathcal{D})$, one obtains at the same time the canonical decomposition of the element ρ (in terms of generators S, T, J) and the centered continued fraction expansion of the three cusps of $\rho(\mathcal{B} \setminus \mathcal{D})$: This is seen by replacing in the Gaussian fraction

$$z = \rho(t) = \cfrac{1}{m_1 + \cfrac{\epsilon_1}{m_2 + \cfrac{\epsilon_2}{\ddots \cfrac{\epsilon_{\ell-2}}{m_{\ell-1} + \cfrac{\epsilon_{\ell-1}}{m_\ell + \epsilon_\ell t}}}}}$$

the complex t by each of the three values

$$t = \infty,\ t = 0,\ t = (1/2),$$

which leads to three centered continued fraction expansions. These expansions may be "improper" though, because the last stage may involve $m - 1/2$ (with an integer $m \ge 3$), while we would have $(m-1)+1/2$ if the expansion was proper. Here, we accept improper continued fraction expansions. Then, it is clear that the three rationals $\rho(i\infty), \rho(0), \rho(1/2)$ are three consecutive convergents, either proper or improper, of the last quantity $\rho(1/2)$.

We thus encapsulate the previous remarks into a definition.

Definition.- We say that an integer c is an (m,ϵ)-*antecedent* of a positive integer d if there exist two integers a and b determining two irreducible fractions a/c and b/d of interval $]0,(1/2)]$ that satisfy the conditions:

(i) a/c is the last convergent of b/d in the centered continued fraction expansion (proper or improper) of b/d,

(ii) The last stage of the centered continued fraction expansion of b/d involves the pair (m,ϵ).

So we have found: *If a unimodular transformation $\rho : z \to (az+b)/(cz+d)$ (with $d>1$) belongs to $\mathbf{G}_+$, then the integer c is an antecedent of integer d and the pair (a,b) is the associate of the pair (c,d).*

Here, we consider the rationals $\rho(1/2)$ whose last stage in their centered continued fraction expansion (either proper or improper) involves the integer 2; so, the 2-antecedents play a central rôle in the sequel, and we call a *grand-antecedent* an antecedent that is not a 2-antecedent.

We first derive a relation between successive antecedents that we use further in an inductive characterization of grand-antecedents and 2-antecedents.

Lemma 1.- *An integer d is an (m,ϵ)-antecedent of integer f if and only if there exists an integer c that satisfies the three conditions:*

(i) $\mathrm{sign}(c) = \epsilon$,

(ii) c is an antecedent of $|d|$,

(iii) $f = m|d| + c$.

A 2-antecedent is always positive. If c is a grand-antecedent of d, so is $-c$.

With Lemma 1, we can finally arrive at the characterization of grand-antecedents and 2-antecedents.

Lemma 2.- *Let $d \geq 2$ be an integer. The set of the grand-antecedents of d, and the set of the 2-antecedents of d are respectively equal to*

$$Z^*(d) \bigcap \mathcal{H}(d) \quad \text{and} \quad Z^*(d) \bigcap \mathcal{K}(d), \qquad (20)$$

where $Z^*(d)$ denotes the set of integers coprime with d, and $\mathcal{H}(d), \mathcal{K}(d)$ are the intervals

$$\mathcal{H}(d) = [-\frac{d}{\phi^2}, \frac{d}{\phi^2}] \quad \text{and} \quad \mathcal{K}(d) = [\frac{d}{\phi^2}, \frac{d}{\phi}], \qquad (21)$$

with ϕ being the golden ratio.

Thus Prop. 3 is now established. Combining its results with Prop. 2, we obtain the form of β as

$$\frac{\pi}{4} \sum_{f \geq 1} \frac{1}{f^2} \sum_{d \in \mathcal{K}(f) \cap Z^*(f)} \frac{1}{d^2} = \frac{\pi}{4\zeta(4)} \sum_{f \geq 1} \frac{1}{f^2} \sum_{d \in \mathcal{K}(f)} \frac{1}{d^2}.$$

We eliminate the condition $(d,f) = 1$ in the second form of β, provided that we divide the result by $\zeta(4) = \pi^4/90$. This remark concludes the proof of Theorem 1.

Distributional Bounds. The tools that we have just developed also enable us to derive qualitatitive information on the probability distribution of the quantity L.

Theorem 2.- *The probability distribution of the number L of iterations of the In-Gauss algorithm admits an exponential tail:*

$$\Pr[L \geq \ell] \leq \frac{\phi^2}{(1+\sqrt{2})^{2\ell-2}} \quad \text{for} \quad \ell \geq 1.$$

Discrete Probabilistic Models. We now digress a little and examine the discrete lattice-point model for the probabilistic analysis. Under that model, one considers numbers of $\mathbf{Q}(i)$ with denominator equal to n, and we naturally exclude real number where the algorithm fails to terminate. We denote this set by $\mathbf{Q}^{(n)}$. Furthermore, we restrict attention to inputs inside $\mathcal{D} \cap \mathbf{Q}^{(n)}$ that are legal inputs to algorithm In-Gauss. We can estimate the expectation $E(L_n)$ of the number of iterations of In-Gauss, when the inputs are uniformly distributed inside $\mathcal{D} \cap \mathbf{Q}^{(n)}$.

Theorem 3.- *In the discrete model, the expectation $E(L_n)$ satisfies*

$$E(L_n) = E(L) + O(\frac{\log n}{n}) = \frac{\pi}{16}\beta + O(\frac{\log n}{n}).$$

3 The standard Gauss algorithm

Our purpose is now to estimate the expectation of the random variable N that is equal to the number of iterations of the standard algorithm of Gauss. More precisely, if

$$z_{-1}, z_0, z_1, \ldots, z_{n-2}, z_{n-1}, z_n$$

is the execution trace of the Gauss algorithm, we shall let $N(z)$ denote the number of iterations performed by this algorithm when presented with input $z = z_{-1}$. So we have

$$N(z) = n \text{ if } z_{-1} = z_0 \text{ and } N(z) = n+1 \text{ if } z_{-1} \neq z_0.$$

We take into account the preliminary step only if it is actually used. The probabilistic model assumes that the input z is uniformly distributed on the unit half disk $\mathcal{C}$.

$$K(c,d) = \left(\frac{2}{3c^2(c+2d)^2} + \frac{1}{3d^2(2c-d)^2}\right)\pi + \left(\frac{1}{c(c+2d)} - \frac{1}{d(2c-d)}\right)\frac{\sqrt{3}}{2(d^2-c^2)}$$
$$+\left(\frac{1}{c^2(2d+c)^2} - \frac{1}{(d^2-c^2)^2}\right)\arcsin\frac{\sqrt{3}}{2}\frac{(d^2-c^2)}{d^2+cd+c^2} - \left(\frac{1}{d^2(2c-d)^2} - \frac{1}{(d^2-c^2)^2}\right)\arcsin\frac{\sqrt{3}}{2}\frac{(d^2-c^2)}{d^2-cd+c^2},$$
$$H(c,d) = \left(\frac{2}{3c^2(2d+c)^2} - \frac{1}{3c^2(2d-c)^2}\right)\pi + \left(\frac{1}{c(2d+c)} - \frac{1}{c(2d-c)}\right)\frac{\sqrt{3}}{2(d^2-c^2)}$$
$$+\left(\frac{1}{c^2(2d+c)^2} - \frac{1}{(d^2-c^2)^2}\right)\arcsin\frac{\sqrt{3}}{2}\frac{(d^2-c^2)}{c^2+cd+d^2} - \left(\frac{1}{c^2(2d-c)^2} - \frac{1}{(d^2-c^2)^2}\right)\arcsin\frac{\sqrt{3}}{2}\frac{(d^2-c^2)}{d^2-cd+c^2}.$$

Table 1. The definitions of functions $K(c,d)$ and $H(c,d)$ of Theorem 4.

Theorem 4.- *The expected number of reduction steps $E(N)$ for the standard reduction algorithm* Gauss *is equal to*

$$E(N) = \frac{5}{3} - \frac{\sqrt{3}}{2\pi} + \frac{4}{\pi}\gamma,$$

where γ is the constant

$$\gamma = \gamma_0 + \frac{1}{\zeta(4)}\sum_{d\geq 3}\left[\sum_{\substack{c=\lceil d\phi^{-1}\rceil \\ c\neq d/2}}^{\lfloor d\phi^{-1}\rfloor} K(c,d) + \sum_{c=1}^{\lfloor d\phi^{-2}\rfloor} H(c,d)\right],$$

with

$$\gamma_0 = \frac{\sqrt{3}}{180} - \frac{16}{225}\arcsin\frac{3\sqrt{3}}{14} + \frac{43\pi}{675},$$

and $H(c,d)$, $K(c,d)$ are given by Table 1. Numerically, we find $E(N) \approx 1.53$.

There are two steps in the proof: (i) Expressing $E(N)$ in terms of the regular semigroup $\mathbf{G}$; (ii) Using elementary geometry and the analysis of $\mathbf{G}$ carried out in the last section in order to estimate the constant γ.

Lemma 3.- *The expectation $E(N)$ satisfies*

$$E(N) = \frac{5}{3} - \frac{\sqrt{3}}{2\pi} + \frac{4}{\pi}\gamma$$

where

$$\gamma = \beta + \sum_{\rho\in\mathbf{G}} |\rho ST^2 JS(\mathcal{F})| - \sum_{\rho\in\mathbf{G}_+} |\rho(\mathcal{F})|.$$

Proof. *A*). We first need to dispose of the preliminary step "PreGauss". We let $E(N^*)$ denote the expected number of reduction steps when z is taken uniformly over $\mathcal{C}_1 = \mathcal{C}\cap\mathcal{B}$. In that case the preparation step of PreGauss is the identity transformation.

The two expectations $E(N)$ and $E(N^*)$ are related as follows:

$$E(N) = E(N^*)\left(\frac{2}{3} + \frac{\sqrt{3}}{\pi}\right) + \left(1 - \frac{3\sqrt{3}}{2\pi}\right). \qquad (22)$$

To justify this, we remark that Pre-Gauss$(z) = z$ if z belongs to $\mathcal{C}_1$. Otherwise, if z belongs to $\mathcal{C}_2$, we have

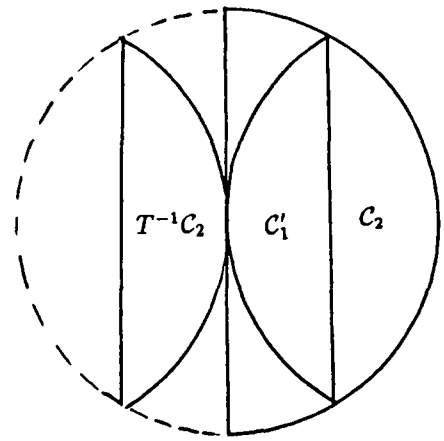

Figure 3. The domains $\mathcal{C}_1, \mathcal{C}_2$ and $\mathcal{C}_1'$.

Pre-Gauss$(z) = 1 - z$ and Pre-Gauss$(\mathcal{C}_2) = \mathcal{C}_1'$, the symmetrical of $\mathcal{C}_1$ with respect to line $\mathrm{Re}(z) = 1/2$.

B). Next we need to examine the situations where one or two more reduction steps are effected after algorithm In-Gauss in order to complete the reduction process.

It is convenient to let $E(L^*)$ denote the expected value of L relative to the In-Gauss algorithm when the input z is taken uniformly over $\mathcal{C}_1$ instead of $\mathcal{D}$. According to Prop. 1, we have

$$E(N^* - L^*) = \Pr[N^* - L^* \geq 1] + \Pr[N^* - L^* \geq 2].$$

We can estimate each term of this sum:

$$\Pr[N^* - L^* \geq 1] = \frac{1}{|\mathcal{C}_1|}\sum_{\rho\in\mathbf{G}} |\rho(\mathcal{C}_1 \setminus \mathcal{D})|$$

$$= \frac{|\mathcal{C}_1| - |\mathcal{D}|}{|\mathcal{C}_1|} + \frac{1}{|\mathcal{C}_1|}\sum_{\rho\in\mathbf{G}_+} |\rho(\mathcal{C}_1 \setminus \mathcal{D})|.$$

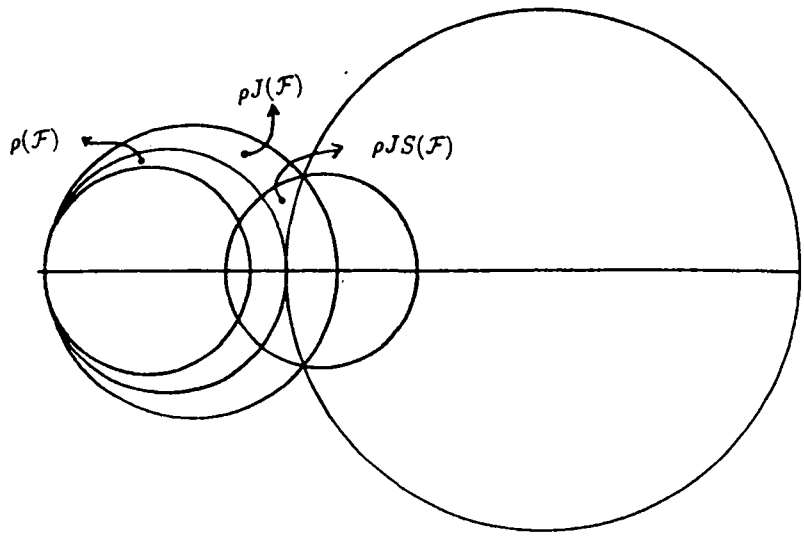

Figure 4. The triangle domain $\rho(\mathcal{F})$ and the adjacent domains $\rho J(\mathcal{F}), \rho JS(\mathcal{F})$.

But, we have $\mathcal{B} \setminus \mathcal{D} = (\mathcal{C}_1 \setminus \mathcal{D}) \bigcup \mathcal{F}$, and we get

$$\Pr[N^* - L^* \geq 1] = 1 - \frac{1}{|\mathcal{C}_1|} \sum_{\rho \in \mathbf{G}_+} |\rho(\mathcal{F})|.$$

On the other hand,

$$\Pr[N^* - L^* \geq 2] = \frac{1}{|\mathcal{C}_1|} \sum_{\rho \in \mathbf{G}} |\rho ST^2 JS(\mathcal{F})|.$$

At this stage, we introduce the constant α defined by

$$\alpha = \sum_{\rho \in \mathbf{G}} |\rho ST^2 JS(\mathcal{F})| - \sum_{\rho \in \mathbf{G}_+} |\rho(\mathcal{F})|.$$

Using it, we obtain:

$$E(N^*) - E(L^*) = 1 + \frac{\alpha}{|\mathcal{C}_1|}.$$

Furthermore, we have

$$E(L^*) = E(L)\frac{|\mathcal{D}|}{|\mathcal{C}_1|} = \frac{\beta}{|\mathcal{C}_1|},$$

and finally, with Eq. (22), we obtain the result. ■

In order to complete the proof of Theorem 4, we need to estimate the constant γ of Lemma 3. A more concise form of that constant is obtained if we regroup terms using two special subsets, $\mathbf{H}^+$ and $\mathbf{K}$, of $\mathbf{G}$.

An element ρ of $\mathbf{G}_+$ can be written as $\rho = \tau ST^m J_\epsilon$, with $\tau \in \mathbf{G}$ and (m, ϵ) satisfying conditions (D). We individuate two cases according to the value of the last pair (m, ϵ); if it is equal to $(2, +1)$, we say that ρ is an element of $\mathbf{K}$; if $\epsilon = +1$ and $m \geq 3$, we say that ρ is an element of $\mathbf{H}^+$. An element ρ of $\mathbf{H}^+$ has a twin defined to be equal to ρJ; an element ρ of $\mathbf{K}$ is a *priori* "alone", but we define its twin as ρJS. This notion is guided by the geometry of adjacent triangular regions, see Fig. 4. The twin function is involutive.

Remark that we have a good transfer of these definitions on the pair built on the denominator coefficients (c, d) of ρ. Elements of $\mathbf{K}$ have their c that is an element of $\mathcal{K}(d)$, while, for elements of $\mathbf{H}^+$, coefficient c is a positive element of $\mathcal{H}(d)$. So, using Lemma 2, we obtain a good characterization of these two subsets of $\mathbf{G}$. (Here, we identify the element ρ of $\mathbf{G}$ with its pair of denominator coefficients.)

$\rho = (c, d)$ belongs to $\mathbf{K}$ iff

$$\gcd(c, d) = 1 \text{ and } \frac{d}{\phi^2} \leq c \leq \frac{d}{\phi}$$

$\rho = (c, d)$ belongs to $\mathbf{H}^+$ iff

$$\gcd(c, d) = 1 \text{ and } 0 < c \leq \frac{d}{\phi^2}.$$

The twin of an element (c, d) of $\mathbf{H}^+$ is $(-c, d)$ while the twin of an element (c, d) of $\mathbf{K}$ is $(-d, c)$.

Furthermore, we remark that the disk $\mathcal{D}$ coincides with $ST^2(\mathcal{R})$. Thus, we get a new expression for the constant β:

$$\beta = \sum_{\tau \in \mathbf{G}} |\tau(\mathcal{D})| = \sum_{\rho \in \mathbf{K}} |\rho(\mathcal{R})|.$$

We use now the two subsets $\mathbf{H}^+$ and $\mathbf{K}$ to write the sum that expresses the constant γ as the sum of two constants $\gamma(\mathbf{H}^+)$ and $\gamma(\mathbf{K})$ that we define as follows

$$\gamma(\mathbf{H}^+) = \sum_{(c,d) \in \mathbf{H}^+} H(c, d)$$

with $H(c,d) = -|\rho(\mathcal{F})| - |\rho J(\mathcal{F})|$

$$\gamma(\mathbf{K}) = \sum_{(c,d)\in\mathbf{K}} K(c,d)$$

with $K(c,d) = |\rho(\mathcal{R})| + |\rho JS(\mathcal{F})| - |\rho(\mathcal{F})|$.

The terms $K(c,d)$ or $H(c,d)$ are then computed using the geometry of circular triangles of the type $\rho(\mathcal{F})$, see Fig. 4. The proof of Theorem 4 is now complete.

Conclusion. We have performed a precise average case analysis of the Gauss reduction algorithm. The average complexity was found to equal a certain constant under the continuous model where the complex input z varies uniformly over the unit disk. (see Theorems 1 and 4). For the corresponding discrete model, the notable result (Theorem 3) is that the average cost over inputs of a fixed size is asymptotically constant, and thus is *asymptotically independent of input size.*

Therefore, there is a ratio from worst case to average case complexity which is of the order of $\log N$ when operating with inputs of the order of N: The algorithm behaves on average appreciably better than in the worst case. Our worst–case to average–case complexity ratio of $\log N$ is expected to "propagate" inside other algorithms based on the reduction of Gauss, most notably the LLL algorithm. This should entail a practical reduction of complexity by a factor of $\log N$ at least when numbers of the order of N are processed by these algorithms. Such phenomena have not been studied yet, we are not even aware of the existence of detailed empirical studies.

The number of iterations performed by the LLL algorithm in dimension equal to d with inputs of norm bounded by M is known to satisfy the worst case bound

$$L^{(d)} \leq \frac{d(d-1)}{2} \log_t M,$$

with $t > 1$ a control parameter. In view of our results, it is then tempting to conjecture that there is a constant $\beta^{(d)}$ such that the expected cost of the LLL reduction algorithm (measured by the number of iteration steps) is asymptotic to $\beta^{(d)}$. It is likely that an argument similar to the one used in the proof of Theorem 2 regarding exponential tail should constitute a major step. An appreciably harder problem would consist in establishing uniform bounds valid for all dimensions; this essentially amounts to quantifying the dependence of constant $\beta^{(d)}$ with respect to d.

Finally, the exponential tail result of Theorem 2 is of intrinsic interest. Apart from showing that costly runs are rather unlikely, it entails that the result relative to constant average case complexity remains valid under a fairly large class of probabilistic models, for instance all those with bounded density. Thus we expect our theoretical conclusions to be of some practical relevance as well because of this model independence property.

Acknowledgements. This work was supported in part by PRC Mathématiques et Informatique, and in part by the Basic Research Action of the EC under contract No 3075 (Project ALCOM).

References

[1] DIXON, J. D. The number of steps in the Euclidean algorithm. *Journal of Number Theory 2* (1970), 414–422.

[2] DUPRÉ, A. Sur le nombre de divisions à effectuer pour obtenir le plus grand commun diviseur entre deux nombres entiers. *Journal de Mathématiques 11* (1846), 41–74.

[3] FORD, L. R. *Automorphic Functions*, second ed. Chelsea Pub. Co., New York, 1957. Reprinted from first edition, 1929.

[4] HARDY, G. H., AND WRIGHT, E. M. *An Introduction to the Theory of Numbers*, fifth ed. Oxford University Press, 1979.

[5] HEILBRONN, H. On the average length of a class of continued fractions. In *Number Theory and Analysis* (1969), P. Turan, Ed., pp. 87–96.

[6] KANNAN, R. Algorithmic geometry of numbers. *Annual Reviews in Computer Science 2* (1987), 231–267.

[7] KNUTH, D. E. *The Art of Computer Programming*, 2nd ed., vol. 2: Seminumerical Algorithms. Addison-Wesley, 1981.

[8] LAMÉ, G. Note sur la limite du nombre de divisions dans la recherche du plus grand commun diviseur entre deux nombres entiers. *Comptes-Rendus de l'Académie des Sciences XIX* (1845), 867–870.

[9] 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.

[10] SCHOENEBERG, B. *Elliptic Modular Functions*, vol. 203 of *Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen.* Springer-Verlag, 1974.

[11] SERRE, J.-P. *A Course in Arithmetic.* Graduate Texts in Mathematics. Springer Verlag, 1973.

[12] VALLÉE, B. La réduction des réseaux. Autour de l'algorithme de Lenstra, Lenstra, Lovàsz. *Theoretical Informatics and Applications 23*, 3 (1989), 345–376.

[13] VALLÉE, B. Gauss' algorithm revisited. *Journal of Algorithms* (1991). To appear. Also available as Technical Report 1989–7 of Laboratoire A3L, University of Caen.

Imprimé en France
par
l'Institut National de Recherche en Informatique et en Automatique

ISSN 0249-6399