Multiple GCD's. Probabilistic analysis of the plain algorithm

Valérie Berthé, Jean Creusefond, Loïck Lhote, Brigitte Vallée

Multiple GCD's. Probabilistic analysis of the plain algorithm

Valérie Berthé
LIAFA (CNRS and Université Paris Diderot), France
berthe@liafa.univ-paris-diderot.fr

Jean Creusefond
INSA Rouen, France
jean.creusefond@insa-rouen.fr

Loïck Lhote
GREYC (CNRS, ENSICAEN, and Université of Caen), France
loick.lhote@ensicaen.fr

Brigitte Vallée
GREYC (CNRS, ENSICAEN, and Université of Caen), France
brigitte.vallee@unicaen.fr

ABSTRACT

The paper provides a probabilistic analysis of an algorithm which computes the gcd of ℓ inputs (with $\ell \geq 2$), with a succession of $\ell - 1$ phases, each of them being the Euclid algorithm on two entries. This algorithm is both basic and natural, and two kinds of inputs are studied –polynomials over the finite field $\mathbb{F}_q$ and integers–. The analysis exhibits the precise probabilistic behaviour of the main parameters, namely the number of iterations in each phase and the evolution of the length of the current gcd along the execution. We first provide an average-case analysis. Then we make it even more precise by a distributional analysis. Our results rigorously exhibit two phenomena: (i) there is a strong difference between the first phase, where most of the computations are done and the remaining phases; (ii), there is a strong similarity between the polynomial and integer cases, as can be expected.

Keywords

GCD algorithm, average-case analysis, analytic combinatorics, generating functions, dynamical analysis

1. INTRODUCTION

Computing gcd's is a main operation, perhaps the fifth main arithmetic operation. Indeed, in many symbolic computation systems, a large proportion of the time is devoted to computing gcd's on numbers or polynomials in order to keep fractions under an irreducible form (see [19, 20]). Here, we deal with the case when inputs are either integers or polynomials over a finite field $\mathbb{F}_q$. When there are only two inputs (polynomials or integers), various methods have been designed to compute gcd's. The Euclid Algorithm or its variants play a central role here, and Knuth writes in [14] that the Euclid algorithm can be called "the granddaddy of all algorithms, because it is the oldest nontrivial algorithm that has survived to the present day".

In this paper, we are concerned by the case when the gcd of ℓ inputs $x_1, \ldots, x_\ell$, with $\ell \geq 2$ has to be computed. Since it is easy to compute the gcd of two entries, the most straightforward algorithm which is also described in Knuth's book [14] is a sequence of $\ell - 1$ gcd computations on two entries: one lets $y_1 := x_1$, then, for $k \in [2..\ell]$, one successively computes $y_k := \gcd(x_k, y_{k-1}) = \gcd(x_1, x_2, \ldots, x_k)$. The "total" gcd is $y_\ell := \gcd(x_1, x_2, \ldots, x_\ell)$, and it is obtained after $\ell - 1$ phases. We call it the plain ℓ-Euclid algorithm.

This algorithm cannot be easily extended for computing small Bezout coefficients and we do not claim it is efficient. However, a first step in analysis of algorithms is to understand and precisely analyze even the simplest algorithms; Such an analysis then provides a basis of comparison for other algorithms of the same class.

To the best of our knowledge, the plain ℓ–Euclid algorithm has not been yet analyzed[1]. The situation contrasts with the case $\ell = 2$, where the classical Euclid algorithm and its main variants running on integers or on polynomials are now precisely analyzed. See [2, 10, 13] for analyses on polynomials, [11, 6] for the first analyses on integers and [12, 18, 1, 15] for more recent ones. As usual, the results are similar in the two cases. The length is the degree of the polynomial, or the logarithm of the integer, and the length of a pair is the maximum of its components. With respect to the length, the mean number of iterations is linear, the arithmetic complexity is quadratic. There are also precise results on the distribution of the number of iterations, which is asymptotically gaussian. Observe that the analysis is much more difficult in the integer case. In the polynomial case, classical tools of analytic combinatorics, described in [8], may be used, whereas the analysis in the integer case is based on a methodology which mixes analytic combinatorics and

[1]The analysis of the plain algorithm was proposed as an exercise in Knuth's book (second edition) [14], and quoted as a difficult one (HM48). However, for reasons we do not understand, the exercise disappears in the third edition....Too bad!

dynamical systems [18].

For the ℓ-Euclid algorithm, it is natural to choose, as the length of the input, the sum of the length of its components. Changing the input length yields a different probabilistic behavior, even when $\ell = 2$. We show the following: in the case $\ell = 2$, the mean number of iterations remains linear with respect to the (new) length but the distribution of number of iterations is now asymptotically uniform. In the case $\ell \geq 3$, our analysis exhibits a strong difference between the first phase and the following phases. In the first phase, the number of iterations has a linear mean and follows a beta law (the same as the law followed by the minimum of $\ell - 1$ reals i.i.d. in the unit interval). In the following phases, the number of iterations is constant on the average and follows a geometric law. These results can be expected, as Knuth wrote "In most cases, the length of the partial gcd decreases rapidly during the first few phases of the calculation, and this will make the remainder of the computation quite fast" [14]. Our analysis shows that, in most cases, "almost all the calculation" is done during the first phase. This will make possible to compare (in the extended version of this paper) the ℓ-Euclid algorithm to another strategy proposed in [3].

Plan of the paper. Sections 2, 3, 4 are devoted to the polynomial case, and Section 5 to the integer case. Section 2 describes the algorithm and states the main results, then, Sections 3 and 4 are devoted to the proofs, mainly based on analytical combinatorics: we use generating functions, built in Section 3, and used in Section 4. Section 5 explains the strong similarity between the two cases (polynomial and integer); it presents the useful generating functions (here of Dirichlet type), states the main results (only for the average-case) and briefly describes the main steps of the proof. We also explain why distributional results may be expected.

2. MAIN RESULTS FOR POLYNOMIALS.

We consider the ring $\mathbb{F}_q[X]$ of polynomials over the finite field $\mathbb{F}_q$ with q elements, and the degree of a non zero polynomial x is denoted by $\mathrm{d}(x)$.

The plain ℓ-Euclid algorithm on $\mathbb{F}_q[X]$ deals with a sequence of ℓ nonzero polynomials and computes their greatest common divisor y via a sequence of $\ell - 1$ gcd computations between two polynomials, as we previously explained.

Each phase is a 2–Euclid Algorithm which performs a gcd computation between two polynomials, with a sequence of Euclidean divisions, as recalled as follows:

The Euclid Algorithm on (a_1, a_2) with $\mathrm{d}(a_1) > \mathrm{d}(a_2)$

$$\begin{array}{lcccll} a_1 & = & m_1 a_2 & + & a_3 & 0 < \mathrm{d}(a_3) < \mathrm{d}(a_2) \\ a_2 & = & m_2 a_3 & + & a_4 & 0 < \mathrm{d}(a_4) < \mathrm{d}(a_3) \\ \dots & = & \dots & + & & \\ a_{p-1} & = & m_{p-1} a_p & + & a_{p+1} & 0 < \mathrm{d}(a_{p+1}) < \mathrm{d}(a_p) \\ a_p & = & m_p a_{p+1} & + & 0 & \end{array}$$

Then, $\gcd(a_1, a_2)$ is the last non-zero remainder a_{p+1}. It can be chosen monic. The number of steps (here equal to p) is one of the main parameters of interest.

This paper aims at precisely understanding the random behavior of the plain algorithm. Since the algorithm is a succession of phases, it is important to describe each phase of index k ($k \in [1..\ell - 1]$), with the following parameters:

(i) the number L_k of divisions performed during the k-th phase,

(ii) the degree D_k of the gcd y_k at the beginning of the k-th phase.

We are also interested in the analysis of global parameters.

(iii) The algorithm may be interrupted as soon as $D_k = 0$, and Π is the number of useful phases. Namely:

$\Pi = 0$ if $D_1 = 0$ and $\Pi := \max\{k | D_k > 0\}$ if $D_1 > 0$.

(iv) The total number of divisions of the interrupted algorithm is defined as

$L = 0$ if $\Pi = 0$ and $L = L_1 + L_2 + \ldots + L_\Pi$ if $\Pi > 0$.

The possible inputs are all the sequences $\underline{x}$ formed of ℓ polynomials, and we limit ourselves to monic polynomials, without loss of generality. Then the set of inputs is $\Omega = \mathcal{U}^\ell$, where $\mathcal{U}$ is the set of monic polynomials. Here, the size of the input $\underline{x}$ is the total degree of the sequence (and not the maximum, as it is often the case), and we let

$$\mathrm{d}(\underline{x}) = \mathrm{d}(x_1, x_2, \ldots, x_\ell) := \mathrm{d}(x_1) + \mathrm{d}(x_2) + \ldots + \mathrm{d}(x_\ell).$$

The subset Ω_n formed with the inputs of size n is a finite set, and it is endowed with the uniform probability. Now, the parameters of interest become random variables and we are interested in their probabilistic features.

With analytic combinatorics methodology, we prove the following two results. The first one deals with the expected values, whereas the second one describes asymptotic limit laws.

2.1 Average-case analysis.

Theorem 1 below exhibits a strong difference between the first phase and the following ones. It shows that, on average, the first phase performs a linear number of iterations which involves the entropy $2q/(q-1)$ of the Euclid dynamical system; then, even if the degree of the first gcd is linear with respect to the input size, the degree of the gcd is proven to be of constant order after the first phase on the average. Then, the number of divisions L_k which will be performed in the following phases, together with the degrees D_k of the following gcd's will be of constant order.

THEOREM 1. *[Expectations]. When the set Ω_n is endowed with the uniform distribution, the following holds:*

(a) The expectation of the number of iterations L_1 during the first phase is linear with respect to the size n and satisfies

$$\mathbb{E}_n[L_1] = \frac{q-1}{2q}\frac{n}{\ell} + \frac{3q+1}{4q} + O\left(\frac{1}{n}\right).$$

(b) For any $k \in [2..\ell - 1]$, the expectation of the number of iterations L_k during the k-th phase is asymptotic to a constant, and satisfies

$$\mathbb{E}_n[L_k] = \frac{q^k - 1}{q^k - q}\left[1 + O\left(\frac{1}{n}\right)\right].$$

(c) The expectation of the degree of the first polynomial x_1 is linear with respect to the size n and satisfies

$$\mathbb{E}_n[D_1] = \frac{n}{\ell}.$$

(d) For any $k \in [2..\ell - 1]$, the expectation of the degree D_k of the gcd y_k at the beginning of the k-th phase is asymptotic to a constant, and satisfies

$$\mathbb{E}_n[D_k] = \frac{1}{q^{k-1} - 1}\left[1 + O\left(\frac{1}{n}\right)\right].$$

2.2 Limit Laws.

The following result makes more precise the results obtained in Theorem 1, and explains more deeply the difference between the first phase ($k = 1$) and the following phases. For $k = 1$, the expected degrees of the first two polynomials x_1 and x_2 are linear, and the number of divisions is closely related to $\min(\mathrm{d}(x_1), \mathrm{d}(x_2))$. Then, it is natural to expect beta laws for the first phase, more precisely a beta law of parameter $(1, \ell - 1)$, since it is the law of the minimum Y of $\ell - 1$ random variables i.i.d. on the unit interval, which satisfies $\mathbb{P}[Y \geq x] = (1 - x)^{\ell-1}$. Such a law has a density equal to $(\ell - 1)(1 - x)^{\ell-2}$. For $\ell = 2$, this is the uniform law. For the following phases, geometric laws are expected, since the means of the gcd degrees are of constant order.

The next result describes the asymptotic distribution of L_k [Assertion (a)] and D_k [Assertion (b)]. At first glance, the results do not seem to heavily depend on the index k of the phase. However, it is not true, since the two ratios $p_k := (q-1)/(q^k-1)$ (in case L), $r_k := q^{1-k}$ (in case D), are equal to 1 for $k = 1$ and strictly less than 1 for $k \geq 2$.

THEOREM 2. *[Limit laws.] When the set Ω_n is endowed with the uniform distribution, the following holds:*

(a) The number of iterations L_1 during the first phase asymptotically follows a beta law of parameter $(1, \ell - 1)$ on the interval $[0, (q-1)/(2q)]$ whereas, the number of iterations L_k during each following phase asymptotically follows a geometric law with ratio $p_k := (q-1)/(q^k-1)$. One has:

$$\mathbb{P}_n\left[L_k > n/(k+1)\right] = 0, \quad \text{for any } k.$$

For any k, the distribution of L_k satisfies when $n \to \infty$,

$$\mathbb{P}_n[L_k > m] = \left(\frac{q-1}{q^k-1}\right)^m \left[1 + O\left(\frac{m}{n}\right)\right] \quad \text{for } m = o(n),$$

and for $m/n \to c$ with $c \in]0, \frac{1}{k+1}\frac{q^k-1}{q^k}[$

$$\mathbb{P}_n[L_k > m] = \left(\frac{q-1}{q^k-1}\right)^m \left(1 - \frac{(k+1)q^k}{q^k-1}c\right)^{\ell-1} \left[1 + O\left(\frac{1}{n}\right)\right].$$

(b) The degree D_1 of the first polynomial x_1 asymptotically follows a beta law of parameter $(1, \ell-1)$ on the interval $[0, 1]$, whereas the degree D_k of the gcd y_k at the beginning of each following phase asymptotically follows a geometric law with ratio $r_k := q^{1-k}$. One has : $\mathbb{P}_n\left[D_k > n/k\right] = 0$ for any k. For any k, the distribution of D_k satisfies when $n \to \infty$,

$$\mathbb{P}_n[D_k \geq m] = q^{(1-k)m}\left[1 + O\left(\frac{m}{n}\right)\right] \quad \text{for } m = o(n),$$

and for $m/n \to c$ with $c \in]0, 1/k[$

$$\mathbb{P}_n[D_k \geq m] = q^{(1-k)m}\left(1 - kc\right)^{\ell-1}\left[1 + O\left(\frac{1}{n}\right)\right].$$

2.3 Global parameters.

The interrupted algorithm stops as soon as the gcd y_k is of degree 0. Let $\Pi(x_1, \ldots, x_\ell)$ be the number of useful phases. Since the event $[\Pi \geq k]$ coincides with the event $[D_k \geq 1]$ for $k \in [1..\ell - 1]$, Theorem 2 leads to the following:

THEOREM 3. *When the set Ω_n is endowed with the uniform distribution, the distribution of the number Π of useful phases satisfies $\mathbb{P}_n[\Pi \geq 0] = 1$, $\mathbb{P}_n[\Pi \geq \ell] = 0$,*

$$\mathbb{P}_n[\Pi \geq k] = q^{1-k}\left[1 + O\left(\frac{1}{n}\right)\right] \quad \text{for } k \in [1..\ell - 1].$$

The following result studies the total number L of divisions performed by the interrupted version (see the proof in the appendix). With Theorems 1 and 3, the mean is easy to study. Moreover, Theorems 1 and 2 prove that the variance of L_1 is quadratic whereas the variance of L_k (for $k \geq 2$) is of constant order. Then, the linearity of the mean with the Markov inequality entail that L has the same asymptotic distribution as L_1.

THEOREM 4. *When the set Ω_n is endowed with the uniform distribution, the total number of divisions L performed by the interrupted version of the ℓ-Euclid algorithm has an expected value $\mathbb{E}_n[L]$ equal to*

$$\frac{q-1}{2q}\frac{n}{\ell} + \frac{3q+1}{4q} + \sum_{k=2}^{\ell-1}\left[\frac{q}{q^k} + \frac{q-1}{q^k-1}\right] + O\left(\frac{\ell^2}{n}\right).$$

Moreover, the number L asymptotically follows a beta distribution of parameter $(1, \ell-1)$ on the interval $[0, (q-1)/(2q)]$.

3. GENERATING FUNCTIONS

3.1 General setting

We use the analytic combinatorics methodology, and deal with its main tool, the generating functions. We use a variable z_i to mark the degree of the i–th polynomial x_i, and the generating function $F(z_1, z_2, \ldots z_\ell)$ of the set $\Omega = \mathcal{U}^\ell$, relative to the size d, is defined as

$$F(z_1, z_2, \ldots z_\ell) := \sum_{\underline{x} \in \mathcal{U}^\ell} z_1^{\mathrm{d}(x_1)}\, z_2^{\mathrm{d}(x_2)} \ldots z_\ell^{\mathrm{d}(x_\ell)}.$$

It is equal to the product $U(z_1)\,U(z_2)\ldots U(z_\ell)$, where $U(z)$ is the generating function of the set $\mathcal{U}$ of the monic polynomials relative to the size d, namely

$$U(z) = \sum_{x \in \mathcal{U}} z^{\mathrm{d}(x)} = \sum_{n \geq 0} q^n z^n = \frac{1}{1 - qz}.$$

Most of the time, we limit ourselves to the case when all the variables z_i are equal, and we write $F(z)$ instead of $F(z, \cdots, z)$. For studying a parameter (or a cost C) on $\Omega = \mathcal{U}^d$, a main tool is the bivariate generating function relative to some cost C, obtained by introducing a further variable u to mark the cost, and defined as

$$F(z, u) := \sum_{\underline{x} \in \mathcal{U}^d} z^{\mathrm{d}(\underline{x})} u^{C(\underline{x})}.$$

We are first interested in the mean value of parameter C, and we deal with the cumulative generating function

$$\widehat{F}(z) := \frac{\partial F}{\partial u}(z, u)|_{u=1}, \qquad \text{and} \quad \mathbb{E}_n[C] = \frac{[z^n]\widehat{F}(z)}{[z^n]F(z)}. \quad (1)$$

The probability distribution of the cost C can be studied with the generating function $F(z, u)$, via the relation

$$\mathbb{P}_n[C = i] = \frac{[z^n\, u^i]F(z, u)}{[z^n]F(z)}.$$

Then, the probabilities $\mathbb{P}_n[C \geq m]$ are expressed with the "cumulative bivariate generating functions" defined as

$$\widehat{F}^{[m]}(z) := \sum_{i \geq m}[u^i]F(z, u), \quad \text{and } \mathbb{P}_n[C \geq m] = \frac{[z^n]\widehat{F}^{[m]}(z)}{[z^n]F(z)}. \quad (2)$$

3.2 Another expression for the generating function

We first obtain an alternative expression for the generating function $F(z)$ with a product of $\ell-1$ factors, each of them describing a phase of the algorithm.

PROPOSITION 1. *The generating function of the set $\Omega = \mathcal{U}^\ell$ with the size equal to the total degree decomposes as*

$$F(z) = U(z)^\ell = U(z^\ell) \cdot \prod_{k=1}^{\ell-1} T(z, z^k)$$

and involves the phase-function T defined as

$$T(z,t) = \frac{U(z)+U(t)-1}{1-G(zt)}, \tag{3}$$

the generating function $U(z)$ of monic polynomials, and the generating function $G(z)$ of general polynomials with strictly positive degree, i.e.,

$$U(z) = \frac{1}{1-qz}, \qquad G(z) = \frac{(q-1)qz}{1-qz}.$$

PROOF. The Euclid algorithm first compares the degrees of x_1 and x_2. There are three cases:

$$\mathrm{d}(x_1) = \mathrm{d}(x_2), \qquad \mathrm{d}(x_1) > \mathrm{d}(x_2), \qquad \mathrm{d}(x_1) < \mathrm{d}(x_2).$$

In the first case, there is an extra subtraction, which can be viewed as a division with a quotient equal to 1.

In all the cases, the gcd $y := \gcd(x_1, x_2)$ together with the sequence of quotients $(m_1, m_2, \dots m_p)$ completely determines the input pair (x_1, x_2). More precisely, one writes $(x_1, x_2) = (y\,\widehat{x}_1, y\,\widehat{x}_2)$ with a coprime pair $(\widehat{x}_1, \widehat{x}_2)$ and the execution of the Euclid algorithm on the pair $(\widehat{x}_1, \widehat{x}_2)$ produces the same sequence $(m_1, m_2, \dots m_p)$ as the pair (x_1, x_2). The first quotient m_1 is monic (this is due to the fact that x_1 and x_2 are monic) and the remainder of the sequence $\Sigma = (m_2, \dots m_p)$ is formed with general polynomials m_i (no longer monic) with $\mathrm{d}(m_i) \geq 1$. As previously, the total degree of Σ is $\mathrm{d}(\Sigma) = \mathrm{d}(m_2) + \ldots + \mathrm{d}(m_p)$.

We now focus on the first quotient m_1, and we consider the three following possible cases:

(a) If $\mathrm{d}(x_1) = \mathrm{d}(x_2)$, then $m_1 = 1$.

(b) If $\mathrm{d}(x_1) > \mathrm{d}(x_2)$ then
$\mathrm{d}(m_1) \geq 1, \quad \mathrm{d}(\widehat{x}_2) = \mathrm{d}(\Sigma), \quad \mathrm{d}(\widehat{x}_1) = \mathrm{d}(m_1) + \mathrm{d}(\Sigma)$.

(c) If $\mathrm{d}(x_1) < \mathrm{d}(x_2)$ then
$\mathrm{d}(m_1) \geq 1, \quad \mathrm{d}(\widehat{x}_1) = \mathrm{d}(\Sigma), \quad \mathrm{d}(\widehat{x}_2) = \mathrm{d}(m_1) + \mathrm{d}(\Sigma)$.

All these remarks provide an alternative expression of the product $U(z_1)\,U(z_2)$. Indeed, we use the two relations

$$z_1^{\mathrm{d}(x_1)} z_2^{\mathrm{d}(x_2)} = (z_1 z_2)^{\mathrm{d}(y)} \cdot z_1^{\mathrm{d}(\widehat{x}_1)} z_2^{\mathrm{d}(\widehat{x}_2)},$$

$$\sum_{\widehat{x}_1, \widehat{x}_2} z_1^{\mathrm{d}(\widehat{x}_1)} z_2^{\mathrm{d}(\widehat{x}_2)} = \left[1 + \sum_{m_1} z_1^{\mathrm{d}(m_1)} + z_2^{\mathrm{d}(m_1)}\right] \left[\sum_{\Sigma} (z_1 z_2)^{\mathrm{d}(\Sigma)}\right],$$

together with the conditions previously described on the first quotient m_1, the sequence Σ and the gcd y. The first factor gives rise to the generating function

$$1 + (U(z_1) - 1) + (U(z_2) - 1) = U(z_1) + U(z_2) - 1$$

which involves the generating function $U(z)$ of monic polynomials, whereas the second factor is expressed with the generating function $G(z)$ of general polynomials with a strictly positive degree, under the form $1/(1 - G(z_1 z_2))$. We have then proven the following alternative form for the product

$$U(z_1)\,U(z_2) = U(z_1 z_2) \cdot T(z_1, z_2).$$

When we replace this expression into the total product

$$U(z_1)\,U(z_2) \dots U(z_\ell) = F(z_1, z_2, \dots, z_\ell)$$

and iterate the transformation, we obtain an alternative expression for the generating function $F(z_1, z_2, \dots, z_\ell)$ with a product of $\ell - 1$ factors, each of them involving the phase-function T at points z_k and $t_k = z_1 \dots z_k$

$$F(z_1, z_2, \dots z_\ell) = U(t_\ell) \cdot \prod_{k=1}^{\ell-1} T(z_{k+1}, t_k).$$

It can be useful in some studies to keep all the variables z_i, but here, we let $z_1 = z_2 = \ldots = z_\ell$, and we obtain an expression of the generating function $F(z)$. □

3.3 Generating functions for parameters of interest

When studying the parameter L_k (number of steps in the k-th phase), we use an extra variable u which marks each step of the k-th iteration, and we deal with the generating function

$$T(z,t,u) = u\frac{U(z)+U(t)-1}{1-uG(zt)} \qquad \text{with } t = z^k,$$

which replaces $T(z, z^k)$ inside $F(z)$.

When studying the parameter D_k (degree of the gcd at the beginning of the k-th phase), we use an extra variable u which marks the degree of the gcd y_k, and we deal with the generating function

$$U(t,u) = \frac{1}{1-qut} \qquad \text{with } t = z^k,$$

which replaces $U(z^k)$ inside $F(z)$. Finally:

PROPOSITION 2. *For any $k \in [1..\ell-1]$, the bivariate generating function $L_k(z,u)$ relative to the number of divisions during the k-th phase satisfies*

$$L_k(z,u) = U(z)^\ell \cdot \frac{T(z, z^k, u)}{T(z, z^k)}.$$

For any $k \in [1..\ell-1]$, the bivariate generating function $D_k(z,u)$ relative to the degree of the k-th gcd y_k at the beginning of the k-th phase satisfies

$$D_k(z,u) = U(z)^\ell \cdot \frac{U(z^k, u)}{U(z^k)}.$$

With Proposition 1, the functions $T(z, z^k, u)$ and $U(z^k, u)$ admit precise expressions, and Proposition 2 leads to the expression of the bivariate generating functions $L_k(z,u)$ and $D_k(z,u)$

$$\frac{L_k(z,u)}{U(z)^\ell} = u\,\frac{1-G(z^{k+1})}{1-uG(z^{k+1})}, \qquad \frac{D_k(z,u)}{U(z)^\ell} = \frac{1-qz^k}{1-quz^k}. \tag{4}$$

Finally, taking the derivative with respect to u, we obtain the cumulative generating functions

$$\frac{\widehat{L}_k(z)}{U(z)^\ell} = \frac{1-qz^{k+1}}{1-q^2z^{k+1}}, \qquad \frac{\widehat{D}_k(z)}{U(z)^\ell} = \frac{qz^k}{1-qz^k}. \tag{5}$$

Extracting in (4) the coefficient of $[u^i]$ in the bivariate generating functions and taking the sum over $i \geq m$ gives

$$\frac{\widehat{L}_k^{[m]}(z)}{U(z)^\ell} = G(z^{k+1})^{m-1}, \qquad \frac{\widehat{D}_k^{[m]}(z)}{U(z)^\ell} = (qz^k)^m. \tag{6}$$

4. FINAL STEPS FOR THE PROOFS.

We have obtained in (5) and (6) the expressions of useful generating functions, that are always here fractional functions. It is then possible to use (1) and (2) to obtain an exact expression of the expectation and probability distribution of the parameters D_k and L_k. However, we are mainly interested in the asymptotic behaviour (as $n \to \infty$) of these probabilistic features. Singularity analysis describes the possible transfer between the behavior of a generating function, viewed as an analytic function, near its dominant singularity (the singularity closest to 0) and the asymptotic behavior of its coefficients. More precisely, the position and the nature of the dominant singularity play a fundamental role.

4.1 Average-case analysis.

Here, the main tools are the cumulative generating functions (5). They both admit a dominant pole in $z = 1/q$ but the order of the pole is different according to the phase. For the first phase ($k = 1$), the pole $z = 1/q$ is of order $\ell + 1$ whereas for the other phases ($k \geq 2$), this pole remains of order ℓ.

We now use the following result, here quite trivial since we deal with rational fractions.

LEMMA 1. *Consider a function $f(z) = g(z)/(1-qz)^j$ with $j \geq 2$ which involves a function $g(z)$ which is analytic in the disk $|z| > 1/q$ and satisfies $g(1/q) \neq 0$. Then,*

$$[z^n]f(z) = g\left(\frac{1}{q}\right)\binom{n+j-1}{j-1}q^n - g'\left(\frac{1}{q}\right)\binom{n+j-2}{j-2}q^{n-1},$$

with a remainder term in $O\left(n^{j-3}q^n\right)$. The previous estimate also holds if g admits simple isolated poles on the punctured circle $\{|z| = 1/q, z \neq 1/q\}$, as soon as $j \geq 3$.

We first consider the case when the phase index k is at least equal to 2. In this case, with the expressions of the cumulative generating functions $\widehat{D}_k(z)$ and $\widehat{L}_k(z)$ given in (5), the lemma applies with $j = \ell$ and

$$g(z) = \frac{1-qz^{k+1}}{1-q^2z^{k+1}}, \quad g\left(\frac{1}{q}\right) = \frac{q^k-1}{q^k-q} \quad \text{(case } L\text{)},$$

$$g(z) = \frac{qz^k}{1-qz^k} \quad g\left(\frac{1}{q}\right) = \frac{1}{q^{k-1}-1} \quad \text{(case } D\text{)}.$$

Consider now the case when the phase index k equals 1. In this case, the integer j equals $\ell + 1$, and

$$g(z) = qz \quad \text{(case } D\text{)}, \qquad g(z) = \frac{1-qz^2}{1+qz} \quad \text{(case } L\text{)}.$$

In both cases L and D, the lemma applies. In case D, we obtain:

$$[z^n]\widehat{D}_1(z) = q^n\binom{n+\ell-1}{\ell}.$$

We remark that, in case L, the function g admits $z = -1/q$ as a simple pole and we obtain $[z^n]\widehat{L}_1(z) =$

$$= \frac{q-1}{2q}q^n\binom{n+\ell}{\ell} + \frac{q+3}{4}q^{n-1}\binom{n+\ell-1}{\ell-1} + O\left(n^{\ell-2}q^n\right).$$

In all the cases, the normalization by $\text{card}(\Omega_n) = \binom{n+\ell-1}{\ell-1}q^n$ proves Theorem 1.

4.2 A general framework for limit laws.

The "cumulative bivariate generating functions" which are useful for the study of the distributions at the k-th phase are, due to (6), of the form

$$U(z)^\ell \cdot A_k(z)^m,$$

where the function $A_k(z)$ depends on the index of the phase, and the type of parameter. One has

$$A_k(z) = qz^k \quad \text{(type } D\text{)}, \tag{7}$$

$$A_k(z) = \frac{(q-1)qz^{k+1}}{1-qz^{k+1}} = G(z^{k+1}) \quad \text{(type } L\text{)}. \tag{8}$$

In all the cases, the generating functions are expressed as a product of the function $U(z)^\ell$ which has a pole of order ℓ at $z = 1/q$, with a "large power" of a function $A(z)$. The term "large power" is used since the exponent m may depend on the size n. As the following proposition shows it, there are two main cases, according as the value of A at $z = 1/q$ is equal to 1 or not. The first case happens for $k = 1$ and leads to beta distribution, whereas the second case happens for $k \geq 2$ and leads to geometric distributions.

The following result is proven in the appendix.

PROPOSITION 3. *Consider the function*

$$F^{[m]}(z) = \frac{1}{(1-z)^\ell} \cdot A(z)^m,$$

where $A(z)$ is analytic on the disk $|z| \leq \rho$ with $\rho > 1$ and satisfies $a := A(1) \neq 0$, $b := A'(1) > 0$ and for $|z|$ close enough to 1, $|A(z)| \leq A(|z|)$.
Then, the coefficient of z^n in $F^{[m]}(z)$ satisfies the following:

(*a*) *When $m/n \to 0$, then*

$$[z^n]F^{[m]}(z) = \frac{n^{\ell-1}}{(\ell-1)!} \cdot a^m\left[1 + O\left(\frac{m}{n}\right)\right].$$

(*b*) *When $m/n \to c$ for some $c \in]0, a/b[$ then*

$$[z^n]F^{[m]}(z) = \frac{n^{\ell-1}}{(\ell-1)!}\,a^m\left(1-\frac{b}{a}c\right)^{\ell-1}\left[1 + O\left(\frac{1}{n}\right)\right].$$

4.3 End of the proof of Theorem 2.

The probabilities $\mathbb{P}_n[L_k > m]$, $\mathbb{P}_n[D_k > m]$ are related to the coefficient of z^n in the generating functions described in (6), and the functions $A_k(z)$ of interest are given in (7) and (8). Since, in case D, $A_k(z)$ is multiple of z^k, and, in case L, $A_k(z)$ is multiple of z^{k+1}, we first deduce the equalities

$$\mathbb{P}_n\left[D_k > n/k\right] = 0, \qquad \mathbb{P}_n\left[L_k > n/(k+1)\right] = 0.$$

After a change of variable $z \to z/q$, the hypotheses of Proposition 3 are fulfilled for $z \mapsto A_k(z/q)$, and one has,

$$a_k = q^{1-k}, \quad \frac{b_k}{a_k} = k \qquad \text{(case } D\text{)},$$

$$a_k = \frac{q-1}{q^k-1}, \quad \frac{b_k}{a_k} = (k+1)\frac{q^k}{q^k-1}, \qquad \text{(case } L\text{)}.$$

For $k = 1$, the constants a_k are equal to 1, whereas they are strictly less than 1 for $k \geq 2$. Finally, the normalization by $\text{card}(\Omega_n)$,

$$\text{card}(\Omega_n) = \frac{n^{\ell-1}}{(\ell-1)!}q^n\left[1 + O\left(\frac{1}{n}\right)\right],$$

proves Assertions (*a*) and (*b*) of Theorem 2.

5. THE INTEGER CASE.

We now briefly explain how a similar study can be performed in the case of the Euclid algorithm on integers. As usual (see for instance [15, 18]), the polynomial study shows the road, and similar results are expected in the integer study, even if they are often more difficult to obtain and less precise.

5.1 The plain algorithm on integers.

In the integer case, the ℓ–plain Euclid algorithm has exactly the same structure as in the polynomial case. It is composed with $\ell - 1$ phases, each of them being the Euclid algorithm which performs the gcd computation between two integers. Its execution is described as in Figure of Section 2, and the degree is just replaced by the integers themselves. The main parameters of interest are the same, and D_k is now the length of the gcd y_k, namely its logarithm.

The possible inputs are all the sequences $\underline{x}$ formed of ℓ integers, and we limit ourselves to positive integers, without loss of generality. Then the set of inputs is $\Omega = \mathbb{N}^\ell$, where $\mathbb{N}$ is the set of positive integers. Here, the size of the input $\underline{x}$ equals the product of its components $x_1 \cdot x_2 \cdot \ldots \cdot x_\ell$, and the length of the input is the sum of the lengths of its components.

Here again, the subset Ω_N formed with the inputs of size at most N is a finite set, and it is endowed with the uniform probability.

5.2 Dirichlet generating functions.

The generating functions are now of Dirichlet type, and the basic one is the generating function of the set $\mathbb{N}^\ell$. We deal with ℓ–uples $\underline{x}$ of integers $\underline{x} = (x_1, x_2, \ldots x_\ell)$ and consider the generating function

$$F(s_1, s_2, \ldots s_\ell) = \sum_{\underline{x} \in \mathbb{N}^\ell} \frac{1}{x_1^{s_1}} \frac{1}{x_2^{s_2}} \cdots \frac{1}{x_\ell^{s_\ell}} = \zeta(s_1) \ldots \zeta(s_\ell),$$

where the ζ function is the generating function of $\mathbb{N}$,

$$\zeta(s) := \sum_{n \geq 1} \frac{1}{n^s}.$$

The main case of interest $s_1 = s_2 = \ldots s_\ell = s$ gives rise to

$$F(s) := F(s, \ldots, s) = \zeta(s)^\ell.$$

However, we begin with the generic case when the ℓ–uple $s = (s_1, \ldots, s_\ell)$ is general.

The first step derives, as previously, a decomposition for the product $\zeta(s_1)\,\zeta(s_2)$, of the form

$$\zeta(s_1)\,\zeta(s_2) = \zeta(s_1 + s_2) \cdot T(s_1, s_2),$$

where $T(s_1, s_2)$ is the phase generating function which describes the Euclid algorithm on two integers. In the integer case, the 2-Euclid algorithm is described with the transfer operator, related to the underlying dynamical system, and introduced by Ruelle in [16] in a general setting. This operator deals here with the Gauss map S defined by $S(x) := 1/x - [1/x]$, and the set $\mathcal{G}$ of its inverse branches, namely

$$\mathcal{G} := \left\{ h_m(t) : t \mapsto \frac{1}{m+t};\quad m \geq 1 \right\}.$$

It depends on a complex parameter s, and acts for $\Re s > 1$ on functions defined on the unit interval as

$$\mathbf{G}_s[f](t) = \sum_{h \in \mathcal{G}} |h'(t)|^{s/2} f \circ h(t). \tag{9}$$

The following result provides an analog of Proposition 1.

PROPOSITION 4. *The generating function of $\Omega = \mathbb{N}^\ell$ with the size "product of inputs" can be written as*

$$F(s) = \zeta(s)^\ell = \zeta(\ell s) \cdot \prod_{k=1}^{\ell-1} T(s, ks),$$

and involves the phase-function T defined as

$$T(s,t) = \frac{1}{2}\left[(I - \mathbf{G}_{s+t})^{-1} \circ (\mathbf{G}_s + \mathbf{G}_t)[1](0) \right], \tag{10}$$

where $\mathbf{G}_s$ is the transfer operator relative to the Euclid dynamical system, defined in (9).

PROOF. The Euclid algorithm first compares the two integers x_1 and x_2. There are three cases:

$$x_1 = x_2, \qquad x_1 > x_2, \qquad x_1 < x_2.$$

In all the cases, the gcd $y := \gcd(x_1, x_2)$ together with the sequence of quotients $(m_1, m_2, \ldots m_p)$ completely determines the input pair (x_1, x_2). More precisely, one writes $(x_1, x_2) = (y\,\widehat{x}_1, y\,\widehat{x}_2)$ with a coprime pair $(\widehat{x}_1, \widehat{x}_2)$ and the execution of the Euclid algorithm on the pair $(\widehat{x}_1, \widehat{x}_2)$ produces the same sequence $(m_1, m_2, \ldots m_p)$ as the pair (x_1, x_2), with now remainders $\widehat{x}_i$ which satisfy $x_i = y\widehat{x}_i$.

The execution of the Euclid algorithm on the pair $(\widehat{x}_1, \widehat{x}_2)$ with $\widehat{x}_1 > \widehat{x}_2$ builds continued fraction expansions

$$\frac{\widehat{x}_2}{\widehat{x}_1} = h \circ g(0), \qquad \frac{\widehat{x}_3}{\widehat{x}_2} = g(0).$$

Here, $h := h_{m_1}$ is related to the first quotient and $g = h_{m_2} \circ h_{m_3} \circ \ldots \circ h_{m_p}$ is related to the sequence $(m_2, m_3, \ldots, m_p)$. Since the two pairs $(\widehat{x}_1, \widehat{x}_2)$ and $(\widehat{x}_2, \widehat{x}_3)$ are coprime, the denominators of each rational are expressed with derivatives,

$$\frac{1}{\widehat{x}_1^2} = |(h \circ g)'(0)| = |h'(g(0))| \cdot |g'(0)|, \qquad \frac{1}{\widehat{x}_2^2} = |g'(0)|.$$

Then, in the case when $\widehat{x}_1 \geq \widehat{x}_2$, the sum

$$\sum_{\widehat{x}_1 \geq \widehat{x}_2} \frac{1}{\widehat{x}_1^{s_1}} \frac{1}{\widehat{x}_2^{s_2}} = 1 + \sum_{h,g} |h'(g(0))|^{s_1/2} \cdot |g'(0)|^{(s_1+s_2)/2},$$

can be expressed with the transfer operator $\mathbf{G}_s$ as[2]

$$\frac{1}{2}\left[(I - \mathbf{G}_{s_1+s_2})^{-1} \circ \mathbf{G}_{s_1}[1](0) + 1 \right].$$

The case $\widehat{x}_2 \geq \widehat{x}_1$ can be dealt with exchanging the roles of $\widehat{x}_1$ and $\widehat{x}_2$. Finally, the relations

$$\sum_{x_1, x_2} \frac{1}{x_1^{s_1}} \frac{1}{x_2^{s_2}} = \sum_v \frac{1}{y^{s_1+s_2}} \sum_{\widehat{x}_1, \widehat{x}_2} \frac{1}{\widehat{x}_1^{s_1}} \frac{1}{\widehat{x}_2^{s_2}}$$

entail the equality $\zeta(s_1)\zeta(s_2) = \zeta(s_1 + s_2) \cdot T(s_1, s_2)$, where T is defined in (10). When we replace this expression into the total product

$$\zeta(s_1)\,\zeta(s_2) \ldots \zeta(s_\ell) = F(s_1, s_2, \ldots, s_\ell),$$

and iterate the transformation, we obtain an alternative expression for the generating function $F(s_1, s_2, \ldots, s_\ell)$ with a product of $\ell - 1$ factors, each of them involving the phase-function T at points s_k and $t_k = s_1 + \ldots + s_k$,

$$F(s_1, s_2, \ldots s_\ell) = \zeta(t_\ell) \cdot \prod_{k=1}^{\ell-1} T(s_{k+1}, t_k).$$

[2]The factor $(1/2)$ is here to take into account the fact that any rational of $]0, 1]$ admits two continued fraction expansions: the proper one and the improper one.

It may be useful in some studies to keep all the variables s_i, but, here again, we let $s_1 = s_2 = \ldots = s_\ell = s$, and we obtain the expression of the generating function $F(s)$. □

5.3 Dirichlet generating functions for parameters of interest

When studying the parameter L_k (number of steps in the k-th phase), we use an extra variable u which marks each step of the k-th iteration, and we deal with the generating function

$$2T(s,t,u) = u(1-u\mathbf{G}_{s+t})^{-1} \circ (\mathbf{G}_s + \mathbf{G}_t)[1](0),$$

with $t = ks$, which replaces $2T(s,ks)$ inside $F(s)$.
When studying the parameter D_k (length of the gcd at the beginning of the k-th phase), we use an extra variable u which marks the length of the gcd y_k, and we deal with the generating function

$$\zeta(t,u) = \sum_{n\geq 1} \frac{u^{\log n}}{n^t} = \zeta(t - \log u) \qquad \text{with } t = ks,$$

which replaces $\zeta(ks)$ inside $F(s)$. Finally, we obtain an analog of Proposition 2.

PROPOSITION 5. *For any $k \in [1..\ell-1]$, the bivariate generating function $L_k(s,u)$ relative to the number of divisions during the k-th phase satisfies*

$$L_k(s,u) = \zeta(s)^\ell \cdot \frac{T(s,ks,u)}{T(s,ks)}.$$

For any $k \in [1..\ell-1]$, the bivariate generating function $D_k(z,u)$ relative to the size of the k-th gcd y_k at the beginning of the k-th phase satisfies

$$D_k(s,u) = \zeta(s)^\ell \cdot \frac{\zeta(ks - \log u)}{\zeta(ks)}.$$

With Proposition 4, the function $T(s,ks,u)$ admits a precise expression. Using Proposition 5, and taking the derivative with respect to u, we obtain the cumulative generating functions

$$\frac{\widehat{D}_k(s)}{\zeta(s)^\ell} = \frac{\zeta'(ks)}{\zeta(ks)}, \qquad \frac{\widehat{L}_k(s)}{\zeta(s)^\ell} = \frac{\widehat{T}(s,ks)}{T(s,ks)}. \tag{11}$$

Remark that the Dirichlet series

$$\widehat{T}(s,ks) := \frac{\partial T}{\partial u}(s,ks,u)|_{u=1}$$

involves two occurrences of the quasi-inverse $(I-\mathbf{G}_{(k+1)s})^{-1}$.

5.4 Average-case analysis results.

The following result is an exact analog of Theorem 1. In particular, in Assertion (a), the entropy $\pi^2/(6\log 2)$ of the integer Euclidean system replaces its polynomial analog $(2q)/(q-1)$ on $\mathbb{F}_q[X]$.

THEOREM 5. *[Expectations]. When the set Ω_N is endowed with the uniform distribution, the following holds:*
(a) The expectation of the number of iterations L_1 during the first phase is linear with respect to the length $\log N$ and satisfies

$$\mathbb{E}_N[L_1] = \frac{6\log 2}{\pi^2}\left(\frac{1}{\ell}\log N\right)\left[1 + O\left(\frac{1}{\log N}\right)\right].$$

(b) For any $k \in [2..\ell-1]$, the expectation of the number of iterations L_k during the k-th phase is asymptotic to a constant a_k which is expressed with the operator $\mathbf{G}_s$ at $s = k+1$,

$$\mathbb{E}_N[L_k] = a_k\left[1 + O\left(\frac{1}{\log N}\right)\right].$$

(c) The expectation of the length of the first integer u_1 is linear with respect to the length $\log N$ and satisfies

$$\mathbb{E}_N[D_1] = \frac{1}{\ell}\log N.$$

(d) For any $k \in [2..\ell-1]$, the expectation of the length of the gcd y_k at the beginning of the k-th phase is asymptotic to a constant, and satisfies

$$\mathbb{E}_N[D_k] = \frac{\zeta'(k)}{\zeta(k)}\left[1 + O\left(\frac{1}{\log N}\right)\right].$$

5.5 Main principles for the analysis.

We have obtained in (11) the expressions of the cumulative generating functions. It is now possible to "extract" coefficients of these Dirichlet series, with an analog of (1).

However, for Dirichlet generating functions, it is (very often) only possible to study the sum of the coefficients for $n \leq N$, and this is why we deal with the set Ω_N of inputs with size at most N. Then the mean value of cost C on Ω_N is obtained from the cumulative generating function as

$$\mathbb{E}_N[C] = \frac{\sum_{n\leq N}[n^{-s}]\widehat{F}(s)}{\sum_{n\leq N}[n^{-s}]F(s)}.$$

As previously, singularity analysis performs a transfer between the behavior of a Dirichlet generating function, viewed as an analytic function, near its dominant singularity (here, the singularity with the largest real part) and the asymptotic behavior of its coefficients. The position and the nature of the dominant singularity play a fundamental role.

However, this transfer is more difficult for Dirichlet series. As previously, the basic tool is the Cauchy formula, but, here, the circles centered at 0 are replaced by vertical lines, which are not compact. For this short version, we use here the following Tauberian theorem, due to Delange [5, 17], which deals with Dirichlet series with positive coefficients, but does not provide any remainder term. This is why we only prove here, in this short version, a weak version of Theorem 5, without remainder terms.

THEOREM 6. *Let $F(s) = \sum_{n\geq 1} a_n n^{-s}$ be a Dirichlet series with non negative coefficients such that $F(s)$ converges for $\Re(s) > \sigma > 0$. Assume the following:*
(i) $F(s)$ is analytic on $\Re(s) = \sigma, s \neq \sigma$,
(ii) for some $\gamma \geq 0$, one has

$$F(s) = A(s)(s-\sigma)^{-\gamma-1} + C(s),$$

where A, C are analytic at σ, with $A(\sigma) \neq 0$.
Then, as $N \to \infty$,

$$\sum_{n\leq N} a_n = \frac{A(\sigma)}{\sigma\Gamma(\gamma+1)}\, N^\sigma \log^\gamma N \quad [1+\epsilon(N)], \quad \epsilon(N) \to 0.$$

5.6 Sketch of proof for Theorem 5.

We apply Theorem 6 with $\sigma = 1$ to the cumulative functions defined in (11). We then study the series $\zeta(s)^\ell, \zeta'(s)$, $\widehat{T}(s,ks), T(s,ks)$. All is known about the first two, and this entails a easy application of Theorem 6 in case D.

We focus now on case L, where we use precise results on the operator $\mathbf{G}_s$, when it acts on a convenient Banach space $\mathcal{B}$ (not described here). First, the function $\widehat{\zeta}_s = \mathbf{G}_s[1](0)$ has a pole of order 1 at $s = 1$, and near $s = 1$, one has $\widehat{\zeta}_s \sim C/(s-1)$, where C belongs to $\mathcal{B}$. Then, near $s = 1$,

$$2^{\delta(k,1)}\, T(s, ks) \sim 1/(s-1)(I - \mathbf{G}_{(k+1)s})^{-1}[C](0),$$

for any k, ($\delta(i,j)$ is the Kronecker symbol).

Now, the following is known: for $\Re t > 2$, the operator $\mathbf{G}_t$ has a spectral radius strictly less than 1, and the quasi-inverse $(I - \mathbf{G}_t)^{-1}$ is analytic there. On the line $\Re t = 2$, the quasi-inverse $(I - \mathbf{G}_t)^{-1}$ is analytic except at $t = 2$ where it admits a simple pole with a residue which involves the entropy. As $T(s, ks)$ (resp. $\widehat{T}(s, ks)$) contains one (resp. two) occurrence(s) of the quasi-inverse, the following holds:
– for $k \geq 2$, the series $T(s, ks)$ and $\widehat{T}(s, ks)$ have a simple pole at $s = 1$,
– for $k = 1$, the series $T(s, s)$ has a pole of order 2 at $s = 1$, and the series $\widehat{T}(s, s)$ has a pole of order 3 at $s = 1$.
Finally, the function $\widehat{L}_k(s)$ satisfies the hypotheses of Theorem 6, with $\sigma = 1$; the exponent γ equals ℓ for $k \geq 2$ and $\ell + 1$ for $k = 1$. This concludes the "proof" in case L.

5.7 And now?

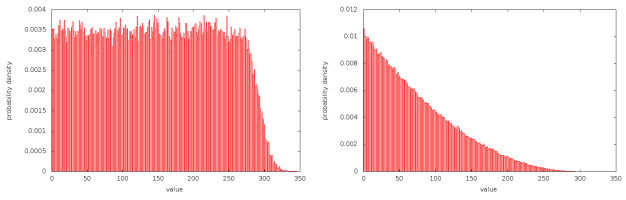
We can also study the number of useful phases, and recover the classical result $\mathbb{P}_N[D_k = 0] \sim 1/\zeta(k)$. If we wish to obtain remainder terms in Theorem 5, and an analog of distributional results obtained in Theorem 2, we need the analog of Proposition 3. For this purpose, we have to deal with the Perron Formula, as for previous distributional analyses performed in integer case.

The Perron Formula provides remainder terms as soon as the Dirichlet series of interest possess a "tameness" region on the left of the vertical line $\Re s = 1$, where $s = 1$ is their only pole and they are of polynomial growth for $|\Im s| \to \infty$. Classical results [17] entail such tameness properties for the zeta function and its derivatives, and results due to Dolgopyat-Baladi-Vallée [7, 1] prove that they also hold for the transfer operator $\mathbf{G}_{2s}$. Then, it would be possible to obtain the analog of Theorem 2 where n, m are replaced by $\log N, \log M$ and the ratios p_k, r_k which respectively appear in Assertions (a) and Assertions (b) of Theorem 2 are replaced by

$$\widehat{p}_k := \lambda\left((k+1)/2\right) \ \text{(case } L\text{)}, \quad \widehat{r}_k := \exp(1-k) \ \text{(case } D\text{)}.$$

Here, $\lambda(s)$ is the dominant eigenvalue of the operator $\mathbf{G}_{2s}$ (when acting on the functional space $\mathcal{B}$), which plays a central role in many analyses of Euclidean type. It satisfies $\lambda(s) = 1$, and, for $k = 3$, we recover the constant $\lambda(2)$ which plays a central role in Euclidean dynamics, notably in the analysis of the Gauss lattice reduction algorithm [4, 9].

5.8 Some experiments



The figure shows the experimental densities of L_1 for $\ell = 2$ (left) and $\ell = 4$ (right) and clearly exhibits the uniform limit law (left) and the beta limit law (right). These experiments are obtained with $5 \cdot 10^5$ executions on random integer inputs $(x_1, \ldots, x_\ell)$ whose (total) length satisfies $\log_{10} N = 10^4$.

Acknowledgements. Thanks to the two ANR Projects:
– ANR BOOLE (ANR 2009 BLAN 0011)
– ANR MAGNUM (ANR 2010 BLAN 0204)

6. REFERENCES

[1] V. Baladi and B. Vallée. Euclidean algorithms are gaussian. *Journal of Number Theory, Volume 110, Issue*, 110:331–386, 2006.

[2] V. Berthé and H. Nakada. On continued fraction expansions in positive characteristic: Equivalence relations and some metric properties. *Expositiones Mathematicae*, 18:257–284, 2000.

[3] G. Cooperman, S. Feisel, J. von zur Gathen, and G. Havas. Gcd of many integers, Cocoon'99, LNCS.

[4] H. Daudé, P. Flajolet, and B. Vallée. An average-case analysis of the gaussian algorithm for lattice reduction. *Combinatorics, Probability & Computing*, 6(4):397–433, 1997.

[5] H. Delange. Généralisation du théorème d'ikehara. *Ann. Sc. ENS*, 71:213–422, 1954.

[6] J. D. Dixon. The number of steps in the Euclidean algorithm. *Journal of Number Theory*, 2:414–422, 1970.

[7] D. Dolgopyat. On decay of correlations in Anosov flows. *Ann. of Math.*, 147(2):357–390, 1998.

[8] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009.

[9] P. Flajolet and B. Vallée. Continued fraction algorithms, functional operators, and structure constants. *Theor. Comput. Sci.*, 194(1-2):1–34, 1998.

[10] C. Friesen and D. Hensley. The statistics of continued fractions for polynomials over a finite field. *Proc. Amer. Math. Soc*, 124:2661–2673, 1996.

[11] H. Heilbronn. On the average length of a class of continued fractions. In P. Turan, editor, *Number Theory and Analysis*, pages 87–96, 1969.

[12] D. Hensley. The number of steps in the Euclidean algorithm. *Journal of Number Theory*, 2(49):149–182, 1994.

[13] A. Knopfmacher and J. Knopfmacher. The exact length of the euclidean algorithm in $F_q[X]$. *Mathematika*, 35:297–304.

[14] D. E. Knuth. *Seminumerical Algorithms*, volume 2 of *The Art of Computer Programming*. Addison-Wesley, 3rd edition, 1998.

[15] L. Lhote and B. Vallée. Gaussian laws for the main parameters of the Euclid algorithms. *Algorithmica*, 50(4):497–554, 2008.

[16] D. Ruelle. *Thermodynamic Formalism*. Cambridge University Press, 2004.

[17] G. Tenenbaum. *Introduction à la théorie analytique des nombres*, volume 13. Institut Élie Cartan, Nancy, France, 1990.

[18] B. Vallée. Euclidean dynamics. *Discrete and Continuous Dynamical Systems*, 1(15):281–352, 2006.

[19] J. von zur Gathen and J. Gerhard. *Modern Computer Algebra*. Cambridge University Press, 2003.

[20] C. Yap. *Fundamental Problems of Algorithmic Algebra*. Oxford University Press, 2000.

7. APPENDIX

7.1 Proof of Proposition 3

The residue theorem entails

$$[z^n]f_m(z) = -\mathrm{Res}\left(\frac{f_m(z)}{z^{n+1}}, z=1\right) + \frac{1}{2i\pi}\int_{C_r}\frac{f_m(z)}{z^{n+1}}dz$$

where C_r is the circle with radius $r>1$ with r close enough to 1. Then, with hypotheses on A, the following upper bound holds when m is of the form $m=cn$,

$$\left|\frac{1}{2i\pi}\int_{C_r}\frac{f_m(z)}{z^{n+1}}dz\right| \le \left(\frac{A(r)^c}{r}\right)^n \frac{1}{r(r-1)^\ell}.$$

The function $r \to (1/r)A(r)^c$ equals a^c at $r=1$ and the conditions on c, $a=A(1)$ and $b=A'(1)$ prove that it is strictly decreasing when r is close to 1. Hence, for $r>1$ close enough to 1, we let $\rho = A(r)^c/r = \theta a^c$ with $\theta<1$ and

$$\left|\frac{1}{2i\pi}\int_{C_r}\frac{f_m(z)}{z^{n+1}}dz\right| = a^m O(\theta^n).$$

Furthermore, the residue at $z=1$ equals

$$\mathrm{Res}\left(\frac{f_m(z)}{z^{n+1}}, z=1\right) = \frac{(-1)^\ell}{(\ell-1)!}\left[\frac{d^{\ell-1}}{dz^{\ell-1}}\frac{A(z)^m}{z^{n+1}}\right]_{z=1}$$

The successive derivatives of $1/z^{n+1}$ satisfy

$$\left[\frac{d^i}{dz^i}\frac{1}{z^{n+1}}\right]_{z=1} = (-1)^i\cdot n^i\left[1+O\left(\frac{1}{n}\right)\right].$$

There are two cases for the derivatives of $A(z)^m$.
Case $m/n \to c$ with $c>0$. Then

$$\left[\frac{d^j}{dz^j}A(z)^m\right]_{z=1} = b^j\cdot m^j\cdot a^{m-j}\left[1+O\left(\frac{1}{m}\right)\right],$$

and, inserting these relations into the Leibnitz formula for the derivative provides the estimate

$$\mathrm{Res}\left(\frac{f_m(z)}{z^{n+1}}, z=1\right) = -a^m\frac{n^{\ell-1}}{(\ell-1)!}\left(1-\frac{b}{a}c\right)^{\ell-1}\left[1+O\left(\frac{1}{n}\right)\right].$$

This concludes the proof of Assertion (b).
Case $m=\epsilon(n)\cdot n$ with $\epsilon(n)\to 0$. Then

$$\left[\frac{d^j}{dz^j}A(z)^m\right]_{z=1} = a^m n^j\left(\frac{b}{a}\right)^j O(\epsilon(n)^j)$$

and, inserting these relations into the Leibnitz formula for the derivative provides the estimate

$$\mathrm{Res}\left(\frac{f_m(z)}{z^{n+1}}, z=1\right) = -a^m\frac{n^{\ell-1}}{(\ell-1)!}\left[1+O(\epsilon(n))\right]\left[1+O\left(\frac{1}{n}\right)\right].$$

7.2 Proof of Theorem 4

The variable L equals $\widetilde{L} := L_1+\ldots+L_\Pi$ if $\Pi\ge 1$ and equals 0 if $\Pi=0$. Then

$$L = 0\cdot\mathbf{1}_{[\Pi=0]} + \widetilde{L}\cdot\mathbf{1}_{[\Pi\ge 1]} = \widetilde{L}\cdot\mathbf{1}_{[\Pi\ge 1]}$$

When $\Pi\ge 1$, the variable $\widetilde{L}$ is defined as

$$\widetilde{L} = \sum_{k=1}^{\Pi} L_k = \sum_{k=1}^{\ell-1} L_k\cdot\mathbf{1}_{[k\le\Pi]},$$

so that $\widetilde{L}\cdot\mathbf{1}_{[\Pi\ge 1]} = \widetilde{L}$ and then finally $L=\widetilde{L}$. We then study the variables

$$\widetilde{L}_k = L_k\cdot\mathbf{1}_{[k\le\Pi]}.$$

In particular, the difference $L_k-\widetilde{L}_k$ satisfies

$$L_k-\widetilde{L}_k = L_k\cdot\mathbf{1}_{[k>\Pi]},$$

and the relations $[k>\Pi]=[D_k=0]\subset[L_k=1]$, entail the equality $L_k\cdot\mathbf{1}_{[k>\Pi]}=\mathbf{1}_{[k>\Pi]}$, and, finally

$$\mathbb{E}_n[L_k]-\mathbb{E}_n[\widetilde{L}_k] = \mathbb{E}_n[\mathbf{1}_{[k>\Pi]}] = \mathbb{P}_n[k>\Pi] = 1-\mathbb{P}_n[k\le\Pi].$$

Then Theorem 3 applies and we obtain for $k\ge 2$,

$$\mathbb{E}_n[\widetilde{L}_k] = \frac{q-1}{q^k-1}+\frac{q}{q^k}+O\left(\frac{1}{n}\right),$$

and for $k=1$,

$$\mathbb{E}_n[\widetilde{L}_1] = \mathbb{E}_n[L_1]+O\left(\frac{1}{n}\right).$$

We conclude with the linearity of the mean.

We now prove the beta limit law. We split the random variable L into two random variables:
– the *main* random variable L_1, which admits a beta limit law (Theorem 2),
– the *remainder* random variable $R=L-L_1$ which is "negligible" w.r.t L_1; more precisely, Theorem 1 entails the estimate $\mathbb{E}_n[|R|]=o(\mathbb{E}_n[L_1])$.
The next proposition shows that, in this situation, the sum $L=L_1+R$ asymptotically follows the same beta law as L_1. This provides an extension of the result obtained in [15] for gaussian laws.

PROPOSITION 6. *Consider two random variables X and Y defined on Ω. Assume that there exist*
(i) a function $f:[0,a]\to[0,1]$, strictly increasing, Lipschitz, with $f(0)=0$ and $f(a)=1$,
(ii) three sequences $\gamma_n\to\infty$, $\epsilon_n\to 0$, $r_n\to 0$,
such that, one has, for all $c\in]0,a[$, for all n,

$$\mathbb{P}_n[X<c\cdot\gamma_n] = f(c)+\epsilon_n \qquad \mathbb{E}_n[|Y|] = r_n\cdot\gamma_n.$$

Then the following holds, for all $c\in]0,a[$, for all n,:

$$\mathbb{P}_n[X+Y<c\cdot\gamma_n] = f(c)+O\left(\epsilon_n+\sqrt{r_n}\right).$$

The random variable $X+Y$ asymptotically follows the same law as X.

PROOF. Consider a sequence δ_n which will be made precise later, and define the two events E and F as

$$E=[X+Y<c\cdot\gamma_n], \qquad F=[|Y|\le\delta_n].$$

The hypotheses on Y and the Markov inequality lead to

$$\mathbb{P}_n[E\cap F^c]\le\mathbb{P}_n[F^c] = O\left(\frac{r_n\gamma_n}{\delta_n}\right). \tag{12}$$

On the other side, the following inclusions hold,

$$[X\le c\cdot\gamma_n-\delta_n]\cap F\subset E\cap F\subset[X\le c\cdot\gamma_n+\delta_n]. \tag{13}$$

The rightmost inclusion in (13) and the Lipschitz condition on f entail the upper bound

$$\mathbb{P}_n[E\cap F]\le f\left(c+\frac{\delta_n}{\gamma_n}\right)+O(\epsilon_n) = f(c)+O\left(\epsilon_n+\frac{\delta_n}{\gamma_n}\right). \tag{14}$$

The leftmost inclusion in (13), together with (12) and the Lipschitz condition entail the lower bound

$$\mathbb{P}_n[E \cap F] \geq f(c) + O\left(\epsilon_n + \frac{r_n\gamma_n}{\delta_n} + \frac{\delta_n}{\gamma_n}\right). \qquad (15)$$

With relations (12), (14) and (15), we obtain

$$\mathbb{P}_n[E] = f(c) + O\left(\epsilon_n + \frac{r_n\gamma_n}{\delta_n} + \frac{\delta_n}{\gamma_n}\right).$$

Then the optimal choice $\delta_n = \gamma_n\sqrt{r_n}$ concludes the proof. □

To derive the proof of Theorem 4, the previous proposition was applied with

$$\epsilon_n = \Theta\left(\frac{1}{n}\right), \quad r_n = \Theta\left(\frac{1}{n}\right), \quad \gamma_n = n, \quad a = \frac{q-1}{2q}$$

and

$$f(c) = 1 - \left(1 - \frac{c}{a}\right)^{\ell-1}.$$