Combinatorics, Probability and Computing (2022), **31**, pp. 166–183
doi:10.1017/S0963548321000274

CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Average-case complexity of the Euclidean algorithm with a fixed polynomial over a finite field

Nardo Giménez[1], Guillermo Matera[1,2,3,*], Mariana Pérez[3,4], and Melina Privitelli[3,5,†]

[1]Universidad Nacional de General Sarmiento, Instituto del Desarrollo Humano, J.M. Gutiérrez 1150 (B1613GSX) Los Polvorines, Buenos Aires, Argentina, [2]Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales, Departamento de Matemática, Ciudad Universitaria, Pabellón I (1428) Buenos Aires, Argentina, [3]Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina, [4]Universidad Nacional de Hurlingham, Instituto de Tecnología e Ingeniería, Av. Gdor. Vergara 2222 (B1688GEZ), Villa Tesei, Buenos Aires, Argentina and [5]Universidad Nacional de General Sarmiento, Instituto de Ciencias, J.M. Gutiérrez 1150 (B1613GSX) Los Polvorines, Buenos Aires, Argentina
*__Corresponding author. Emails:__ gmatera@campus.ungs.edu.ar, materaguillermo@gmail.com

(Received 26 December 2019; revised 21 November 2020; accepted 3 May 2021; first published online 6 July 2021)

Abstract

We analyse the behaviour of the Euclidean algorithm applied to pairs (g,f) of univariate nonconstant polynomials over a finite field $\mathbb{F}_q$ of q elements when the highest degree polynomial g is fixed. Considering all the elements f of fixed degree, we establish asymptotically optimal bounds in terms of q for the number of elements f that are relatively prime with g and for the average degree of $\gcd(g,f)$. We also exhibit asymptotically optimal bounds for the average-case complexity of the Euclidean algorithm applied to pairs (g,f) as above.

Keywords: Finite fields; rational points; Euclidean algorithm; symmetric functions; resultant; average-case complexity

2020 MSC Codes: Primary 12E20; Secondary 14G05, 14G15, 68W40

1. Introduction

Let $\mathbb{F}_q$ be the finite field of q elements, where q is a prime power, let T be an indeterminate over $\mathbb{F}_q$ and $\mathbb{F}_q[T]$ the ring of univariate polynomials in T with coefficients in $\mathbb{F}_q$. In this paper we are concerned with the polynomial gcd problem for elements of $\mathbb{F}_q[T]$, namely the problem of computing the greatest common divisor of two non-zero polynomials in $\mathbb{F}_q[T]$.

The fundamental computational tool for this problem is the Euclidean algorithm, and many variants of it are known in the literature (see, e.g. [6]). It is well known that the Euclidean algorithm in $\mathbb{F}_q[T]$ requires a number of polynomial divisions which is linear in the degree of the input polynomials. In particular, we are interested in its average-case complexity, which has been the subject of several papers. The paper [13] establishes the average-case complexity of the Euclidean algorithm and some variants of it, based on explicit counting. In [14], the average-case complexity of variants of the Euclidean algorithm is considered using generating functions. Finally, [2,12] analyse the average-case complexity and related costs of the Euclidean algorithm and variants using tools of analytic combinatorics such as bivariate generating functions.

[†] The authors were partially supported by the grants PIP CONICET 11220130100598 and PIO CONICET-UNGS 14420140100027

© The Author(s), 2021. Published by Cambridge University Press.

https://doi.org/10.1017/S0963548321000274 Published online by Cambridge University Press

All these results consider the average, for fixed degrees $e > d > 0$, over the set of pairs $(g, f) \in \mathbb{F}_q[T] \times \mathbb{F}_q[T]$ with g monic of degree e and f either of degree at most d or of degree less than e, assuming the uniform distribution of pairs. Nevertheless, there are important tasks, which rely heavily on the computation of gcd's and lie outside the scope of these analyses. For example, a critical step in the standard algorithm for finding the roots in $\mathbb{F}_q$ of a polynomial $f \in \mathbb{F}_q[T]$ with $\deg f < q$ consists of computing $\gcd(T^q - T, f)$ (see, e.g. [6]). As the first element in the pair $(T^q - T, f)$ is a fixed polynomial, average-case analyses as before do not contribute to the analysis of the complexity of this problem.

In this paper we consider, for fixed degrees $e > d > 0$ and a fixed (arbitrary) $g \in \mathbb{F}_q[T]$ monic of degree e, the average-case complexity of the Euclidean algorithm over the set of pairs (g,f) with $f \in \mathbb{F}_q[T]$ monic of degree d, endowed with the uniform probability. We shall be interested in the case $q \gg e$; in this sense, all our results may be regarded as asymptotic in q.

We discuss a number of issues concerning this case of the Euclidean algorithm. Our first result shows that the average degree $\mathbb{E}[\mathcal{X}_g]$ of $\gcd(g, f)$ for a random element f of $\mathbb{F}_q[T]$, monic of degree d, decreases fast as q tends to infinity. Further, we prove that the decrease rate depends on the factorisation pattern of g. We have the following result (see Theorem 4.5 for a precise statement).

Theorem 1.1. *Let e, d be integers with $e > d > 0$, let $g \in \mathbb{F}_q[T]$ be a monic polynomial of degree e having factorisation pattern $(\lambda_1, \ldots, \lambda_e)$, with $\lambda_1 + 2\lambda_2 + \cdots + e\lambda_e = e$, and k the least index with $\lambda_k > 0$. Denote by λ_k^* the number of distinct irreducible factors of g in $\mathbb{F}_q[T]$ of degree k. If $k \leq d$, then the average degree $\mathbb{E}[\mathcal{X}_g]$ of the greatest common divisor of g and a random monic element $f \in \mathbb{F}_q[T]$ of degree d satisfies*

$$\left| \mathbb{E}[\mathcal{X}_g] - \frac{k\lambda_k^*}{q^k} \right| = \mathcal{O}\left(\frac{1}{q^{k+1}} \right).$$

The average degree of the gcd of a random pair of elements in $\mathbb{F}_q[T]$ of degrees e and d as above is $(1 - q^{-d})/(q - 1)$ (see [13, Corollary 2.6]). Our result, although not as precise as the latter, confirms that in our case the average degree of the gcd is $\mathcal{O}(q^{-1})$ (for fixed d,e).

We also show that, with high probability, g and a random monic polynomial f of $\mathbb{F}_q[T]$ of degree d are relatively prime. In fact, we have the following result (see Theorem 4.2).

Theorem 1.2. *With assumptions and notations as in Theorem 1, the probability $\mathcal{P}_0$ that* $\gcd(g, f) = 1$ *when f runs through all the monic elements of $\mathbb{F}_q[T]$ of degree d satisfies the estimate*

$$\left| \mathcal{P}_0 - \left(1 - \frac{\lambda_k^*}{q^k} \right) \right| = \mathcal{O}\left(\frac{1}{q^{k+1}} \right).$$

This may be compared with the probability $1 - 1/q$ that a random pair of elements of $\mathbb{F}_q[T]$ of degrees e and d are relatively prime (see, e.g. [13, Proposition 2.4]).

To prove Theorem 1.2 we observe that a monic $f \in \mathbb{F}_q[T]$ of degree d is relatively prime with g if and only if the resultant $\mathrm{res}(g, f)$ does not vanish. This leads us to consider the set of zeros in $\mathbb{F}_q^d$ of the 'generic' resultant $\mathrm{res}(g, T^d + S_{d-1}T^{d-1} + \cdots + S_0)$, which represents the set of vectors of coefficients of the f such that $\gcd(g, f) = 1$. This generic resultant can be factored in terms of the factorisation pattern of g, which explains the dependence of the estimate of Theorem 1.2 on the factorisation pattern of g. Further, for the proof of Theorem 1.1 we estimate the number of monic $f \in \mathbb{F}_q[T]$ of degree d with $\deg\gcd(g, f) \geq i$ for $1 \leq i \leq d$. Observe that the case $i = 1$ is closely related to the number of f such that $\gcd(g, f) = 1$.

Finally, we analyse the average number $\mathbb{E}[t_g^{\mathrm{div}}]$, $\mathbb{E}[t_g^{\div}]$ and $\mathbb{E}[t_g^{-,\times}]$ of polynomial divisions, divisions in $\mathbb{F}_q$, and additions/multiplications in $\mathbb{F}_q$, performed by the Euclidean algorithm. We have the following result (see Theorem 5.4).

https://doi.org/10.1017/S0963548321000274 Published online by Cambridge University Press

Theorem 1.3. *Let e, d be positive integers such that $q > d(2e - d + 1)/2$ and $e > d$. Let $g \in \mathbb{F}_q[T]$ be a monic polynomial of degree e and $\mathsf{w} \in \{\mathsf{div}, \div, -, \times\}$. The average number $\mathbb{E}[t_g^{\mathsf{w}}]$ of operations w performed on (uniformly distributed) monic inputs from $\mathbb{F}_q[T]$ of degree d is bounded in the following way:*

$$\left|\frac{\mathbb{E}[t_g^{\mathsf{div}}]}{d+1} - 1\right| \leq \frac{de}{q}, \qquad \left|\frac{\mathbb{E}[t_g^{\div}]}{e+d+1} - 1\right| \leq \frac{de}{q}, \qquad \left|\frac{\mathbb{E}[t_g^{-,\times}]}{de} - 1\right| \leq \frac{de}{q}.$$

The main terms in these bounds agree with those in the corresponding ones for random pairs of polynomials of degrees e and d with $e > d$, according to [13, Theorem 2.1].

A critical point to prove Theorem 1.3 is a lower bound on the number of polynomials f for which the Euclidean algorithm performs the highest possible number of steps, namely d. It is well known that, on input two generic polynomials of degrees $e > d > 0$, the Euclidean algorithm performs d steps, and the degrees of the successive remainders decrease by one in each step (see, e.g. [11]). To establish such a lower bound we compare the 'formal' execution of the Euclidean algorithm on polynomials whose roots are indeterminates with its actual execution on g and a monic polynomial f of degree d. It turns out that any f for which the actual execution is not a specialisation of the formal execution must annihilate a leading coefficient of a remainder of the formal execution. Combining a description of this coefficient in terms symmetric functions due to [11] with an upper bound on the number of zeros with coordinates in $\mathbb{F}_q$ of multivariate polynomials with coefficients in $\mathbb{F}_q$, we establish the lower bound.

The paper is organised as follows. In Section 2, we recall the description of remainders and quotients arising in the formal execution of the Euclidean algorithm in terms of symmetric functions. In Section 3, we use this machinery to estimate the degrees of the leading coefficients of the remainders of the formal execution and consider the behaviour of the Euclidean algorithm under specialisations. In Section 4, we estimate the number of polynomials f for which $\gcd(g, f)$ has at least a given degree, which is used to prove Theorems 1.1 and 1.2. Finally, in Section 5, we use the results of Sections 3 and 4 to establish the results on the average-case complexity.

2. Basic notions and notations

Let $\mathbb{F}_q$ be the finite field of q elements and $\overline{\mathbb{F}}_q$ its algebraic closure. Let $X_1, \ldots, X_n$ be indeterminates over $\overline{\mathbb{F}}_q$. For $\mathbb{K} = \mathbb{F}_q$ or $\mathbb{K} = \overline{\mathbb{F}}_q$, we denote by $\mathbb{K}[X_1, \ldots, X_n]$ the ring of multivariate polynomials in $X_1, \ldots, X_n$ with coefficients in $\mathbb{K}$. By $\mathbb{A}^n$ we denote the affine n–dimensional space $\mathbb{A}^n := \overline{\mathbb{F}}_q^n$, endowed with its Zariski topology over $\mathbb{K}$, for which a closed set is the zero locus of a set of polynomials of $\mathbb{K}[X_1, \ldots, X_n]$. A subset $V \subset \mathbb{A}^n$ is an *affine variety defined over* $\mathbb{K}$ (or an affine $\mathbb{K}$–variety) if it is the set of common zeros in $\mathbb{A}^n$ of polynomials $F_1, \ldots, F_m \in \mathbb{K}[X_1, \ldots, X_n]$. We shall denote by $\mathcal{V}(F_1, \ldots, F_m)$ the affine $\mathbb{K}$–variety consisting of the common zeros of $F_1, \ldots, F_m$.

A $\mathbb{K}$–variety V is *irreducible* if it cannot be expressed as a finite union of proper $\mathbb{K}$–subvarieties of V. Any $\mathbb{K}$–variety V can be expressed as an irredundant union $V = \mathcal{C}_1 \cup \cdots \cup \mathcal{C}_s$ of irreducible $\mathbb{K}$–varieties, unique up to reordering, called the *irreducible* $\mathbb{K}$– *components* of V. We say that V has *pure dimension* r if every irreducible $\mathbb{K}$–component of V has dimension r. A $\mathbb{K}$–variety of $\mathbb{A}^n$ of pure dimension $n - 1$ is called a $\mathbb{K}$– *hypersurface*. A $\mathbb{K}$–hypersurface of $\mathbb{A}^n$ can also be described as the set of zeros of a single non-zero polynomial of $\mathbb{K}[X_1, \ldots, X_n]$.

The *degree* $\deg V$ of an irreducible $\mathbb{K}$–variety V is the maximum of the cardinality $|V \cap L|$ of $V \cap L$, considering all the linear spaces L of codimension $\dim V$ such that $|V \cap L| < \infty$. More generally, following [8] (see also [5]), if $V = \mathcal{C}_1 \cup \cdots \cup \mathcal{C}_s$ is the decomposition of V into irreducible $\mathbb{K}$–components, we define the degree of V as

https://doi.org/10.1017/S0963548321000274 Published online by Cambridge University Press

$$\deg V := \sum_{i=1}^{s} \deg \mathcal{C}_i.$$

The degree of a $\mathbb{K}$–hypersurface V is the degree of a polynomial of minimal degree defining V. In particular, the degree of a linear variety is equal to 1.

Let $\mathbb{A}^n(\mathbb{F}_q)$ be the n–dimensional $\mathbb{F}_q$–vector space $\mathbb{F}_q^n$. For an $\mathbb{K}$–affine variety $V \subset \mathbb{A}^n$, the set of $\mathbb{F}_q$–rational points $V(\mathbb{F}_q)$ of V is defined as $V(\mathbb{F}_q) := V \cap \mathbb{A}^n(\mathbb{F}_q)$. For an affine $\mathbb{K}$–variety $V \subset \mathbb{A}^n$ of dimension r and degree $d \geq 0$, we have the following bound (see, e.g. [3, Lemma 2.1]):

$$|V(\mathbb{F}_q)| \leq d\, q^r. \tag{2.1}$$

2.1. Symmetric functions and the Euclidean algorithm

Next, we gather the terminology and results we will use concerning the description of the Euclidean algorithm in terms of symmetric functions, following [11].

We call a finite set of indeterminates $\boldsymbol{A}$ over $\mathbb{F}_q$ an *alphabet* and denote its cardinality by $|\boldsymbol{A}|$. The *elementary symmetric functions* $\Lambda^i(\boldsymbol{A})$ and the *complete functions* $S^i(\boldsymbol{A})$ $(i \geq 0)$ are defined by means of the following identities of formal power series in the variable z:

$$\prod_{a \in \boldsymbol{A}} (1 + za) = \sum_{i \geq 0} \Lambda^i(\boldsymbol{A}) z^i, \quad \prod_{a \in \boldsymbol{A}} \frac{1}{1 - za} = \sum_{i \geq 0} S^i(\boldsymbol{A}) z^i.$$

We further define $\Lambda^i(\boldsymbol{A}) := 0$ and $S^i(\boldsymbol{A}) := 0$ for $i < 0$. Observe that $\Lambda^i(\boldsymbol{A}) = 0$ if $i > |\boldsymbol{A}|$. Writing $\boldsymbol{A} + \boldsymbol{B}$ for the disjoint union of two alphabets $\boldsymbol{A}$ and $\boldsymbol{B}$, we have the following *Cauchy formulas*:

$$\Lambda^i(\boldsymbol{A} + \boldsymbol{B}) = \sum_{j+k=i} \Lambda^j(\boldsymbol{A}) \Lambda^k(\boldsymbol{B}), \quad S^i(\boldsymbol{A} + \boldsymbol{B}) = \sum_{j+k=i} S^j(\boldsymbol{A}) S^k(\boldsymbol{B}), \tag{2.2}$$

Define $S^i(\boldsymbol{A} - \boldsymbol{B})$ $(i \geq 0)$ by means of the identity

$$\frac{\prod_{b \in \boldsymbol{B}} (1 - zb)}{\prod_{a \in \boldsymbol{A}} (1 - za)} = \sum_{i \geq 0} S^i(\boldsymbol{A} - \boldsymbol{B}) z^i,$$

and set $S^i(\boldsymbol{A} - \boldsymbol{B}) := 0$ for $i < 0$. Define $S^i(-\boldsymbol{A}) := (-1)^i \Lambda^i(\boldsymbol{A})$ for any integer i. Thus, besides (2.2) we have

$$S^i(\boldsymbol{A} - \boldsymbol{B}) = \sum_{j+k=i} S^j(\boldsymbol{A}) S^k(-\boldsymbol{B}). \tag{2.3}$$

We shall express polynomials using this terminology. Indeed, let $n := |\boldsymbol{A}|$ and identify a single indeterminate T with the alphabet $\{T\}$. Since $S^i(T) = T^i$ for $i \geq 0$, according to (2.3) we have that

$$S^n(T - \boldsymbol{A}) = \sum_{i=0}^{n} S^{n-i}(-\boldsymbol{A}) T^i$$

is the polynomial in T having $\boldsymbol{A}$ as its set of roots.

2.2. Remainders as symmetric functions

Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two alphabets of cardinalities e and d respectively, with $e > d$. When the Euclidean algorithm is applied to the generic polynomials $S^e(T - \boldsymbol{A})$ and $S^d(T - \boldsymbol{B})$ we obtain d remainders $\mathcal{R}_1, \ldots, \mathcal{R}_d$ and quotients $q_1, \ldots, q_d$ satisfying the following identities:

https://doi.org/10.1017/S0963548321000274 Published online by Cambridge University Press

$$\begin{aligned} S^e(T-\boldsymbol{A}) &= q_1 S^d(T-\boldsymbol{B}) + \mathcal{R}_1,\\ S^d(T-\boldsymbol{B}) &= q_2\mathcal{R}_1 + \mathcal{R}_2,\\ \mathcal{R}_1 &= q_3\mathcal{R}_2 + \mathcal{R}_3,\\ &\vdots\\ \mathcal{R}_{d-2} &= q_d\mathcal{R}_{d-1} + \mathcal{R}_d. \end{aligned} \tag{2.4}$$

Here $\deg_T q_1 = e-d$, $\deg_T q_i = 1$ for $2 \le i \le d$, and $\deg_T \mathcal{R}_i = d-i$ for $1 \le i \le d$. It turns out that all the remainders $\mathcal{R}_i$ are elements of the ring $\mathbb{F}_q[\boldsymbol{A},\boldsymbol{B}][T]$. Further, we may express these remainders in terms of symmetric functions [11, equation (3.1.4)]:

$$\mathcal{R}_k = \det \begin{pmatrix} S^0(-\boldsymbol{A}) & \cdots & S^{e-d+2k-2}(-\boldsymbol{A}) & S^{e+k-1}(T-\boldsymbol{A})\\ \vdots & & \vdots & \vdots\\ S^{-k+1}(-\boldsymbol{A}) & \cdots & S^{e-d+k-1}(-\boldsymbol{A}) & S^e(T-\boldsymbol{A})\\ S^0(-\boldsymbol{B}) & \cdots & S^{e-d+2k-2}(-\boldsymbol{B}) & S^{e+k-1}(T-\boldsymbol{B})\\ \vdots & & \vdots & \vdots\\ S^{d-e-k+1}(-\boldsymbol{B}) & \cdots & S^{k-1}(-\boldsymbol{B}) & S^d(T-\boldsymbol{B}) \end{pmatrix}. \tag{2.5}$$

3. The formal execution of the Euclidean algorithm and specialisations

Our methodology consists in comparing the formal execution of the Euclidean algorithm on the input elements $S^e(T-\boldsymbol{A}) \in \mathbb{F}_q[\boldsymbol{A}][T]$ and $S^d(T-\boldsymbol{B}) \in \mathbb{F}_q[\boldsymbol{B}][T]$ with its execution on two polynomials $g,f \in \mathbb{F}_q[T]$ of degrees e and d, respectively.

As expressed in Section 2.2, on input $S^e(T-\boldsymbol{A})$ and $S^d(T-\boldsymbol{B})$, the Euclidean algorithm performs d steps. Further, if the Euclidean algorithm is applied to polynomials $g,f \in \mathbb{F}_q[T]$ with $\deg g = e$ and $\deg f = d$, and performs d steps, then the degrees of the successive remainders decrease by 1 each step, and the sequence of quotients and remainders associated to g and f coincides with the specialisation of the sequence associated to $S^e(T-\boldsymbol{A})$ and $S^d(T-\boldsymbol{B})$. We refer to the latter as the *formal sequence*.

Our focus is on the set of pairs (g,f) for which the corresponding sequence does not have the degree pattern of the formal sequence. We show that, for fixed g, any f with 'bad' behaviour annihilates a leading coefficient of the partial specialisation of a remainder of the formal sequence in the vector of coefficients of g. Combining this with a precise analysis of the structure of such leading coefficients will allow us to conclude that, for fixed g and random f, 'most' sequences have the degree pattern of the formal sequence.

For this purpose, we devote Section 3.1 to the analysis of the leading coefficients of the remainders of the formal sequence. Then, in Section 3.2, we fix g and establish a characterisation of the set of f with 'bad' behaviour, which leads us to the quantitative result concerning such a set of f.

3.1. Degree bounds for the remainders in the formal execution

In the sequel, for $1 \le k \le d$ we denote by $F_k \in \mathbb{F}_q[\boldsymbol{A},\boldsymbol{B}]$ the leading coefficient of $\mathcal{R}_k$, considered as an element of $\mathbb{F}_q[\boldsymbol{A},\boldsymbol{B}][T]$. Let $\mathbf{S} := (S^i(-\boldsymbol{B}) : 1 \le i \le d)$ and $\mathbf{R} := (S^i(-\boldsymbol{A}) : 1 \le i \le e)$. Observe that both $\mathbf{S}$ and $\mathbf{R}$ are algebraically independent sets over $\mathbb{F}_q$.

https://doi.org/10.1017/S0963548321000274 Published online by Cambridge University Press

Proposition 3.1. *F_k is a non-zero element of $\mathbb{F}_q[\boldsymbol{R}][\boldsymbol{S}]$ of degree $\deg_{\boldsymbol{S}} F_k = e-d+k$. Further, $\pm F_k$ is monic of degree $e-d+k$ in $S^k(-\boldsymbol{B})$.*

Proof. Taking into account the linearity of the determinant of the right-hand side of (2.5) with respect to the last column, we conclude that

$$F_k = \det \begin{pmatrix} S^0(-\boldsymbol{A}) & \cdots & S^{e-d+2k-2}(-\boldsymbol{A}) & S^{e-d+2k-1}(-\boldsymbol{A}) \\ \vdots & & \vdots & \vdots \\ S^{-k+1}(-\boldsymbol{A}) & \cdots & S^{e-d+k-1}(-\boldsymbol{A}) & S^{e-d+k}(-\boldsymbol{A}) \\ S^0(-\boldsymbol{B}) & \cdots & S^{e-d+2k-2}(-\boldsymbol{B}) & S^{e-d+2k-1}(-\boldsymbol{B}) \\ \vdots & & \vdots & \vdots \\ S^{d-e-k+1}(-\boldsymbol{B}) & \cdots & S^{k-1}(-\boldsymbol{B}) & S^k(-\boldsymbol{B}) \end{pmatrix}, \tag{3.1}$$

where the entries in the last column of the determinant in the right-hand side of (3.1) are the coefficients of T^{d-k} of the corresponding entries in the last column of (2.5).

We expand this determinant along the last column. Let $N := e-d+2k$. For $1 \leq j \leq N$, let A_j denote the $(N-1)$th minor of the matrix of (3.1) obtained by deleting the last column and the j-row of this matrix. We have

$$\begin{aligned} F_k = &\pm S^{e-d+2k-1}(-\boldsymbol{A})A_1 \pm \cdots \pm S^{e-d+k}(-\boldsymbol{A})A_k \\ &\pm S^{e-d+2k-1}(-\boldsymbol{B})A_{k+1} \pm \cdots \pm S^k(-\boldsymbol{B})A_N. \end{aligned} \tag{3.2}$$

Since the $S^h(-\boldsymbol{B})$ only occur in the last $N-k$ rows of the matrix in (3.1), it follows that

$$\deg_{\boldsymbol{S}} F_k \leq N-k = e-d+k. \tag{3.3}$$

Further, the $S^h(-\boldsymbol{B})$ only occur in the last $N-k$ rows of A_j for $1 \leq j \leq k$, and in the last $N-1-k$ rows of A_j for $k+1 \leq j \leq N$. We deduce that

$$\begin{aligned} &\deg_{\boldsymbol{S}} A_j \leq N-k = e-d+k \quad \text{for } 1 \leq j \leq k, \\ &\deg_{\boldsymbol{S}} A_j \leq N-1-k = e-d+k-1 \quad \text{for } k+1 \leq j \leq N. \end{aligned} \tag{3.4}$$

Claim 1. *A_N is a non-zero element of $\mathbb{F}_q[\boldsymbol{R}][\boldsymbol{S}]$ with $\deg_{\boldsymbol{S}} A_N = e-d+k-1$. Further, A_N is monic of degree $e-d+k-1$ in $S^k(-\boldsymbol{B})$.*

Proof of Claim. Observe that the determinantal expression of A_N is

$$A_N = \det \begin{pmatrix} S^0(-\boldsymbol{A}) & \cdots & S^{e-d+2k-2}(-\boldsymbol{A}) \\ \vdots & & \vdots \\ S^{-k+1}(-\boldsymbol{A}) & \cdots & S^{e-d+k-1}(-\boldsymbol{A}) \\ S^0(-\boldsymbol{B}) & \cdots & S^{e-d+2k-2}(-\boldsymbol{B}) \\ \vdots & & \vdots \\ S^{d-e-k+2}(-\boldsymbol{B}) & \cdots & S^k(-\boldsymbol{B}) \end{pmatrix}. \tag{3.5}$$

More precisely, $A_N = \det(a_{ij})_{1 \leq i,j \leq N-1}$ with $a_{ij} := S^{j-i}(-\boldsymbol{A})$ for $1 \leq i \leq k$ and $a_{ij} := S^{j-i+k}(-\boldsymbol{B})$ for $k+1 \leq i \leq N-1$. In particular, the diagonal of the matrix of (3.5) has $S^0(-\boldsymbol{A})$ in the first k columns and $S^k(-\boldsymbol{B})$ in the last $N-1-k$ columns. Write

$$A_N = \sum_{\sigma} \pm a_{1\sigma_1} a_{2\sigma_2} \cdots a_{N-1\sigma_{N-1}},$$

where $\sigma = (\sigma_1, \ldots, \sigma_{N-1})$ runs over all permutations of $\{1, 2, \ldots, N-1\}$. By (3.4) we have $\deg_{\boldsymbol{S}} A_N \leq e-d+k-1$. To prove the equality, consider a permutation $(\sigma_1, \ldots, \sigma_{N-1}) \neq (1, 2, \ldots, N-1)$. Then there exists an index i with $\sigma_i < i$. Suppose first that $1 \leq i \leq k$. Then $a_{i\sigma_i} = S^{\sigma_i - i}(-\boldsymbol{A}) = 0$ since $\sigma_i - i < 0$. In such a case, we have $a_{1\sigma_1} a_{2\sigma_2} \cdots a_{N-1\sigma_{N-1}} = 0$. Now suppose that $k+1 \leq i \leq N-1$. Then $a_{i\sigma_i} = S^{\sigma_i - i + k}(-\boldsymbol{B}) \neq S^k(-\boldsymbol{B})$ since $k + \sigma_i - i < k$. Therefore, $a_{1\sigma_1} a_{2\sigma_2} \cdots a_{N-1\sigma_{N-1}}$ has degree at most $N-2-k$ in $S^k(-\boldsymbol{B})$. On the other hand, the term $a_{11} \cdots a_{(N-1)(N-1)} = (S^k(-\boldsymbol{B}))^{N-1-k}$ is monic of degree $N-1-k = e-d+k-1$ in $S^k(-\boldsymbol{B})$. We conclude that A_N is monic of degree $e-d+k-1$ in $S^k(-\boldsymbol{B})$. This proves the claim. □

Observe that $S^k(-\boldsymbol{B})$ is not an entry of the last row of A_j for $j < N$. It follows that $S^k(-\boldsymbol{B})$ occurs as an entry of A_j at most $N-1-k$ times for $j < N$, and then

$$\deg_{S^k(-\boldsymbol{B})} A_j \leq N-1-k = e-d+k-1 \quad \text{for } j < N. \tag{3.6}$$

As the first $N-1$ terms of the right-hand side of (3.2) are either of the form $S^h(-\boldsymbol{A})A_j$, or $S^\ell(-\boldsymbol{B})A_j$ for some $\ell \neq k$ and $j < N$, we deduce that each of these terms have degree at most $e-d+k-1$ in $S^k(-\boldsymbol{B})$. On the other hand, by the claim the last term $S^k(-\boldsymbol{B})A_N$ is monic of degree $e-d+k$ in $S^k(-\boldsymbol{B})$. We conclude that $\pm F_k$ is monic of degree $e-d+k$ in $S^k(-\boldsymbol{B})$. This together with (3.3) proves the proposition. □

3.2. Specialisations of the formal execution

The next result shows that, if the sequence of remainders associated to polynomials $g, f \in \mathbb{F}_q[T]$ with $\deg g = e$ and $\deg f = d$ fails to have the degree pattern of the formal case, the first remainder where such a failure occurs is still a specialisation of the corresponding one of the formal execution.

Lemma 3.2. *For a specialisation $\boldsymbol{A} \mapsto \overline{a}$ and $\boldsymbol{B} \mapsto \overline{b}$ in $\mathbb{F}_q$, denote by $r_1, \ldots, r_k$ the first k remainders of the application of the Euclidean algorithm to $S^e(T-\overline{a})$ and $S^d(T-\overline{b})$. If deg $r_i = d-i$ for $1 \leq i \leq k-1$, then $r_i = \mathcal{R}_i(\overline{a}, \overline{b})$ for $1 \leq i \leq k$. Further, if deg $\mathcal{R}_i(\overline{a}, \overline{b}) = d-i$ for $1 \leq i \leq k-1$, then $r_i = \mathcal{R}_i(\overline{a}, \overline{b})$ for $1 \leq i \leq k$.*

Proof. Substituting $\overline{a}$ for $\boldsymbol{A}$ and $\overline{b}$ for $\boldsymbol{B}$ in the first identity of (2.4) we easily see that $\mathcal{R}_1(\overline{a}, \overline{b})$ is the remainder in the division of $S^e(T-\overline{a})$ by $S^d(T-\overline{b})$, which proves that $r_1 = \mathcal{R}_1(\overline{a}, \overline{b})$. Let $j > 1$ and assume inductively that $r_i = \mathcal{R}_i(\overline{a}, \overline{b})$ for $1 \leq i \leq j < k$. Thus $\deg \mathcal{R}_i(\overline{a}, \overline{b}) = d-i$ for $1 \leq i \leq j$. Taking into account that F_i is the leading coefficient of $\mathcal{R}_i$ for $1 \leq i \leq k$, we deduce that $F_i(\overline{a}, \overline{b}) \neq 0$ for $1 \leq i \leq j$. Since $q_{j+1} \in \mathbb{F}_q[\boldsymbol{A}, \boldsymbol{B}]_{F_j}[T]$, where $\mathbb{F}_q[\boldsymbol{A}, \boldsymbol{B}]_{F_j}$ is the localisation of $\mathbb{F}_q[\boldsymbol{A}, \boldsymbol{B}]$ at F_j, we can substitute $\overline{a}$ for $\boldsymbol{A}$ and $\overline{b}$ for $\boldsymbol{B}$ in the $(j+1)$th equation of (2.4) to obtain

$$r_{j-1} = q_{j+1}(\overline{a}, \overline{b}) r_j + \mathcal{R}_{j+1}(\overline{a}, \overline{b}).$$

Since

$$\deg_T \mathcal{R}_{j+1}(\overline{a}, \overline{b}) \leq \deg_T \mathcal{R}_{j+1} = d-j-1 < \deg_T \mathcal{R}_j(\overline{a}, \overline{b}),$$

https://doi.org/10.1017/S0963548321000274 Published online by Cambridge University Press

we conclude that $\mathcal{R}_{j+1}(\overline{a}, \overline{b})$ is the remainder in the division of r_{j-1} by r_j. In other words, $r_{j+1} = \mathcal{R}_{j+1}(\overline{a}, \overline{b})$, which completes the proof of the first assertion of the lemma. The second assertion is proved with a similar argument. □

Let $\overline{a} \in \overline{\mathbb{F}}_q^e$ be the tuple of roots of g in any order, so that $g = S^e(T - \overline{a})$. Let $G_k := F_k(\mathbf{R}(\overline{a}), \mathbf{S}(-\boldsymbol{B}))$ denote the polynomial obtained by substituting $\overline{a}$ for $\boldsymbol{A}$ in F_k. Since the set $\mathbf{R} := (S^i(-\boldsymbol{A}):1 \leq i \leq e)$ consists of the first e elementary symmetric functions in $\boldsymbol{A}$, it follows that $\mathbf{R}(\overline{a})$ belongs to $\mathbb{F}_q^e$, and thus G_k belongs to $\mathbb{F}_q[\mathbf{S}]$. Further, Proposition 3.1 shows that $\pm G_k$ is a non-zero polynomial with $\deg_{\mathbf{S}} G_k = e - d + k$, which is monic in $S^k(-\boldsymbol{B})$ with $\deg_{S^k(-\boldsymbol{B})} G_k = e - d + k$.

We end this section with a result that will be crucial to establish lower bounds for the average-case complexity of the Euclidean algorithm. In the next section we prove that, for a fixed $g \in \mathbb{F}_q[T]$ with $\deg g = e$ and a random $f \in \mathbb{F}_q[T]$ with $\deg f = d$, the polynomials f and g are relatively prime with high probability. In this sense, we call a polynomial $f \in \mathbb{F}_q[T]$ with $\deg f = d$ *generic* (with respect to g) if the remainder sequence in the Euclidean algorithm applied to the pair (g,f) has length d. In particular, in such a remainder sequence $(r_1, \ldots, r_d)$ we have $\deg(r_k) = d - k$ for $1 \leq k \leq d$. The next result establishes a lower bound on the number generic monic elements in $\mathbb{F}_q[T]$ of degree d.

Proposition 3.3. *Let $\mathcal{G} \subset \mathbb{F}_q[T]$ be the set of monic elements of degree d which are generic in the sense above. Then*

$$|\mathcal{G}| \geq q^d \left(1 - \frac{d(2e - d + 1)}{2q}\right).$$

In particular, for $q > d(2e - d + 1)/2$ the set $\mathcal{G}$ is non-empty.

Proof. Let $f := T^d + s_1 T^{d-1} + \cdots + s_d \in \mathcal{G}$ and let $(r_1, \ldots, r_d)$ be the sequence of remainders in the Euclidean algorithm applied to the pair (g,f). By hypothesis $\deg(r_j) = d - j$ for $1 \leq j \leq d$, which by Lemma 3.2 is equivalent to the condition $G_j(s_1, \ldots, s_d) \neq 0$ for $1 \leq j \leq d$. It follows that

$$\mathcal{G} = \bigcap_{j=1}^{d} \left(\mathbb{F}_q^d \setminus \mathcal{V}(G_j)(\mathbb{F}_q)\right) = \mathbb{F}_q^d \setminus \bigcup_{j=1}^{d} \mathcal{V}(G_j)(\mathbb{F}_q).$$

As a consequence,

$$|\mathcal{G}| = q^d - \left|\bigcup_{j=1}^{d} \mathcal{V}(G_j)(\mathbb{F}_q)\right| \geq q^d - \sum_{j=1}^{d} |\mathcal{V}(G_j)(\mathbb{F}_q)|.$$

According to (2.1), we have

$$\sum_{k=1}^{d} |\mathcal{V}(G_k)(\mathbb{F}_q)| \leq q^{d-1} \sum_{k=1}^{d} (e - d + k) = q^{d-1} \frac{d(2e - d + 1)}{2},$$

which readily implies the proposition. □

4. Analysis of the average degree in the Euclidean algorithm

Let e, d be positive integers with $e > d$. For any $m \geq 0$, we denote by $\mathbb{F}_q[T]_m$ the set of monic polynomials of degree m with coefficients in $\mathbb{F}_q$. From now on we fix $g \in \mathbb{F}_q[T]_e$ and consider the random variable

$$\mathcal{X}_g : \mathbb{F}_q[T]_d \to \{0, \ldots, d\}, \quad \mathcal{X}_g(f) = \deg(\gcd(g, f)),$$

https://doi.org/10.1017/S0963548321000274 Published online by Cambridge University Press

defined by the degree of the greatest common divisor gcd (g,f), where $\mathbb{F}_q[T]_d$ is endowed with the uniform probability. Applying the Euclidean algorithm to a pair (g,f) with $f \in \mathbb{F}_q[T]_d$ we obtain a positive integer k with $1 \leq k \leq d$, a unique polynomial quotient sequence $(q_1, \ldots, q_{k+1})$ and a unique polynomial remainder sequence $(r_1, \ldots, r_k)$, satisfying the following conditions:

$$\begin{array}{ll} g = f \cdot q_1 + r_1, & \deg(r_1) < \deg(f), \\ f = r_1 \cdot q_2 + r_2, & \deg(r_2) < \deg(r_1), \\ \vdots & \vdots \\ r_{k-2} = r_{k-1} \cdot q_k + r_k, & \deg(r_k) < \deg(r_{k-1}), \\ r_{k-1} = r_k \cdot q_{k+1}. & \end{array}$$

In this section we study the average degree of the gcd, namely the expected value of $\mathcal{X}_g$:

$$\mathbb{E}[\mathcal{X}_g] = \sum_{i=0}^{d} i \, \frac{|B_i|}{q^d} = \sum_{i=1}^{d} i \, \frac{|B_i|}{q^d} = \sum_{i=1}^{d} \sum_{j=i}^{d} \frac{|B_j|}{q^d}, \tag{4.1}$$

where $B_i := \{f \in \mathbb{F}_q[T]_d : \mathcal{X}_g(f) = i\}$ for $0 \leq i \leq d$.

For this purpose, we obtain asymptotically optimal estimates for the cardinality of the sets $\bigcup_{j=i}^{d} B_j$ with $1 \leq i \leq d$. Taking into account that $\bigcup_{j=1}^{d} B_j$ is the set of $f \in \mathbb{F}_q[T]_d$ such that the resultant of f and g vanishes, we rely on an analysis of the resultant res$(g, T^d + S_{d-1}T^{d-1} + \cdots + S_0)$ of g with a generic monic polynomial of degree d. More precisely, the number of zeros of this resultant in $\mathbb{F}_q^d$ equals the cardinality of the set $\bigcup_{j=1}^{d} B_j$.

According to the well-known product formula for resultants, this resultant can be factored depending on the factorisation pattern of g, and thus its number of zeros in $\mathbb{F}_q^d$ can be expressed in terms of such a factorisation pattern. Combining simple combinatorics arguments with (2.1) we obtain our estimates for $\big|\bigcup_{j=i}^{d} B_j\big|$ in terms of the factorisation pattern of g. In Section 4.1 we consider the case $i = 1$, which as a byproduct yields a proof of Theorem 1.2. Then in Section 4.2 we consider the case $2 \leq i \leq d$, and prove Theorem 1.1.

4.1. *The set of elements which are relatively prime with g*

We start with an estimate on $\sum_{j=1}^{d} |B_j| = \big|\bigcup_{j=1}^{d} B_j\big|$. For this purpose, observe that

$$\bigcup_{j=1}^{d} B_j = \{f \in \mathbb{F}_q[T]_d : \mathrm{res}(g,f) = 0\},$$

where res$(\cdot,\cdot)$ denotes resultant. We recall that g has factorisation pattern $(\lambda_1, \ldots, \lambda_e) \in \mathbb{Z}_{\geq 0}^e$, with $\lambda_1 + 2\,\lambda_2 + \cdots + e\,\lambda_e = e$, if g has λ_i irreducible factors in $\mathbb{F}_q[T]$ of degree i (counting multiplicities) for $1 \leq i \leq e$. We shall also consider the *reduced* factorisation pattern $(\lambda_1^*, \ldots, \lambda_e^*) \in \mathbb{Z}_{\geq 0}^e$ of g, where λ_i^* is the number of distinct irreducible factors of g in $\mathbb{F}_q[T]_i$ for $1 \leq i \leq e$, and denote by g^* the square-free part of g, namely the product of all distinct irreducible factors of g (without multiplicities). In particular, we have that $(\lambda_1^*, \ldots, \lambda_e^*)$ is the factorisation pattern of g^*. We have the following result.

Proposition 4.1. *Let e,d be integers with $e > d > 0$. Let g be an element of $\mathbb{F}_q[T]_e$, g^* its square-free part and $(\lambda_1^*, \ldots, \lambda_e^*)$ the factorisation pattern of g^*. Let k be the least integer with $\lambda_k^* > 0$. If $k \leq d$, then*

https://doi.org/10.1017/S0963548321000274 Published online by Cambridge University Press

$$\lambda_k^*\, q^{d-k} - \binom{\lambda_k^*}{2} q^{\max\{d-2k,0\}} \leq \left| \bigcup_{j=1}^{d} B_j \right| \leq \lambda_k^*\, q^{d-k} + \sum_{i=k+1}^{d} \lambda_i^*\, q^{d-i}.$$

Proof. For $f \in \mathbb{F}_q[T]_d$ we have $\mathrm{res}(g, f) = 0$ if and only if $\mathrm{res}(g^*, f) = 0$. As a consequence, we shall consider the resultant $\mathrm{res}(g^*, f)$. Denote by $g_i := \prod_{j=1}^{\lambda_i^*} g_{i,j}$ the product of all irreducible factors of g^* of degree i for $1 \leq i \leq e$. Let $\boldsymbol{S} := (S_{d-1}, \ldots, S_0)$ be a vector of indeterminates and

$$F(\boldsymbol{S}, T) := T^d + S_{d-1}T^{d-1} + \cdots + S_0.$$

The product formula for the resultant (see, e.g. [1, Theorem 4.16]) implies

$$\mathrm{res}(g^*, F(\boldsymbol{S}, T)) = \prod_{i=k}^{e} \mathsf{R}_i := \prod_{i=k}^{e} \mathrm{res}(g_i, F(\boldsymbol{S}, T)) = \prod_{i=k}^{e} \prod_{j=1}^{\lambda_i^*} \mathrm{res}(g_{i,j}, F(\boldsymbol{S}, \boldsymbol{T})).$$

Now, for any i with $k \leq i \leq d$ and $\lambda_i^* > 0$, we have

$$\mathsf{R}_i := \prod_{j=1}^{\lambda_i^*} \mathsf{R}_{i,j}, \qquad \mathsf{R}_{i,j} := \mathrm{res}(g_{i,j}, F(\boldsymbol{S}, T)).$$

Since $g_{i,j}$ is an irreducible element of $\mathbb{F}_q[T]$, for $\boldsymbol{s} \in \mathbb{F}_q^d$ we have $\mathsf{R}_{i,j}(\boldsymbol{s}) = 0$ if and only if $g_{i,j}$ divides $F(\boldsymbol{s}, T)$. Further, as $\{F(\boldsymbol{s}, T) : \boldsymbol{s} \in \mathbb{F}_q^d\} \subset \mathbb{F}_q[T]_d$, we conclude that there is a bijection between the set of $\mathbb{F}_q$-rational zeros of $\mathsf{R}_{i,j}$ and the set of multiples in $\mathbb{F}_q[T]_d$ of $g_{i,j}$. As the latter has cardinality q^{d-i}, we conclude that $|\mathcal{V}(\mathsf{R}_{i,j})(\mathbb{F}_q)| = q^{d-i}$. Therefore,

$$|\mathcal{V}(\mathsf{R}_i)(\mathbb{F}_q)| = \left| \bigcup_{j=1}^{\lambda_i^*} \mathcal{V}(\mathsf{R}_{i,j})(\mathbb{F}_q) \right| \leq \lambda_i^*\, q^{d-i}.$$

On the other hand, for $i > d$ with $\lambda_i^* > 0$, there is no element of $\mathbb{F}_q[T]_d$ having a non-trivial common factor with g_i defined over $\mathbb{F}_q$. This implies that the set $\mathcal{V}(\mathsf{R}_i)(\mathbb{F}_q)$ is empty, namely

$$|\mathcal{V}(\mathsf{R}_i)(\mathbb{F}_q)| = 0.$$

Now we focuss on the case $i = k$. If $\mathsf{R}_k := \prod_{j=1}^{\lambda_k^*} \mathsf{R}_{k,j}$, we have

$$|\mathcal{V}(\mathsf{R}_k)(\mathbb{F}_q)| \leq |\mathcal{V}(\mathrm{res}(g, F(\boldsymbol{S}, T)))(\mathbb{F}_q)| \leq |\mathcal{V}(\mathsf{R}_k)(\mathbb{F}_q)| + \sum_{i=k+1}^{d} \lambda_i^*\, q^{d-i}.$$

Our previous argument shows that $\mathcal{V}(\mathsf{R}_k)(\mathbb{F}_q)$ is a union of λ_k^* sets $\mathcal{V}(\mathsf{R}_{k,j})(\mathbb{F}_q)$ of cardinality q^{d-k}, which are pairwise distinct. Further, $\boldsymbol{s} \in \mathcal{V}(\mathsf{R}_{k,j_1})(\mathbb{F}_q) \cap \mathcal{V}(\mathsf{R}_{k,j_2})(\mathbb{F}_q)$ for $j_1 \neq j_2$ if and only if both g_{k,j_1} and g_{k,j_2} divide $F(\boldsymbol{s}, T)$. As g_{k,j_1} and g_{k,j_2} are two distinct irreducible elements of $\mathbb{F}_q[T]$, this holds if and only if $g_{k,j_1} \cdot g_{k,j_2}$ divides $F(\boldsymbol{s}, T)$. It follows that

$$|\mathcal{V}(\mathsf{R}_{k,j_1})(\mathbb{F}_q) \cap \mathcal{V}(\mathsf{R}_{k,j_2})(\mathbb{F}_q)| = \begin{cases} q^{d-2k} & \text{for } d \geq 2k, \\ 0 & \text{for } d < 2k. \end{cases}$$

In particular, the Bonferroni inequalities imply

$$\lambda_k^*\, q^{d-k} - \binom{\lambda_k^*}{2} q^{\max\{d-2k,0\}} \leq |\mathcal{V}(\mathsf{R}_k)(\mathbb{F}_q)| = \left| \bigcup_{j=1}^{\lambda_k^*} \mathcal{V}(\mathsf{R}_{k,j})(\mathbb{F}_q) \right| \leq \lambda_k^*\, q^{d-k}.$$

From this the statement of the proposition readily follows. □

https://doi.org/10.1017/S0963548321000274 Published online by Cambridge University Press

As an immediate consequence of Proposition 4.1 we obtain an estimate on the probability that a random element of $\mathbb{F}_q[T]_d$ is relatively prime with g.

Theorem 4.2. *Let e,d be integers with $e > d > 0$. Let g be an element of $\mathbb{F}_q[T]_e$, g^* its square-free part and $(\lambda_1^*, \ldots, \lambda_e^*)$ the factorisation pattern of g^*. Let k be the least integer with $\lambda_k^* > 0$. If $k \leq d$, then the probability $\mathcal{P}_0 := |B_0|/q^d$ that a random element $f \in \mathbb{F}_q[T]_d$ and g are relatively prime is bounded in the following way:*

$$1 - \frac{\lambda_k^*}{q^k} - \sum_{i=k+1}^{d} \frac{\lambda_i^*}{q^i} \leq \mathcal{P}_0 \leq 1 - \frac{\lambda_k^*}{q^k} + \binom{\lambda_k^*}{2} \frac{1}{q^{\min\{2k,d\}}}.$$

In particular, for $q > 2e$ we have $\mathcal{P}_0 > \frac{1}{2}$.

Proof. Observe that

$$|B_0| = \left|\{f \in \mathbb{F}_q[T]_d : \gcd(g,f) = 1\}\right| = \left|\mathbb{F}_q^d \setminus \bigcup_{j=1}^{d} B_j\right| = q^d - \left|\bigcup_{j=1}^{d} B_j\right|.$$

Then the statement readily follows from Proposition 4.1. □

4.2. The average degree of the gcd with g

Now we estimate the cardinality of the sets $\bigcup_{j=i}^{d} B_j$ with $2 \leq i \leq d$. If the square-free part g^* of $g \in \mathbb{F}_q[T]_e$ has a factorisation pattern $(\lambda_1^*, \ldots, \lambda_e^*)$ as in Theorem 4.2, then all its irreducible factors have degree at least k. It follows that $B_1 \cup \cdots \cup B_{k-1}$ is the empty set, which implies the following corollary.

Corollary 4.3. *With hypotheses as in Theorem 4.2, for $1 \leq i \leq k$, we have*

$$\lambda_k^* \, q^{d-k} - \binom{\lambda_k^*}{2} q^{\max\{d-2k,0\}} \leq \left|\bigcup_{j=i}^{d} B_j\right| \leq \lambda_k^* \, q^{d-k} + \sum_{j=k+1}^{d} \lambda_j^* \, q^{d-j}.$$

Next we bound the sum of the cardinalities of $\bigcup_{j=i}^{d} B_j$ for $i \geq k+1$.

Proposition 4.4. *Let $g \in \mathbb{F}_q[T]_e$ have factorisation pattern $(\lambda_1, \ldots, \lambda_e)$ and let k be the least index with $\lambda_k > 0$. We have*

$$\sum_{i=k+1}^{d} \left|\bigcup_{j=i}^{d} B_j\right| \leq \sum_{i=k+1}^{d} (i-k)\, q^{d-i} \sum_{\substack{h_k \leq \lambda_k, \ldots, h_i \leq \lambda_i \\ k\, h_k + \cdots + i\, h_i = i}} \binom{\lambda_k}{h_k} \cdots \binom{\lambda_i}{h_i}.$$

Proof. Observe that

$$B_i \cup \cdots \cup B_d = \{f \in \mathbb{F}_q[T]_d : \deg\ \gcd(g,f) \geq i\}.$$

Fix a factor $m \in \mathbb{F}_q[T]_j$ of degree $j \geq i$ of g. Then the set L_m of multiples $f \in \mathbb{F}_q[T]_d$ of m has cardinality $|L_m| = q^{d-j}$. As a consequence, letting m vary over the set of factors in $\mathbb{F}_q[T]_j$ of g we conclude that

$$\left|\bigcup_{j=i}^{d} B_j\right| \leq \sum_{j=i}^{d} \eta_j \, q^{d-j},$$

where η_j is the number of distinct factors of g in $\mathbb{F}_q[T]_j$ for $i \le j \le d$. It follows that

$$\sum_{i=k+1}^{d} \left| \bigcup_{j=i}^{d} B_j \right| \le \sum_{i=k+1}^{d} (i-k)\, \eta_i\, q^{d-i}.$$

It remains to express the η_i in terms of $\lambda_1, \ldots, \lambda_d$. For this purpose, we observe that

$$\eta_i \le [X^i]\left(\prod_{j=k}^{i} (1+X^j)^{\lambda_j} \right) = \sum_{\substack{h_k \le \lambda_k, \ldots, h_i \le \lambda_i \\ k\, h_k + \cdots + i\, h_i = i}} \binom{\lambda_k}{h_k} \cdots \binom{\lambda_i}{h_i},$$

where $[X^i]f$ denotes the coefficient of X^i in the monomial expansion of $f \in \mathbb{K}[X]$. This proves the proposition. □

Now we obtain an estimate for the average degree of gcd (g, f) for random $f \in \mathbb{F}_q[T]_d$.

Theorem 4.5. *Let e, d be integers with $e > d > 0$, g an element of $\mathbb{F}_q[T]_e$ with factorisation pattern $(\lambda_1, \ldots, \lambda_e)$ and k the least index with $\lambda_k > 0$. Denote by λ_k^* the number of distinct irreducible factors of g in $\mathbb{F}_q[T]_k$. If $k \le d$, then the average degree $\mathbb{E}[\mathcal{X}_g]$ of the greatest common divisor of g and a random element f of $\mathbb{F}_q[T]_d$ is bounded in the following way:*

$$\frac{k\,\lambda_k^*}{q^k} - \binom{\lambda_k^*}{2} \frac{k}{q^{\min\{2k,d\}}} \le \mathbb{E}[\mathcal{X}_g] \le \frac{k\,\lambda_k^*}{q^k} + \sum_{i=k+1}^{d} \frac{i}{q^i} \sum_{\substack{h_k \le \lambda_k, \ldots, h_i \le \lambda_i \\ k\, h_k + \cdots + i\, h_i = i}} \binom{\lambda_k}{h_k} \cdots \binom{\lambda_i}{h_i}.$$

Proof. According to (4.1),

$$\mathbb{E}[\mathcal{X}_g] = \sum_{i=1}^{d} \sum_{j=i}^{d} \frac{|B_j|}{q^d}.$$

By Corollary 4.3, for $1 \le i \le k$,

$$\lambda_k^*\, q^{d-k} - \binom{\lambda_k^*}{2} q^{\max\{d-2k,0\}} \le \left| \bigcup_{j=i}^{d} B_j \right| \le \lambda_k^*\, q^{d-k} + \sum_{i=k+1}^{d} \lambda_i\, q^{d-i}.$$

By Proposition 4.4, we have

$$\sum_{i=k+1}^{d} \left| \bigcup_{j=i}^{d} B_j \right| \le \sum_{i=k+1}^{d} (i-k)\, q^{d-i} \sum_{\substack{h_k \le \lambda_k, \ldots, h_i \le \lambda_i \\ k\, h_k + \cdots + i\, h_i = i}} \binom{\lambda_k}{h_k} \cdots \binom{\lambda_i}{h_i}.$$

We conclude that

$$\begin{aligned} \frac{k\,\lambda_k^*}{q^k} - \binom{\lambda_k^*}{2} \frac{k}{q^{\min\{2k,d\}}} \le \mathbb{E}[\mathcal{X}_g] &\le \frac{k\,\lambda_k^*}{q^k} + \sum_{i=k+1}^{d} \frac{k\,\lambda_i}{q^i} \\ &\quad + \sum_{i=k+1}^{d} \frac{i-k}{q^i} \sum_{\substack{h_k \le \lambda_k, \ldots, h_i \le \lambda_i \\ k\, h_k + \cdots + i\, h_i = i}} \binom{\lambda_k}{h_k} \cdots \binom{\lambda_i}{h_i} \\ &\le \frac{k\,\lambda_k^*}{q^k} + \sum_{i=k+1}^{d} \frac{i}{q^i} \sum_{\substack{h_k \le \lambda_k, \ldots, h_i \le \lambda_i \\ k\, h_k + \cdots + i\, h_i = i}} \binom{\lambda_k}{h_k} \cdots \binom{\lambda_i}{h_i}, \end{aligned}$$

which proves the theorem. □

https://doi.org/10.1017/S0963548321000274 Published online by Cambridge University Press

To simplify the upper bound of Theorem 4.5 we recall that the inner sum in such an upper bound is actually an upper bound for the number η_i of distinct factors of g in $\mathbb{F}_q[T]_i$, namely

$$\eta_i \leq [X^i]\left(\prod_{j=k}^{i}(1+X^j)^{\lambda_j}\right) = \sum_{\substack{h_k\leq\lambda_k,\dots,h_i\leq\lambda_i\\ k\,h_k+\cdots+i\,h_i=i}} \binom{\lambda_k}{h_k}\cdots\binom{\lambda_i}{h_i},$$

with equality when g is square-free. Using the generalised Vandermonde identity (see, e.g. [7, page 248]), we have

$$\eta_i \leq \sum_{\substack{h_k\leq\lambda_k,\dots,h_i\leq\lambda_i\\ k\,h_k+\cdots+i\,h_i=i}} \binom{k\,\lambda_k}{k\,h_k}\cdots\binom{i\,\lambda_i}{i\,h_i} \leq \binom{k\,\lambda_k+\cdots+i\,\lambda_i}{i}. \tag{4.2}$$

On the other hand, taking into account that the expansion of the analytic function $h:\mathbb{C}\to\mathbb{C}$, $h(z) := \prod_{j=k}^{i}(1+z^j)^{\lambda_j}$ has non-negative coefficients at 0, from, e.g. [4, Proposition IV.1] we conclude that

$$\eta_i \leq h(1) = 2^{\lambda_k+\cdots+\lambda_i}. \tag{4.3}$$

The accuracy of (4.2) and (4.3) depends on the actual factorisation pattern $(\lambda_1,\dots,\lambda_e)$. For example, if $g\in\mathbb{F}_q[T]_e$ is a polynomial with an 'equal-degree factorization' (that is, $k\lambda_k = e$), then for large k the bound (4.3) is preferable, while for large λ_k the bound (4.2) is more accurate.

Finally, for the results on the average-case complexity of the Euclidean algorithm we shall use a further upper bound on $\mathbb{E}[\mathcal{X}_g]$. This bound, although not as precise as the one of Theorem 4.5, has a simple expression which suffices for the purposes of the next section.

Lemma 4.6. *Let e, d be integers with $e > d > 0$ and let $g\in\mathbb{F}_q[T]_e$. Then*

$$\mathbb{E}[\mathcal{X}_g] \leq \frac{de}{q^k}.$$

Proof. Let $(\lambda_1,\dots,\lambda_e)$ be the factorisation pattern of g and let k be the least index with $\lambda_k > 0$. By Proposition 4.1,

$$\mathbb{E}[\mathcal{X}_g] = \sum_{k=1}^{d}\sum_{j=k}^{d}\frac{|B_j|}{q^d} \leq \sum_{k=1}^{d}\sum_{j=1}^{d}\frac{|B_j|}{q^d} \leq d\left(\frac{\lambda_k}{q^k}+\sum_{i=k+1}^{d}\frac{\lambda_i}{q^i}\right) \leq \frac{de}{q^k}.$$

□

5. Average-case analysis of the Euclidean Algorithm

Let e, d be positive integers with $q > d(2e-d+1)/2$ and $e > d$ and let $g\in\mathbb{F}_q[T]_e$ be fixed. In this section we analyse the average-case complexity of the Euclidean algorithm applied to pairs (g,f) with $f\in\mathbb{F}_q[T]_d$, and prove Theorem 1.3.

Given positive integers m, n with $m > n$ and $(f_1,f_2)\in\mathbb{F}_q[T]_m\times(\mathbb{F}_q[T]_n\setminus\{0\})$, and an arithmetic operation $\mathsf{w}\in\{\div,-,\times\}$, by $d^{\mathsf{w}}(f_1,f_2)$ we denote the number of operations w used in the 'synthetic' polynomial division algorithm applied to (f_1,f_2) (see, e.g. [10]). It turns out that

$$d^{\div}(f_1,f_2) = m-n+1, \quad d^{-,\times}(f_1,f_2) = n(m-n+1). \tag{5.1}$$

Endowing $\mathbb{F}_q[T]_d$ with the uniform probability, for any $\mathsf{w}\in\{\div,-,\times\}$ we consider the random variable $t_g^{\mathsf{w}}:\mathbb{F}_q[T]_d\to\mathbb{N}$ which counts the number of operations w that the Euclidean Algorithm

https://doi.org/10.1017/S0963548321000274 Published online by Cambridge University Press

performs on input (g,f) for each $f \in \mathbb{F}_q[T]_d$. Furthermore, $t_g^{\mathsf{div}}(f)$ denotes the number of polynomial divisions involved. Our aim is to study the expected value $\mathbb{E}[t_g^{\mathsf{w}}]$ of t_g^{w} for $\mathsf{w} \in \{\div, \mathsf{div}, -, \times\}$, namely

$$\mathbb{E}[t_g^{\mathsf{w}}] = \frac{1}{q^d} \sum_{f \in \mathbb{F}_q[T]_d} t_g^{\mathsf{w}}(f) = \frac{1}{q^d} \sum_{k=0}^{d} \sum_{f \in B_{d-k}} t_g^{\mathsf{w}}(f).$$

As explained before, applying the Euclidean algorithm to an input (g,f) with $f \in \mathbb{F}_q[T]_d$ we obtain a unique polynomial quotient sequence $(q_1 \ldots q_{h+1})$ and a unique polynomial remainder sequence $(r_1 \ldots r_h)$ satisfying the following conditions:

$$\begin{aligned} g &= f \cdot q_1 + r_1, & \deg(r_1) &< \deg(f), \\ f &= r_1 \cdot q_2 + r_2, & \deg(r_2) &< \deg(r_1), \\ &\vdots & &\vdots \\ r_{h-2} &= r_{h-1} \cdot q_h + r_h, & \deg(r_h) &< \deg(r_{h-1}), \\ r_{h-1} &= r_h \cdot q_{h+1}. \end{aligned} \tag{5.2}$$

The number of polynomial divisions is h, and (5.1) implies that the number of arithmetic operations performed is uniquely determined by the sequence $(\deg(r_1), \ldots, \deg(r_h))$. The upper bounds for $\mathbb{E}[t_g^{\mathsf{w}}]$ are obtained by considering sequences of maximum length for the $f \in B_{d-k}$ for each k, combined with the upper bound of Lemma 4.2 for the average degree of the gcd. On the other hand, for the lower bounds we consider the behaviour of the Euclidean algorithm on inputs f which are generic with respect to g. For these f, the corresponding number of operations is precisely determined, and Proposition 3.2 shows that its number is large enough to yield significant conclusions.

5.1. The estimates on the average-case complexity

In the next lemmas, we estimate the average number of polynomial divisions, divisions in $\mathbb{F}_q$, and the remaining arithmetic operations in $\mathbb{F}_q$, performed by the Euclidean algorithm (5.2). We start with polynomial divisions.

Lemma 5.1. *The average number $\mathbb{E}[t_g^{\mathsf{div}}]$ of polynomial divisions performed by the Euclidean algorithm applied to pairs (g,f) with $f \in \mathbb{F}_q[T]_d$ is bounded as follows:*

$$(d+1)\left(1 - \frac{d(2e-d+1)}{2q}\right) \leq \mathbb{E}[t_g^{\mathsf{div}}] \leq (d+1)\left(1 + \frac{de}{q}\right).$$

Proof. For $f \in B_{d-k}$ with $0 \leq k \leq d$, we claim that $t_g^{\mathsf{div}}(f) \leq k+1$. Indeed, the maximum number of polynomial divisions in (5.2) is achieved from a sequence of remainders of maximum length. Since $f \in B_{d-k}$, in such a sequence the degree of each successive remainder decreases by 1, that is, the sequence has length $h = k$. Taking into account that there is a further division to perform, to check that r_h divides r_{h-1}, we deduce our claim. As $k \mapsto \frac{k+1}{d-k}$ is an increasing function for $k \in [0, d-1]$, we obtain

https://doi.org/10.1017/S0963548321000274 Published online by Cambridge University Press

$$\mathbb{E}[t_g^{\mathsf{div}}] \leq \frac{1}{q^d}\sum_{k=0}^{d}\sum_{f\in B_{d-k}}(k+1) = \frac{1}{q^d}\left(\sum_{k=0}^{d-1}\frac{k+1}{d-k}(d-k)|B_{d-k}| + (d+1)|B_0|\right)$$
$$\leq \frac{d}{q^d}\sum_{k=0}^{d-1}(d-k)|B_{d-k}| + (d+1)\frac{|B_0|}{q^d} \leq d\,\mathbb{E}[\mathcal{X}_g] + d + 1.$$

Using the bound $\mathbb{E}[\mathcal{X}_g] \leq de/q$ of Lemma 4.6, we deduce the upper bound in the statement of the lemma.

Next we show the lower bound. Recall that $f \in \mathbb{F}_q[T]_d$ is generic (with respect to g) if the corresponding remainder sequence is of the form $(r_1, \ldots, r_d)$, where $\deg(r_j) = d - j$ for $1 \leq j \leq d$. For such an f, the number of polynomial divisions is precisely $d+1$. By Proposition 3.3, it follows that

$$\mathbb{E}[t_g^{\mathsf{div}}] \geq \frac{1}{q^d}(d+1)|\mathcal{G}| \geq (d+1)\left(1 - \frac{d(2e-d+1)}{2q}\right).$$

This finishes the proof of the lemma. □

Next we analyse the case $\mathsf{w} = \div$.

Lemma 5.2. *Denote by $\mathbb{E}[t_g^{\div}]$ the average number of divisions performed by the Euclidean algorithm applied to pairs (g,f) with $f \in \mathbb{F}_q[T]_d$. Then*

$$(e+d+1)\left(1 - \frac{d(2e-d+1)}{2q}\right) \leq \mathbb{E}[t_g^{\div}] \leq (e+d+1)\left(1 + \frac{de}{q}\right).$$

Proof. Let $f \in B_{d-k}$ with $0 \leq k \leq d$. According to (5.1), the number of operations $\div$ in each step of (5.2) is

$$\begin{aligned} d^{\div}(g,f) &= \deg(g) - \deg(f) + 1, \\ d^{\div}(f,r_1) &= \deg(f) - \deg(r_1) + 1, \\ &\vdots \\ d^{\div}(r_{h-1},r_h), &= \deg(r_{h-1}) - \deg(r_h) + 1. \end{aligned}$$

Therefore,

$$t_g^{\div}(f) = \deg(g) - \deg(r_h) + h + 1 = e - (d-k) + h + 1 \leq e - d + 2k + 1. \tag{5.3}$$

As $k \mapsto \frac{e-d+2k+1}{d-k}$ is increasing for $k \in [0, d-1]$, from (5.3) we deduce that

$$\mathbb{E}[t_g^{\div}] = \frac{1}{q^d}\sum_{k=0}^{d}\sum_{f\in B_{d-k}} t_g^{\div}(f) \leq \frac{1}{q^d}\sum_{k=0}^{d-1}\frac{e-d+2k+1}{d-k}(d-k)|B_{d-k}| + (e+d+1)\frac{|B_0|}{q^d}$$
$$\leq (e+d-1)\mathbb{E}[\mathcal{X}_g] + e + d + 1.$$

Combining this with Lemma 4.6 readily implies the upper bound.

To prove the lower bound, we argue as in the proof of Lemma 5.1. For a generic $f \in \mathcal{G}$, the remainder sequence is of length d, and therefore $t_g^{\div}(f) = e + d + 1$. It follows that

$$\mathbb{E}[t_g^{\div}] \geq \frac{1}{q^d}(e+d+1)|\mathcal{G}| \geq (e+d+1)\left(1 - \frac{d(2e-d+1)}{2q}\right).$$

This proves the lemma. □

https://doi.org/10.1017/S0963548321000274 Published online by Cambridge University Press

Finally, we consider the remaining case $\mathsf{w} \in \{-, \times\}$. We have the following result.

Lemma 5.3. *Let* $\mathbb{E}[t_g^{-,\times}]$ *be the average number of operations* $\mathsf{w} \in \{-, \times\}$ *performed by the Euclidean algorithm applied to pairs (g,f) with* $f \in \mathbb{F}_q[T]_d$. *Then*

$$de\left(1 - \frac{d(2e-d+1)}{2q}\right) \leq \mathbb{E}[t_g^{-,\times}] \leq de\left(1 + \frac{de}{q}\right).$$

Proof. For $f \in B_{d-k}$ with $0 \leq k \leq d$, by (5.1) the number of operations d^{w} with $\mathsf{w} \in \{-, \times\}$ in each step of (5.2) is

$$\begin{aligned} d^{\mathsf{w}}(g,f) &= \deg(f)(\deg(g) - \deg(f) + 1), \\ d^{\mathsf{w}}(f, r_1) &= \deg(r_1)(\deg(f) - \deg(r_1) + 1), \\ &\vdots \\ d^{\mathsf{w}}(r_{h-1}, r_h) &= \deg(r_h)(\deg(r_{h-1}) - \deg(r_h) + 1). \end{aligned}$$

Denote $r_0 := f$. We claim that the maximum number of operations w performed in the whole Euclidean algorithm is achieved with a sequence of remainders $(r_0, \ldots, r_k)$ with $\deg(r_{j-1}) - \deg(r_j) = 1$ for $1 \leq j \leq k$. Indeed, let $(r_0, \ldots, r_h)$ be a remainder sequence such that $\deg(r_{j-1}) - \deg(r_j) > 1$ for a given j. Denote by $(\alpha_0, \ldots, \alpha_h)$ the corresponding sequence of degrees. We compare the number of operations w performed by the Euclidean algorithm to obtain this sequence with that of a remainder sequence with degree pattern $(\alpha_0, \ldots, \alpha_{j-1}, \alpha_j^*, \alpha_j, \ldots, \alpha_h)$, where $\alpha_{j-1} - \alpha_j^* = 1$. Since the number of w operations is determined by the degree pattern of the remainder sequence under consideration, it suffices to compare the cost of the jth step of the first sequence with the sum of those of the jth and $(j+1)$th steps of the second sequence. In particular, we see that our claim for this case holds provided that

$$\alpha_j\big((\alpha_{j-1} - \alpha_j) + 1\big) \leq \alpha_j^*(\alpha_{j-1} - \alpha_j^* + 1) + \alpha_j(\alpha_j^* - \alpha_j + 1).$$

This can be checked by an easy calculation. Arguing successively in this way, the claim follows.

As a consequence, the maximum number of operations w performed is achieved in a sequence of k remainders $(r_1, \ldots, r_k)$ with $\deg(r_{j-1}) - \deg(r_j) = 1$ for $1 \leq j \leq k$, namely with $\deg(r_j) = d - j$ for $1 \leq j \leq k$. It follows that

$$\begin{aligned} t_g^{\mathsf{w}}(f) &\leq \deg(f)(\deg(g) - \deg(f) + 1) + \sum_{j=1}^{k} \deg(r_j)(\deg(r_{j-1}) - \deg(r_j) + 1) \\ &= d(e-d+1) + 2\sum_{j=1}^{k}(d-j) = d(e-d+1) + k(2d-k-1). \end{aligned} \tag{5.4}$$

Since $k \mapsto \frac{d(e-d+1)+k(2d-k-1)}{d-k}$ is increasing for $k \in [0, d-1]$, by (5.4) we obtain

$$\begin{aligned} \mathbb{E}[t_g^{\mathsf{w}}] &= \frac{1}{q^d} \sum_{k=0}^{d} \sum_{f \in B_{d-k}} t_g^{\mathsf{w}}(f) \\ &\leq \frac{1}{q^d} \sum_{k=0}^{d-1} \frac{d(e-d+1) + k(2d-k-1)}{d-k}(d-k)|B_{d-k}| + de\frac{|B_0|}{q^d} \\ &\leq de\,\mathbb{E}[\mathcal{X}_g] + de. \end{aligned}$$

The upper bound follows easily by Lemma 4.6.

https://doi.org/10.1017/S0963548321000274 Published online by Cambridge University Press

On the other hand, for $f \in \mathcal{G}$, by (5.4) we conclude that $t_g^{\mathsf{w}}(f) = de$. Then Proposition 3.3 implies

$$\mathbb{E}[t_g^{\div}] \geq \frac{1}{q^d} de|\mathcal{G}| \geq de\left(1 - \frac{d(2e-d+1)}{2q}\right),$$

which finishes the proof of the lemma. □

Summarising Lemmas 5.1, 5.2 and 5.3, we have the following result.

Theorem 5.4. *Let e, d be positive integers such that $q > d(2e-d+1)/2$ and $e > d$. Let $g \in \mathbb{F}_q[T]_e$ and $\mathsf{w} \in \{\div, \mathsf{div}, -, \times\}$. The average number $\mathbb{E}[t_g^{\mathsf{w}}]$ of operations w performed on (uniform distributed) inputs from $\mathbb{F}_q[T]_d$ is bounded in the following way:*

$$\left|\frac{\mathbb{E}[t_g^{\mathsf{div}}]}{d+1} - 1\right| \leq \frac{de}{q}, \qquad \left|\frac{\mathbb{E}[t_g^{\div}]}{e+d+1} - 1\right| \leq \frac{de}{q}, \qquad \left|\frac{\mathbb{E}[t_g^{-,\times}]}{de} - 1\right| \leq \frac{de}{q}.$$

6. Conclusions and perspectives

We have developed an average-case analysis of the Euclidean algorithm applied to pairs of polynomials $(g,f) \in \mathbb{F}_q[T]_e \times \mathbb{F}_q[T]_d$ with $e > d$, where g is fixed. Our results show that, on average, the behaviour of the Euclidean algorithm in this case mimics that of the general case where g and f vary over the set $\mathbb{F}_q[T]_e \times \mathbb{F}_q[T]_d$. For this purpose, we have shown that, for fixed g, on 'most' pairs (g,f) as above the Euclidean algorithm performs the highest possible number of steps, namely d. This is shown by comparing the 'formal' execution of the Euclidean algorithm on polynomials whose roots are indeterminates with its actual execution on g and a monic polynomial f of degree d.

For given values of q, e and d with $e > d$, we performed some simulations where we executed the Euclidean algorithm on pairs (g,f), where $g \in \mathbb{F}_q[T]_e$ was fixed with a given factorisation pattern and f ran through a random sample $\mathcal{S} \subset \mathbb{F}_q[T]_d$. The aim was to analyse to what extent the estimates on the error terms underlying the asymptotic main term on the average degree of gcd (g,f) (Theorem 4.5), the probability that gcd $(g,f) = 1$ (Theorem 4.2) and the probability that a random $f \in \mathbb{F}_q[T]_d$ is 'generic' with respect to g (Proposition 3.3) were accurate. The numerical experiments we performed suggest that the estimates of Theorems 4.5 and 4.2 are rather accurate. On the other hand, it seems that the estimate on the number of polynomials which are generic with respect to g is somewhat pessimistic. Our numerical experiments suggest that this number depends on the factorisation pattern of g, while the lower bound of Proposition 3.3 depends only on q, e and d. In this sense, we wonder whether such an estimate can be achieved.

Finally, a critical step in our methodology consists in comparing the 'formal' execution of the Euclidean algorithm with its execution on pairs (g,f) as above. The set of input pairs on which the Euclidean algorithm does not behave as expected are expressed in terms of the set of $\mathbb{F}_q$-rational zeros of certain multivariate polynomials. We think that our methodology might be applied to the analysis of other algorithms, such as the Berlekamp–Massey algorithm for linear feedback shift-register synthesis, whose similarity with the Euclidean algorithm for decoding has already been established in the literature (see, e.g. [9]).

Acknowledgements

The authors wish to thank the anonymous referees for several suggestions, which helped to improve the presentation of the results of this paper.

References

[1] Basu, S., Pollack, R. and Roy, M.-F. (2006) *Algorithms in Real Algebraic Geometry*, 2nd ed. Algorithms Comput. Math., Vol. **10**, Springer.

[2] Berthé, V., Nakada, H., Natsui, R. and Vallée, B. (2014) Fine costs for Euclid's algorithm on polynomials and Farey maps. *Adv. Appl. Math.* **54** 27–65.

[3] Cafure, A. and Matera, G. (2006) Improved explicit estimates on the number of solutions of equations over a finite field. *Finite Fields Appl.* **12** 155–185.

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

[5] Fulton, W. (1984) *Intersection Theory*. Springer.

[6] von zur Gathen, J. and Gerhard, J. (1999) *Modern Computer Algebra*. Cambridge University Press.

[7] Graham, R., Knuth, D. and Patashnik, O.(1994) *Concrete Mathematics: A Foundation for Computer Science*, 2nd ed. Addison–Wesley.

[8] Heintz, J. (1983) Definability and fast quantifier elimination in algebraically closed fields. *Theoret. Comput. Sci.* **24** 239–277.

[9] Heydtmann, A. and Jensen, J. (2000) On the equivalence of the Berlekamp–Massey and the Euclidean algorithms for decoding. *IEEE Trans. Inform. Theory* 46 2614–2624.

[10] Knuth, D. E. (1981) *The Art of Computer Programming II: Semi–Numerical Algorithms*, Vol. **2**. Addison-Wesley.

[11] Lascoux, A. (2003) *Symmetric Functions and Combinatorial Operators on Polynomials*. CBMS Reg. Conf. Ser. *Math.*, Vol. **99**, American Mathematical Society.

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

[13] Ma, K. and von zur Gathen, J. (1990) Analysis of Euclidean algorithms for polynomials over finite fields. *J. Symb. Comput.* 9 429–455.

[14] Norton, G. (1989) Precise analyses of the right- and left-shift greatest common divisor algorithms for GF(q)[x]. *SIAM J. Comput.* **18** 608–624.

Cite this article: Giménez N, Matera G, Pérez M and Privitelli M (2022). Average-case complexity of the Euclidean algorithm with a fixed polynomial over a finite field. *Combinatorics, Probability and Computing* **31**, 166–183. https://doi.org/10.1017/S0963548321000274