The complexity of solving a random polynomial system*

Giulia Gaggero and Elisa Gorla

1 Introduction

Randomness plays a fundamental role within cryptography. For example, it plays a pivotal role within key generation and cryptographic algorithms are subject to randomness tests. In multivariate cryptography, the public key consists of a multivariate polynomial system and the goal of the attacker is to find a solution of the system. In this context, one typically wishes for the system be as close as possible to, or at least appear as, a random system. A random system is expected to be hard to solve, since the Multivariate Quadratic Problem is not only NP-complete, but also known to be hard to solve on average for a wide range of parameters. From this point of view, e.g., a digital signature scheme whose public keys are sufficiently random is expected to be secure.

In this paper, we discuss what it means for a polynomial system to be random and how hard it is to solve a random polynomial system. In Definition 7 we propose a mathematical formulation for the concept of random system. The definition of randomness that we propose, which we call algebraic randomness, is broad enough to include a vast majority of the systems which are of interest in cryptography. We then specify our definition further in Definition 12. One advantage of this definition is that the property of being random according to Definition 12 can be computationally tested, at least in principle.

In Theorem 22, Corollary 26, Corollary 28, Theorem 34, and Proposition 32, we prove upper bounds for the degree of regularity and the solving degree of an algebraically random polynomial system, depending on parameters of the system such as the number of equations, the number of variables, and the degrees of the equations. The usefulness of our bounds is twofold: On the one side, our bounds can be used to directly produce bounds on the complexity of computing a Gröbner basis, hence of solving, many systems which are of interest in cryptography. Bounds on the complexity produced in this way have the advantage of being widely applicable and the disadvantage of not always being close to the actual complexity for each system to which they apply. On the other side, our bounds give us an idea of what security one can hope to achieve for a system with given parameters. Therefore, our bounds can be used as a point of comparison for the optimality of a given public key, in the following sense. Say that, in order to forge a signature produced with a given multivariate digital signature scheme, one has to find a solution of a system of m equations of degree D in n variables. Suppose that such a system is algebraically random. Our results provide an upper bound B for the degree of regularity or the solving degree of such a system. Say that one can compute or estimate by a different method the degree of regularity or the solving degree of the specific system and suppose that this turns out to be C. Clearly, it must always be that $C \leq B$. However, how far C is from B gives us a measure of how close to optimal the digital signature scheme is, for the given choice of parameters. In fact, since our bounds on the degree of regularity are sharp, there are systems of m equations of degree D in n variables whose degree of regularity is exactly B. In other words, if B and C are close, then there is not much space for improvement, since any system with the same parameters as our public key can have degree of regularity or solving degree at most B. If on the contrary B and C are far apart, then potentially there is a lot of space for finding a more robust system with the same parameters, since a system with those parameters can have degree of regularity or solving degree up to B.

The paper is structured as follows. Section 2 contains some useful preliminaries. In Section 3 we discuss the concept of randomness and propose two definitions of randomness for a polynomial system, in Definition 7 and Definition 12. Throughout the paper, we refer to our notion of

*This research was funded by armasuisse through grant no. CYD-C-2020010.

randomness as algebraic randomness. We briefly review cryptographic semiregular sequences and semiregular sequences and show that the latter are algebraically random. In Section 4, after recalling some results from commutative algebra, we prove some bounds on the degree of regularity and the solving degree of an algebraically random system consisting of polynomials the same degree. The bounds appear in Theorem 22, Corollary 26, Corollary 28, Theorem 34, and Proposition 32. Finally, in Section 5 we show how to apply our results to check that the systems associated to GeMSS and Rainbow are not algebraically random.

2 Preliminaries

Let $\mathbb{F}_q$ be the finite field of cardinality q and $R = \mathbb{F}_q[x_1, \ldots, x_n]$ be the polynomial ring over $\mathbb{F}_q$ in n variables. We start by recalling how a typical multivariate one-way function is constructed. Let $f_1, \ldots, f_m \in R$, and consider the evaluation map

$$\begin{array}{rccc} \mathcal{F}: & \mathbb{F}_q^n & \to & \mathbb{F}_q^m \\ & \alpha = (\alpha_1, \ldots, \alpha_n) & \mapsto & (f_1(\alpha_1, \ldots, \alpha_n), \ldots, f_m(\alpha_1, \ldots, \alpha_n)) \end{array}$$

To hide the structure of $\mathcal{F}$, we compose it with two random invertible linear maps $S : \mathbb{F}_q^n \to \mathbb{F}_q^n$ and $T : \mathbb{F}_q^m \to \mathbb{F}_q^m$. We obtain $\mathcal{P} = T \circ \mathcal{F} \circ S$, a set of m polynomials $p_1, \ldots, p_m$ in n variables over $\mathbb{F}_q$. The public key of the multivariate scheme is $\mathcal{P} = (p_1, \ldots, p_m)$ and the private key is $\{\mathcal{F}, S, T\}$. The trapdoor consists of constructing $\mathcal{F}$ such that $\mathcal{F}^{-1}$ is efficiently computable. Notice that $\mathcal{P}$ should be hard to invert without the knowledge of S, T, in particular it should be hard to recover the structure of $\mathcal{F}$ from $\mathcal{P}$.

Finding inverse images with respect to $\mathcal{P}$ corresponds to computing solutions of multivariate polynomial systems. It is well-known that this problem can be solved by computing a Gröbner basis of the system, we refer the interested reader to [13] for more detail. The first algorithm for computing Gröbner bases appeared in the doctoral thesis of Buchberger [11]. Modern algorithms for computing Gröbner bases are based on liner algebra and are more efficient than Buchberger's. Examples of linear-algebra-based algorithms are F_4 [20], F_5 [21], the *XL Algorithm* [18], *MutantXL* [12], and their variants. In all of these systems, one computes the reduced row echelon form of the Macaulay matrix associated to the polynomial equations in a given degree, for one or more degrees.

We now describe the main object of linear-algebra-based algorithms [6, Section 1.2]. Fix a term order on R and let $\mathcal{F} = \{f_1, \ldots, f_m\} \subseteq R$ be a system of homogeneous polynomials. The columns of the **homogeneous Macaulay matrix** M_d of $\mathcal{F}$ are labelled by the monomials of R_d and arranged in decreasing order. The rows of M_d are labelled by polynomials of the form $m_{i,j} f_j$, where $m_{i,j} \in R$ is a monomial such that $\deg(m_{i,j} f_j) = d$. The entry (i, j) of M_d is the coefficient of the monomial of column j within the polynomial corresponding to the i-th row.

Let now $f_1, \ldots, f_m$ be arbitrary (not necessarily homogeneous) polynomials. The columns of the **Macaulay matrix** $M_{\leq d}$ of $\mathcal{F}$ are labelled by the monomials of R of degree $\leq d$, arranged in decreasing order. The rows of $M_{\leq d}$ correspond to polynomials of the form $m_{i,j} f_j$, where $m_{i,j} \in R$ is a monomial such that $\deg(m_{i,j} f_j) \leq d$. The entries of $M_{\leq d}$ are defined as in the homogeneous case. The rationale behind the use of homogeneous Macaulay matrices for homogeneous systems is that, for a homogeneous system, the Macaulay matrix $M_{\leq d}$ is a block matrix with blocks $M_d, \ldots, M_0$ as blocks.

One computes the reduced row echelon form of the Macaulay matrix, or of its homogeneous version, in one or more degrees. For large enough degree, this produces a reduced Gröbner basis with respect to the chosen order. Some algorithms, as e.g. *MutantXL*, use a variation called **mutant strategy** in the non-homogeneous case: If the reduction of the Macaulay matrix $M_{\leq d}$ produces new polynomials $g_1, \ldots, g_\ell$ of degree strictly smaller than d, one appends to the reduction of $M_{\leq d}$ the polynomials $m_{i,j} g_j$, where $m_{i,j} \in R$ is a monomial such that $\deg(m_{i,j} g_j) \leq d$, then computes the reduced row echelon form again. Throughout the paper, we refer to the algorithms that employ the mutant strategy as **mutant algorithms** and to the others as **standard algorithms**. For a description of the basic version of these algorithms, we refer to [13, Section 3.1] and [14, Section 1].

The computational complexity of computing the reduced row echelon form of the Macaulay matrices $M_{\leq d}$ and M_d depends on their size, and hence on the degree d. This motivates the next definition.

Definition 1. *Let $\mathcal{F} = \{f_1, \dots, f_m\} \subseteq R$ and let τ be a term order on R. The* **solving degree** *of $\mathcal{F}$ with respect to τ is the least degree d such that Gaussian elimination on the Macaulay matrix $M_{\leq d}$ produces a τ-Gröbner basis of $\mathcal{F}$. We denote by $\operatorname{solv.deg}^s_\tau(\mathcal{F})$ the solving degree of $\mathcal{F}$ with respect to a standard algorithm and by $\operatorname{solv.deg}^m_\tau(\mathcal{F})$ the solving degree of $\mathcal{F}$ with respect to a mutant algorithm. When τ is the degree reverse lexicographic order, we omit the subscript τ.*

In general, the solving degree is not invariant under coordinate change. Moreover, it may depend on the algorithm used to perform the Gröbner basis computation. In particular, for mutant algorithms it may be smaller than for standard ones. Finally, the solving degree depends on the choice of a term order on R.

The complexity of linear-algebra-based algorithms is dominated by the cost of computing a degree reverse lexicographic Gröbner basis of the system, see [13, Sections 2 and 3] for more detail. Therefore, an upper bound on the solving degree with respect to the reverse lexicographic order yields an upper bound on the complexity of computing a lexicographic Gröbner basis, hence on the complexity of solving the polynomial system.

Let I be a homogeneous ideal of R. For an integer $d \geq 0$, we denote by $I_d = I \cap R_d$ the $\mathbb{F}$-vector space of homogeneous polynomials of degree d in I. For $g \in R$ a polynomial, we denote by g^{top} the homogeneous part of g of largest degree. E.g., if $g = x^3 + 2xy^2 - y + 1 \in \mathbb{F}[x, y]$, then $g^{\text{top}} = x^3 + 2xy^2$. For a polynomial system $\mathcal{F} = \{f_1, \dots, f_m\} \subseteq R$, we denote by $\mathcal{F}^{\text{top}} \subseteq R$ the homogeneous system $\{f_1^{\text{top}}, \dots, f_m^{\text{top}}\}$. Up to doing Gaussian elimination in a matrix whose rows correspond to $f_1, \dots, f_m$, we may suppose that $f_1^{\text{top}}, \dots, f_m^{\text{top}}$ are linearly independent. We will assume this throughout the paper.

The **degree of regularity** was introduced in [2, 5].

Definition 2. *Let $\mathcal{F} \subseteq R$ be a polynomial system. The* **degree of regularity** *of $\mathcal{F}$ is*

$$d_{\text{reg}}(\mathcal{F}) = \begin{cases} \min\{d \geq 0 \mid (\mathcal{F}^{\text{top}})_d = R_d\} & \text{if } (\mathcal{F}^{\text{top}})_d = R_d \text{ for } d \gg 0 \\ +\infty & \text{otherwise.} \end{cases}$$

In the cryptographic literature, the degree of regularity is often used as a proxy for the solving degree. This is the case, e.g., in the specification documents of GeMSS [15]. However, this does not always produce reliable estimates. In fact, there are examples in which the gap between the degree of regularity and the solving degree is large, see e.g. [10, Examples 3.2 and 3.3]. A recent result by Semaev and Tenti [27, 28] however shows that, under suitable assumptions, the solving degree of a system with respect to a standard algorithm is at most twice the degree of regularity. Thanks to this result, an upper bound for the degree of regularity yields a proven upper bound for the solving degree.

Theorem 3 ([28, Corollary 3.67] and [27, Theorem 2.1])**.** *Let $\mathbb{F} = \mathbb{F}_q$. Let $\mathcal{F} = \{f_1, \dots, f_m, x_1^q - x_1, \dots, x_n^q - x_n\} \subseteq R$ be a polynomial system. If $d_{\text{reg}}(\mathcal{F}) \geq \max\{q, \deg(f_1), \dots, \deg(f_m)\}$, then for standard algorithms*

$$\operatorname{solv.deg}(\mathcal{F}) \leq 2d_{\text{reg}}(\mathcal{F}) - 2.$$

Notice that, in almost all systems of cryptographic interest, the field size and the degrees of the polynomials are relatively small. Therefore, one expects that Theorem 3 applies to such systems.

Another recent result by Salizzoni [26] shows that the solving degree of a mutant algorithm is at most the degree of regularity plus one, unless the system contains polynomials of large degree.

Theorem 4 ([26, Proposition 3.10])**.** *Let $\mathcal{F} = \{f_1, \dots, f_m\} \subseteq R$ be a polynomial system, then for mutant algorithms*

$$\operatorname{solv.deg}(\mathcal{F}) \leq \max\{d_{\text{reg}}(\mathcal{F}) + 1, \deg(f_1), \dots, \deg(f_m)\}.$$

Another algebraic invariant connected to the solving degree of a polynomial system is the Castelnuovo-Mumford regularity. We refer the reader to [13, Section 3.4] and [14] for its definition and a discussion on its relation with the other invariants.

Remark 5. *If $\mathbb{F} = \mathbb{F}_q$ and the system $\mathcal{F}$ contains the field equations $x_1^q - x_1, \dots, x_n^q - x_n$, then $(\mathcal{F}^{\text{top}})_d = R_d$ for $d \gg 0$, therefore the degree of regularity is finite.*

Remark 6. *If* $(\mathcal{F}^{\text{top}})_d = R_d$ *for* $d \gg 0$*, then*

$$d_{\text{reg}}(\mathcal{F}) = \text{reg}(\mathcal{F}^{\text{top}}).$$

Therefore:

- *By Theorem 4*

$$\text{solv.deg}(\mathcal{F}) \leq \max\{\text{reg}(\mathcal{F}^{\text{top}}) + 1, \deg(f_1), \ldots, \deg(f_m)\}$$

  for mutant algorithms.

- *If* $d_{\text{reg}}(\mathcal{F}) \geq \max\{q, \deg(f_1), \ldots, \deg(f_m)\}$*, then by Theorem 3*

$$\text{solv.deg}(\mathcal{F}) \leq 2\,\text{reg}(\mathcal{F}^{\text{top}}) - 2$$

  for standard algorithms.

3 Random polynomial systems

Consider a system of m equations of degrees $d_1, \ldots, d_m$ in n variables. Over a finite field, one often defines a random system as a polynomial system whose coefficients are chosen uniformly at random in the given field. In this paper, we propose a different definition which still captures the intuitive idea of randomness, while allowing us to estimate the degree of regularity of a random system.

Over an infinite field, one may use the concept of genericity from algebraic geometry to define randomness. More precisely, fix a nonempty Zarisky-open subset of $\mathbb{P}^{\binom{n+d_1}{n}-1} \times \ldots \times \mathbb{P}^{\binom{n+d_m}{n}-1}$, where $\mathbb{P}^t$ denotes t-dimensional projective space. Define a random system as an element of that open set. This makes sense, since every nonempty Zarisky-open set is dense, hence a system of m equations of degrees $d_1, \ldots, d_m$ in n variables whose coefficients are chosen uniformly at random is generic with high probability according to this definition. The problem of extending this definition to a finite field is that, over a finite field, a nonempty Zarisky-open set is no longer dense, so the connection with our intuitive idea of randomness is lost. Nevertheless, for a finite field $\mathbb{F}_q$, we may define randomness using a Zariski-dense open set defined over the algebraic closure $\overline{\mathbb{F}_q}$. While it is not necessarily the case that almost every polynomial system of given degrees with coefficients in $\mathbb{F}_q$ is random, this is the case whenever q is large enough, or if we consider a finite extension of $\mathbb{F}_q$ of large enough cardinality.

Definition 7. *Let* $\mathbb{F}$ *be a field and let* $\overline{\mathbb{F}}$ *be its algebraic closure. Denote by* $\mathbb{P}^t$ *the* t*-dimensional projective space over* $\overline{\mathbb{F}}$*. Let* $d_1, \ldots, d_m$ *be positive integers and let* $\mathcal{U} \subseteq \mathbb{P}^{\binom{n+d_1-1}{n-1}-1} \times \ldots \times \mathbb{P}^{\binom{n+d_m-1}{n-1}-1}$ *be a nonempty Zariski-open set. For a homogeneous polynomial* f*, denote by* $[f] \in \mathbb{P}^{\binom{n+\deg(f)-1}{n-1}-1}$ *the projective point, whose coordinates are the coefficients of* f*. A homogeneous system* $\mathcal{F} = \{f_1, \ldots, f_m\} \subseteq R$ *with* $\deg(f_i) = d_i$ *for* $1 \leq i \leq m$ *is* **generic with respect to** $\mathcal{U}$ *if* $([f_1], \ldots, [f_m]) \in \mathcal{U}$*. It is* **generic** *if it is generic with respect to* $\mathcal{U}$*, for some nonempty Zariski-open set* $\mathcal{U}$*. An arbitrary system* $\mathcal{F}$ *is* **generic** *or* **generic with respect to** $\mathcal{U}$ *if* $\mathcal{F}^{\text{top}}$ *is.*

In the cryptographic literature, random sequences of polynomials are often assumed to be cryptographic semiregular sequences, see e.g. [2, 4]. This is the case in the cryptoanalysis of several systems, as e.g. [15]. The next definition appears in [2, Definition 3.2.1 and Definition 3.2.4], [5, Definition 5], and [7, Definition 5 and Definition 9].

Definition 8. *Let* $\mathcal{F} = \{f_1, \ldots, f_m\} \subseteq R$ *be a homogeneous system.*

If $\mathbb{F} \neq \mathbb{F}_2$*, we say that* $f_1, \ldots, f_m$ *are a* **cryptographic semiregular sequence** *if for all* $1 \leq i \leq m$ *and all* $g_i \in R$ *such that* $g_i f_i \in (f_1, \ldots, f_{i-1})$ *and* $\deg(g_i f_i) < d_{\text{reg}}(\mathcal{F})$*, one has that* $g_i \in (f_1, \ldots, f_{i-1})$*.*

If $\mathbb{F} = \mathbb{F}_2$*, we say that* $f_1, \ldots, f_m \in R/(x_1^2, \ldots, x_n^2)$ *are a* **cryptographic semiregular sequence** *if for all* $1 \leq i \leq m$ *and all* $g_i \in R/(x_1^2, \ldots, x_n^2)$ *such that* $g_i f_i \in (f_1, \ldots, f_{i-1})$ *and* $\deg(g_i f_i) < d_{\text{reg}}(\mathcal{F} \cup \{x_1^2, \ldots, x_n^2\})$*, one has that* $g_i \in (f_1, \ldots, f_i)$*.*

Arbitrary polynomials $f_1, \ldots, f_m$ *are a* **cryptographic semiregular sequence** *if* $f_1^{\text{top}}, \ldots, f_m^{\text{top}}$ *are a cryptographic semiregular sequence.*

In this paper, we use the word cryptographic semiregular sequence in order to distinguish the concept of semiregularity used in the cryptographic literature from the concept of semiregularity originally introduced by Pardue [24, 25], which inspired it. The original definition by Pardue is given over an infinite field $\mathbb{F}$. As we are interested also in dealing with finite fields, we extend it in the natural way.

Definition 9. *Let $\mathbb{F}$ be an infinite field and let $R = \mathbb{F}[x_1, \ldots, x_n]$. Let I be a homogeneous ideal and let $A = R/I$. A polynomial $f \in R_d$ is* **semiregular** *on A if for every $e \geq d$, the vector space map $A_{e-d} \to A_e$ given by multiplication by f has maximal rank (that is, it is either injective or surjective). If $\mathbb{F}$ is a finite field, let $\overline{R} = \overline{\mathbb{F}}[x_1, \ldots, x_n]$. Then f is* **semiregular** *on A if it is semiregular on $\overline{R}/I\overline{R}$.*

A sequence of homogeneous polynomials $f_1, \ldots, f_m$ is a **semiregular sequence** *if f_i is semiregular on $A/(f_1, \ldots, f_{i-1})$ for all $1 \leq i \leq m$.*

It follows from [25, Proposition 1] that, if $\mathbb{F} \neq \mathbb{F}_2$, then a semiregular sequence is also a cryptographic semiregular sequence. The converse does not hold, as shown in [25], see the example just below [25, Proposition 1].

The next proposition presents a simple situation in which cryptographic semiregular sequences and semiregular sequences coincide.

Proposition 10. *Let $q > 2$ and let $f_1 = x_1^q - x_1, \ldots, f_n = x_n^q - x_n, f_{n+1} = f \in R = \mathbb{F}_q[x_1, \ldots, x_n]$. The sequence $f_1, \ldots, f_{n+1}$ is cryptographic semiregular if and only if the sequence $f_1^{\text{top}} = x_1^q, \ldots, f_n^{\text{top}} = x_n^q, f_{n+1}^{\text{top}} = f^{\text{top}}$ is semiregular.*

Proof. It suffices to show that, if the sequence $f_1, \ldots, f_{n+1}$ is cryptographic semiregular, then the sequence $x_1^q, \ldots, x_n^q, f^{\text{top}}$ is semiregular. We start by observing that the sequence $x_1^q, \ldots, x_n^q$ is regular. Therefore, in order to show that $x_1^q, \ldots, x_n^q, f^{\text{top}}$ is semiregular, it suffices to show that f^{top} is semiregular on $\mathbb{F}_q[x_1, \ldots, x_n]/(x_1^q, \ldots, x_n^q)$. If $f_1, \ldots, f_{n+1}$ is cryptographic semiregular and $d = \deg(f)$, then the Hilbert series of $R/(\mathcal{F}^{\text{top}})$ is $[(1 - z^d)(1 + z + \ldots + z^{q-1})^n]$ by [7, Proposition 6]. Here for $p(z) = \sum_{i=0}^{\infty} p_i z^i \in \mathbb{Z}[[z]]$, we denote by $\delta(p) = \min\{i \geq 0 \mid p_i \leq 0\} - 1$ and define $[p(z)] = \sum_{i=0}^{\delta(p)} p_i z^i$. Then f^{top} is semiregular on $\mathbb{F}_q[x_1, \ldots, x_n]/(x_1^q, \ldots, x_n^q)$ and $\mathcal{F}^{\text{top}}$ is a semiregular sequence by [25, Proposition 1]. ◻

Pardue in [24, 25] shows that Fröberg's Conjecture [22], a conjecture which has attracted a lot of attention within the commutative algebra community and that is widely believed to hold, is equivalent to the following

Conjecture 11. *Let $\mathbb{F}$ be an infinite field. A generic sequence of polynomials of degrees $d_1, \ldots, d_m$ in $R = \mathbb{F}[x_1, \ldots, x_n]$ is semiregular.*

In other words, Fröberg conjectures that the set of semiregular sequences of polynomials of given degrees contains a dense Zarisky-open set. If the conjecture is true, then a sequence of polynomials of given degrees is semiregular with high probability, provided that the ground field has large enough cardinality. It follows that, if $\mathbb{F} = \mathbb{F}_q$ with $q \gg 0$ and Fröberg's Conjecture holds, then a sequence of polynomials of given degrees is a cryptographic semiregular sequence with high probability. In addition, most cryptographic semiregular sequences are also semiregular sequences, as the set of semiregular sequences conjecturally contains a dense open set.

In [2, Section 3.2], Bardet conjectures that a sequence of polynomials with coefficients in $\mathbb{F}_2$ is cryptographic semiregular with high probability. This conjecture is motivated by experimental evidence, see also [4, Conjecture 2]. The conjecture was later disproved by Hodges, Molina, and Schlather, who in [23] prove that there are choices of the parameters for which no cryptographic semiregular sequence exists over $\mathbb{F}_2$. This is the case, e.g., for $m = 1$ and $n > 3d_1$. In the sequel, we propose a notion of randomness which applies to any choice of the system parameters.

In this paper, we propose two dense Zariski-open sets, which can be used to formalize the intuitive idea of a random system. The set $\mathcal{V}$ corresponds to systems of m polynomials in n variables which contain a regular sequence of length n, while the set $\mathcal{U}$ parametrizes systems of m polynomials which contain a regular sequence of n polynomials of the smallest possible degrees. Notice that $\mathcal{V}$ contains $\mathcal{U}$ and, if $m \geq n$, it also contains the set of cryptographic semiregular sequences and that of semiregular sequences.

Definition 12. *Fix $m \geq n \geq 1$ and $1 \leq d_1 \leq \ldots \leq d_m$. For any multiset Δ of cardinality n contained in the multiset $\{d_1, \ldots, d_m\}$, let $\mathcal{U}_\Delta$ be the subset of $\mathbb{P}^{\binom{n+d_1-1}{n-1}-1} \times \ldots \times \mathbb{P}^{\binom{n+d_m-1}{n-1}-1}$ whose points correspond to polynomials $f_1, \ldots, f_m \in R$ such that $(f_1^{\text{top}}, \ldots, f_m^{\text{top}})$ contains a regular sequence in the degrees of Δ. Let*

$$\mathcal{U} = \mathcal{U}_{\{d_1,\ldots,d_n\}} \quad \text{and} \quad \mathcal{V} = \bigcup_{\Delta \subseteq \{d_1,\ldots,d_m\}, |\Delta|=n} \mathcal{U}_\Delta.$$

An ***algebraically random system*** *of m polynomials of degrees $d_1, \ldots, d_m$ in n variables is a system $\mathcal{F} = \{f_1, \ldots, f_m\} \subseteq R$ such that $\deg(f_i) = d_i$ for all $1 \leq i \leq m$ and $([f_1], \ldots, [f_m]) \in \mathcal{U}$.*

It is well-known that $\mathcal{U}_{\{d_1,\ldots,d_n\}}$ is a dense open set for any choice of $1 \leq d_1 \leq \ldots \leq d_n$, if $m = n$. This implies that any $\mathcal{U}_\Delta$ as in Definition 12 is a dense open set, hence also $\mathcal{U}$ and $\mathcal{V}$ are. Notice that the degree of regularity is finite for all systems in $\mathcal{U}_\Delta$ and for any choice of Δ, therefore it is also finite for the systems in $\mathcal{U}$ and $\mathcal{V}$. In particular, the degree of regularity of an algebraically random system is finite. Notice moreover that, if $\mathcal{F}$ is homogeneous, then $\mathcal{V}$ is the set of polynomial systems of degrees $d_1, \ldots, d_m$ for which the degree of regularity is finite.

Notice moreover that, if $m = n$, an algebraically random system is a system for which $\mathcal{F}^{\text{top}}$ is a regular sequence. Since this case is well-studied in the cryptographic literature, in the sequel we often assume $m > n$.

Remark 13. *For any system $\mathcal{F}$ of equations of degree at least q, one has*

$$\mathcal{F} \cup \{x_1^q - x_1, \ldots, x_n^q - x_n\} \in \mathcal{U}_{\{q,\ldots,q\}}.$$

Unlike semiregular systems [23], algebraically random systems exist for any choice of the parameters and over every finite field.

Remark 14. *Algebraic random systems over $\mathbb{F}_q$ exist for any choice of n, m and $1 \leq d_1 \leq \ldots \leq d_m$. This corresponds to the existence of regular sequences of any given degrees in $\mathbb{F}_q[x_1, \ldots, x_n]$, as any system of m equations in $\mathbb{F}_q[x_1, \ldots, x_n]$ of degrees $d_1, \ldots, d_m$ which contains a regular sequence in degrees $d_1, \ldots, d_n$ is algebraically random. Some regular sequences in $\mathbb{F}_q[x_1, \ldots, x_n]$ of degrees $d_1, \ldots, d_n$ are for example*

$$x_1^{d_1} + g_1, x_2^{d_2} + g_2, \ldots, x_n^{d_n} + g_n,$$

where $g_i \in \mathbb{F}_q[x_{i+1}, \ldots, x_n]$ and $\deg(g_i) \leq d_i$.

4 The degree of regularity of a random system

In this section, we establish an upper bound for the degree of regularity of an algebraically random system $\mathcal{F}$ consisting of m polynomials of equal degree D. In combination with Theorem 3 and Theorem 4, this provides us with an upper bound for the solving degree of a system of algebraically random polynomials of the same degree.

Remark 15. *Notice that $\mathcal{F} \in \mathcal{U}_{\{d,\ldots,d\}}$ for a given $d > 0$ if and only if $T \circ \mathcal{F} \circ S \in \mathcal{U}_{\{d,\ldots,d\}}$. In other words, when deciding whether a system of polynomials of equal degree is algebraically random, one can safely ignore the random linear transformations S and T used to disguise the internal system $\mathcal{F}$. This also shows that, for system whose equations are all of the same degree, being an algebraically random system is an intrinsic property of the system and it is not affected by the invertible linear transformations used to disguise the system.*

In [10, Section 4], the authors provide an upper bound for the degree of regularity of a system of quadratic polynomials which contains a regular sequence. In this section, we follow the same basic approach and extend it to systems of polynomials of the same degree and systems of polynomials of the same degree to which one adds the field equations. The cases that we treat in this paper are technically more challenging and require the use of more sophisticated results from commutative algebra.

We start by introducing the family of lex-segment ideals. A conjecture by Eisenbud, Green, and Harris will allow us to reduce to these ideals, when estimating the regularity of ideals generated by algebraically random systems. Throughout the section, we fix the lexicographic term order on R with $x_1 > x_2 > \ldots > x_n$.

Definition 16. *A monomial ideal $I \subseteq R$ is a* **lex-segment ideal** *if it has the property that if $u, v \in R$ are monomials of the same degree such that $u \geq_{\mathrm{lex}} v$ and $v \in I$, then $u \in I$.*

Let C and $c_1 \leq \ldots \leq c_n$ be non negative integers. An ideal $\mathcal{L} \subseteq R$ is a **$(c_1, \ldots, c_n; C)$-LexPlusPowers (LPP) ideal** *if $\mathcal{L} = (x_1^{c_1}, \ldots, x_n^{c_n}) + L$, where L is a lex-segment ideal generated in degree C.*

Notation 17. *Let $I \subseteq R$ be a homogeneous ideal containing a regular sequence of polynomials of degrees $c_1 \leq \ldots \leq c_n$. For each $C \geq 0$, we denote by $\mathrm{LPP}(I; c_1, \ldots, c_n; C)$ the $(c_1, \ldots, c_n; C)$-LPP ideal $\mathcal{L} = (x_1^{c_1}, \ldots, x_n^{c_n}) + L$ such that $\dim(I_C) = \dim(\mathcal{L}_C)$. We make L unique by choosing the largest lex-segment ideal generated in degree C for which the equality $\mathcal{L} = (x_1^{c_1}, \ldots, x_n^{c_n}) + L$ holds.*

Example 18. *Let $I \subseteq \mathbb{F}[x, y, z]$ be generated by a homogeneous regular sequence of polynomials of degrees $1, 3, 4$. The $(1, 3, 4; 3)$-LPP ideal $\mathcal{L}$ such that $\dim(I_3) = 7 = \dim(\mathcal{L}_3)$ is $\mathcal{L} = (x, y^3, z^4)$.*

For any lex-segment ideal L generated in degree 3 with $\dim(L_3) \leq 7$, one has $L \subseteq \mathcal{L}$. Hence one may write

$$\mathcal{L} = (x, y^3, z^4) + L$$

as in Definition 16 and choose e.g. $L = (x^3, x^2y, x^2z, xy^2, xyz, xz^2)$, or $L = (x^3)$. However, Notation 17 prescribes that we choose the largest L with respect to containment, i.e.

$$L = (x^3, x^2y, x^2z, xy^2, xyz, xz^2, y^3)$$

is the lex-segment ideal generated by the first 7 cubic monomials in the lexicographic order.

The next conjecture appears as [19, Conjecture (V_m)]. It has been settled in several cases and it is widely believed to hold within the commutative algebra community. For an introduction to the conjecture and an excellent survey of known cases, we refer the interested reader to [17]. Here we state it in a weak form, which is what we need in the sequel.

Conjecture 19 (Eisenbud-Green-Harris Conjecture)**.** *Let $I \subseteq R$ be a homogeneous ideal containing a regular sequence of polynomials of degrees $c_1 \leq \ldots \leq c_n$. Then*

$$\mathrm{reg}(I) \leq \mathrm{reg}(\mathrm{LPP}(I; c_1, \ldots, c_n; C))$$

for all $C \geq c_n$.

In order to estimate the degree of regularity of our systems, we use the following result by Caviglia and De Stefani.

Proposition 20 ([16, Lemma 2.3])**.** *Let $c_1 \leq \ldots \leq c_n$ and $2 \leq C \leq \sum_{i=1}^{n}(c_i - 1)$. Let $\mathcal{L} = (x_1^{c_1}, \ldots, x_n^{c_n}) + L$ be a $(c_1, \ldots, c_n; C)$-LPP ideal, and assume that $\mathcal{L} \neq (x_1^{c_1}, \ldots, x_n^{c_n})$. Let $u = x_k^{t_k} v$, with $t_k \neq 0$ and $v \in \mathbb{F}[x_{k+1}, \ldots, x_n]$, be the smallest monomial with respect to the lexicographic order which belongs to L and has degree C. Then*

$$\mathrm{reg}(\mathcal{L}) = t_k + \sum_{i=k+1}^{n} (c_i - 1).$$

Our first result is an explicit bound for the degree of regularity of an algebraically random system of polynomials of the same degree. In order to make the proof more readable, we introduce the following notation.

Notation 21. *If $u \in R_D$ is a monomial that only involves the variables $x_k, \ldots, x_n$ and has degree a in x_k, we say that u is a (D, k, a)-type monomial.*

In the next theorem, we provide an explicit formula for the degree of regularity of an LPP ideal with given parameters. Thanks to Conjecture 19, this yields an upper bound for the degree of regularity of an algebraically random system of polynomials of the same degree. In the statement of the theorem, $\sigma_{k,t}$ is the position of the smallest $(D, k, D - t)$-type monomial in the ordered list of monomials of degree D different from $x_1^D, \ldots, x_n^D$, sorted in decreasing lexicographic order. In particular, $\sigma_{1,0} = 0$ and $\sigma_{n-1,D-1}$ is the number of monomials of degree D different from $x_1^D, \ldots, x_n^D$.

Theorem 22. *Assume that Conjecture 19 holds. Let $\mathcal{F} = \{f_1, \dots, f_m\} \subseteq R$ be a polynomial system and assume without loss of generality that $f_1^{\mathrm{top}}, \dots, f_m^{\mathrm{top}}$ are linearly independent of degree D. If $m = n$, then let $k = 0$, $t = D - 1$. Else, let $1 \leq k \leq n-1$ and $1 \leq t \leq D-1$ be such that $m - n$ belongs to the interval $(\sigma_{k,t-1}, \sigma_{k,t}]$, where*

$$\sigma_{k,t} = \sum_{i=1}^{k} \sum_{j=1}^{D-1} \binom{n-i-1+j}{j} - \sum_{j=t+1}^{D-1} \binom{n-k-1+j}{j}.$$

If $\mathcal{F}$ is an algebraically random polynomial system, then

$$d_{\mathrm{reg}}(\mathcal{F}) \leq (D-t) + (n-k)(D-1).$$

Proof. If $m = n$, then $\mathcal{F}^{\mathrm{top}}$ is a regular sequence of n polynomials of degree D, hence

$$d_{\mathrm{reg}}(\mathcal{F}) = n(D-1) + 1.$$

Suppose therefore that $m > n$ and let J be the ideal generated by $\mathcal{F}^{\mathrm{top}}$. Consider the lexicographic order on R and let

$$\mathcal{L} = \mathrm{LPP}(J; D, \dots, D; D) = (x_1^D, \dots, x_n^D) + L,$$

be the LPP ideal with L the largest lex-segment ideal generated in degree D such that $\dim(\mathcal{L}_D) = m$. Since both $f_1^{\mathrm{top}}, \dots, f_m^{\mathrm{top}}$ and $x_1^D, \dots, x_n^D$ are linearly independent, we have

$$\dim\left(\mathcal{L}_D / \langle x_1^D, \dots, x_n^D \rangle\right) = m - n.$$

For $1 \leq k \leq n-1$ and $0 \leq t \leq D-1$, the number of $(D, k, D-t)$-type monomials is

$$\dim(\mathbb{F}_q[x_{k+1}, \dots, x_n]_t) = \binom{n-k-1+t}{t},$$

hence

$$\begin{aligned}\sigma_{k,t} &= \sum_{i=1}^{k-1} \sum_{j=1}^{D-1} \binom{n-i-1+j}{j} + \sum_{j=1}^{t} \binom{n-k-1+j}{j} \\ &= \sum_{i=1}^{k-1} \sum_{j=1}^{D-1} \dim(\mathbb{F}_q[x_{i+1}, \dots, x_n])_j + \sum_{j=1}^{t} \dim(\mathbb{F}_q[x_{k+1}, \dots, x_n])_j\end{aligned}$$

is the number of degree D monomials in $\mathbb{F}_q[x_1, \dots, x_n]$ different from $x_1^D, \dots, x_n^D$ and bigger than or equal to $x_k^{D-t} x_n^t$, the smallest $(D, k, D-t)$-type monomial. In other words, the monomial $x_k^{D-t} x_n^t$ is in position $\sigma_{k,t}$ in the ordered list of degree D monomials in $\mathbb{F}_q[x_1, \dots, x_n]/(x_1^D, \dots, x_n^D)$. Notice moreover that $\sigma_{1,0} = 0$ and $\sigma_{k,0} = \sigma_{k-1,D-1}$ for $2 \leq k \leq n-1$.

If u is the smallest monomial in $\mathcal{L}_D / \langle x_1^D, \dots, x_n^D \rangle$, then u is in position $m - n$ in the ordered list of degree D monomials in $\mathbb{F}_q[x_1, \dots, x_n]/(x_1^D, \dots, x_n^D)$. If $\sigma_{k,t-1} < m - n \leq \sigma_{k,t}$ for some $1 \leq k \leq n-1$ and $1 \leq t \leq D-1$, then u is a $(D, k, D-t)$-type monomial. Since $0 = \sigma_{1,0} < m - n \leq \dim(\mathbb{F}_q[x_1, \dots, x_n]_D)/(x_1^D, \dots, x_n^D)_D = \sigma_{n-1,D-1}$, then $m-n$ always belong to one of the intervals above.

The ideal L is generated by the degree D monomials which are greater than or equal to u, unless the monomial following u in lexicographic decreasing order is a pure power. In that case, $u = x_k x_n^{D-1}$ for some $1 \leq k \leq n-1$, i.e. $m - n = \sigma_{k,D-1}$, and the smallest degree D monomial in L is x_{k+1}^D. Then L is generated by the degree D monomials which are greater than or equal to x_{k+1}^D, which is a $(D, k+1, D)$-type monomial. In both situations

$$\mathrm{reg}(\mathcal{L}) = (D-t) + (n-k)(D-1)$$

by Proposition 20. The thesis now follows from observing that

$$d_{\mathrm{reg}}(\mathcal{F}) = \mathrm{reg}(J) \leq \mathrm{reg}(\mathcal{L}), \tag{1}$$

where the inequality follows from Conjecture 19. □

Example 23. *In this example we show how to compute the bound from Theorem 22 for concrete choices of the parameters. Let* $n = 6$ *and* $D = 3$*, so* $\mathcal{F} = \{f_1, \ldots, f_m\} \subseteq \mathbb{F}_q[x_1, \ldots, x_6]_3$*. Then* $1 \leq k \leq 5$ *and* $0 \leq t \leq 2$*. The values of* $\sigma_{k,t}$ *are*

$\sigma_{k,t}$	$t = 0$	$t = 1$	$t = 2$
$k = 1$	0	5	20
$k = 2$	20	24	34
$k = 3$	34	37	43
$k = 4$	43	45	48
$k = 5$	48	49	50

If $m = 12$*, then* $m - n = 6$ *and* $\sigma_{1,1} < 6 \leq \sigma_{1,2}$*. Hence* $k = 1$*,* $t = 2$*, and* $d_{\text{reg}}(\mathcal{F}) \leq 11$*.*
If $m = 42$*, then* $m - n = 36$ *and* $\sigma_{3,0} < 36 \leq \sigma_{3,1}$*. Hence* $k = 3$*,* $t = 1$*, and* $d_{\text{reg}}(\mathcal{F}) \leq 8$*.*
If $m = 54$*, then* $m - n = 48$ *and* $\sigma_{4,1} < 48 \leq \sigma_{4,2}$*. Hence* $k = 4$*,* $t = 2$*, and* $d_{\text{reg}}(\mathcal{F}) \leq 5$*.*

Remark 24. *The upper bound of Theorem 22 is decreasing as a function of* m*, as one would expect. In particular, as* $m - n$ *passes from an interval* $(\sigma_{k,t-1}, \sigma_{k,t}]$ *to the next, the upper bound decreases by one. The largest value of the bound is obtained in the case* $m = n$*, which corresponds to* $\mathcal{F}^{\text{top}}$ *being a regular sequence. In this case, the value for the bound is well-know and is* $n(D-1)+1$*. The smallest value for the bound is obtained in the case* $m - n = \sigma_{n-1,D-1}$*, which corresponds to* $\langle \mathcal{F}^{\text{top}} \rangle + \langle x_1^D, \ldots, x_n^D \rangle = R_D$*. In this case, the value for the bound is easily seen to be* D*.*

Remark 25. *The upper bound produced in Theorem 22 is sharp for all values of* m, n, D*. In fact, it is met by any system* $\mathcal{F}$ *such that* $(f_1^{\text{top}}, \ldots, f_m^{\text{top}})$ *is a* $(D, \ldots, D; D)$*-LPP ideal.*

Combining Theorem 22 and Theorem 4, we obtain the following.

Corollary 26. *Let* $\mathcal{F} \subseteq R$ *be an algebraically random system of degree* D *polynomials. If* $m = n$*, then let* $k = 0$*,* $t = D - 1$*. Else, let* $1 \leq k \leq n-1$ *and* $1 \leq t \leq D-1$ *be such that* $m - n$ *belongs to the interval* $(\sigma_{k,t-1}, \sigma_{k,t}]$*. If Conjecture 19 holds, then*

$$\text{solv.deg}^m(\mathcal{F}) \leq D - t + 1 + (n-k)(D-1).$$

Proof. The upper bound found in Theorem 22 for the degree of regularity of $\mathcal{F}$ is bigger than or equal to D for all k and t. The thesis then follows from Theorem 4. ☐

We now wish to apply Theorem 22 to a system which contains the field equations.

Remark 27. *After reducing the equations of* $\mathcal{F}$ *modulo the field equations, they have degree at most* $q - 1$ *in each variable, hence total degree at most* $n(q-1)$*. Therefore, when adding the field equations to a polynomial system of degree* D*, we may always assume that*

$$D \leq n(q-1).$$

Combining Theorem 22 and Theorem 3, one obtains the following. By Remark 27, we may assume without loss of generality that $D \leq n(q-1)$.

Corollary 28. *Assume that Conjecture 19 holds. Let* $\mathcal{F} \subseteq R$ *be an algebraically random system of degree* D *polynomials, with* $D \leq n(q-1)$*. If* $m = n$*, then let* $k = 0$*,* $t = D-1$*. Else, let* $1 \leq k \leq n-1$ *and* $1 \leq t \leq D-1$ *be such that* $m - n$ *belongs to the interval* $(\sigma_{k,t-1}, \sigma_{k,t}]$*. If* $d_{\text{reg}}(\mathcal{F}) \geq q$*, then*

$$\text{solv.deg}^s(\mathcal{F} \cup \{x_1^q - x_1, \ldots, x_n^q - x_n\}) \leq 2\min\{n(q-1), (n-k+1)(D-1) - t\}.$$

Proof. First, since $(x_1^q, \ldots, x_n^q)_{n(q-1)+1} = R_{n(q-1)+1}$, one has

$$d_{\text{reg}}(\mathcal{F}^{\text{top}} \cup \{x_1^q, \ldots, x_n^q\}) \leq n(q-1)+1.$$

If $D < q$, then

$$(\mathcal{F}^{\text{top}} \cup \{x_1^q, \ldots, x_n^q\})_{q-1} = (\mathcal{F}^{\text{top}})_{q-1} \neq R_{q-1},$$

since $d_{\mathrm{reg}}(\mathcal{F}) \geq q$ by assumption. Therefore, $d_{\mathrm{reg}}(\mathcal{F} \cup \{x_1^q - x_1, \ldots, x_n^q - x_n\}) \geq q = \max\{D, q\}$. If $q \leq D \leq n(q-1)$, then

$$(\mathcal{F}^{\mathrm{top}} \cup \{x_1^q, \ldots, x_n^q\})_{D-1} = (x_1^q, \ldots, x_n^q)_{D-1} \neq R_{D-1},$$

hence $d_{\mathrm{reg}}(\mathcal{F} \cup \{x_1^q - x_1, \ldots, x_n^q - x_n\}) \geq D = \max\{D, q\}$. This shows that the assumptions of Theorem 3 are satisfied. The thesis now follows by combining Theorem 22, Theorem 3, and the observation that

$$d_{\mathrm{reg}}(\mathcal{F} \cup \{x_1^q - x_1, \ldots, x_n^q - x_n\}) \leq d_{\mathrm{reg}}(\mathcal{F}).$$

□

Remark 29. *While the estimates of Corollary 26 and of Corollary 28 hold for every D, they are most relevant for $D \leq q$. In fact, for $D > q$ we obtain tighter upper bounds on the degree of regularity - hence on the solving degree - of $\mathcal{F} \cup \{x_1^q - x_1, \ldots, x_n^q - x_n\}$ in Theorem 34 by inspecting the degree D part of the system $\mathcal{F} \cup \{x_1^q - x_1, \ldots, x_n^q - x_n\} \in \mathcal{U}_{\{q,\ldots,q\}}$. This corresponds to the fact that, whenever the degree of the equations of the system is larger than or comparable to the field size, it is convenient to add the field equations to the system before computing a Gröbner basis.*

Next we estimate the degree of regularity and the solving degree of $\mathcal{F} \cup \{x_1^q - x_1, \ldots, x_n^q - x_n\}$ in the case when $D \geq q$. By Remark 27, we may assume that $D \leq n(q-1)$.

First, we observe that one can easily derive a lower bound on the degree of regularity.

Remark 30. *Let $\mathcal{F} = \{f_1, \ldots, f_m\}$ be a polynomial system of degree $q \leq D \leq n(q-1)$. Since $(\mathcal{F}^{\mathrm{top}} \cup \{x_1^q, \ldots, x_n^q\})_{D-1} = (x_1^q, \ldots, x_n^q)_{D-1} \neq R_{D-1}$, then*

$$d_{\mathrm{reg}}(\mathcal{F} \cup \{x_1^q - x_1, \ldots, x_n^q - x_n\}) \geq D.$$

We now want to derive an upper bound. We start by computing the number of linearly independent homogeneous polynomials of degree D as a function of n, D, q.

Remark 31. *Assume that $q \leq D \leq n(q-1)$. A standard Hilbert function computation shows that the number of homogeneous polynomials in n variables of degree D which are linearly independent modulo $(x_1^q, \ldots, x_n^q)$ is*

$$\dim(R/(x_1^q, \ldots, x_n^q))_D = \sum_{i=0}^{\lfloor \frac{D}{q} \rfloor} (-1)^i \binom{n}{i} \binom{n+D-1-iq}{n-1}.$$

It follows that the assumption that $f_1^{\mathrm{top}}, \ldots, f_m^{\mathrm{top}}$ are linearly independent modulo $(x_1^q, \ldots, x_n^q)$ holds on a dense open set, whenever

$$m \leq \sum_{i=0}^{\lfloor \frac{D}{q} \rfloor} (-1)^i \binom{n}{i} \binom{n+D-1-iq}{n-1}. \tag{2}$$

If inequality (2) is not satisfied, then $f_1^{\mathrm{top}}, \ldots, f_m^{\mathrm{top}}$ cannot be linearly independent.

The simplest case to treat is that of very overdetermined systems, more specifically the case when $f_1^{\mathrm{top}}, \ldots, f_m^{\mathrm{top}}$ are too many to be linearly independent modulo $(x_1^q, \ldots, x_n^q)$. This happens when m is larger than the bound from Remark 31. The next proposition shows that, in such a situation, the degree of regularity is equal to the degree of the equations of the system.

Proposition 32. *Let $\mathcal{F} = \{f_1, \ldots, f_m\}$ be a polynomial system of degree D, where $q \leq D \leq n(q-1)$. If $m > \sum_{k=0}^{\lfloor \frac{D}{q} \rfloor} (-1)^k \binom{n}{k} \binom{n+D-1-kq}{n-1}$, then there is a dense open set $\mathcal{W}$ such that, if $\mathcal{F} \in \mathcal{W}$, then*

$$d_{\mathrm{reg}}(\mathcal{F} \cup \{x_1^q - x_1, \ldots, x_n^q - x_n\}) = D.$$

Moreover,

$$\mathrm{solv.deg}^s(\mathcal{F} \cup \{x_1^q - x_1, \ldots, x_n^q - x_n\}) \leq 2D - 2$$

and

$$\mathrm{solv.deg}^m(\mathcal{F} \cup \{x_1^q - x_1, \ldots, x_n^q - x_n\}) \leq D + 1.$$

Proof. Since by assumption

$$m > \sum_{k=0}^{\lfloor \frac{D}{q} \rfloor} (-1)^k \binom{n}{k} \binom{n+D-1-kq}{n-1} = \dim(R/(x_1^q, \ldots, x_n^q))_D,$$

then there is an open set $\mathcal{W}$ of m-tuples of polynomials of degree D such that

$$\langle f_1^{\text{top}}, \ldots, f_m^{\text{top}} \rangle + (x_1^q, \ldots, x_n^q)_D = R_D.$$

Since $(\mathcal{F}^{\text{top}} \cup \{x_1^q, \ldots, x_n^q\})_{D-1} = (x_1^q, \ldots, x_n^q)_{D-1} \neq R_{D-1}$, then

$$d_{\text{reg}}(\mathcal{F} \cup \{x_1^q - x_1, \ldots, x_n^q - x_n\}) = D.$$

The rest of the statement now follows from Theorem 3 and Theorem 4. □

The next theorem yields an upper bound on the degree of regularity of an algebraically random system of equations of degree larger than the field size to which we add the field equations. For such a system, the bound is tighter than the one from Corollary 28. We start with a preparatory lemma, whose proof follows directly from the definition.

Lemma 33. *Let u, v be monomials of type (D, k, a) and (D, h, b), respectively. If $u \geq v$, then either $k < h$ or $k = h$ and $a \geq b$. In particular, if u, v, w are monomials such that $u \geq v \geq w$ and u and w have the same type, then v also has the same type as u and w.*

Theorem 34. *Assume that Conjecture 19 holds. Let $\mathcal{F} = \{f_1, \ldots, f_m\}$ be a polynomial system of degree D, with $q \leq D \leq n(q-1)$. Assume that $m \leq \sum_{i=0}^{\lfloor \frac{D}{q} \rfloor} (-1)^i \binom{n}{i} \binom{n+D-1-iq}{n-1}$ and that $f_1^{\text{top}}, \ldots, f_m^{\text{top}}$ are linearly independent modulo $(x_1^q, \ldots, x_n^q)_D$. Let $1 \leq t \leq q-1$ and $1 \leq k \leq n-1$ be such that m belongs to the interval $(\sigma_{k,t-1}, \sigma_{k,t}]$, where*

$$\sigma_{k,t} = \sum_{i=1}^{k-1} \sum_{j=1}^{q-1} \eta_{i,j} + \sum_{j=1}^{t} \eta_{k,j}$$

and

$$\eta_{k,t} = \sum_{i=0}^{\lfloor \frac{D+t}{q} \rfloor - 1} (-1)^i \binom{n-k}{i} \binom{n-k-1+D-(i+1)q+t}{n-k-1}.$$

Let

$$B = q - t + (n-k)(q-1).$$

If $m = \sigma_{k,t}$ and either $t \neq q-1$ and $D \geq 2q-t-1$, or $t = q-1$ and $D \geq 2q-1$, then

$$d_{\text{reg}}(\mathcal{F} \cup \{x_1^q - x_1, \ldots, x_n^q - x_n\}) \leq B - 1,$$

$$\text{solv.deg}^s(\mathcal{F} \cup \{x_1^q - x_1, \ldots, x_n^q - x_n\}) \leq 2(B-2),$$

and

$$\text{solv.deg}^m(\mathcal{F} \cup \{x_1^q - x_1, \ldots, x_n^q - x_n\}) \leq B + 1.$$

In any other case

$$d_{\text{reg}}(\mathcal{F} \cup \{x_1^q - x_1, \ldots, x_n^q - x_n\}) \leq B,$$

$$\text{solv.deg}^s(\mathcal{F} \cup \{x_1^q - x_1, \ldots, x_n^q - x_n\}) \leq 2(B-1),$$

and

$$\text{solv.deg}^m(\mathcal{F} \cup \{x_1^q - x_1, \ldots, x_n^q - x_n\}) \leq B + 1.$$

Proof. We start by proving that $\eta_{k,t}$ is the number of $(D,k,q-t)$-type monomials in R which are linearly independent modulo $(x_1^q,\dots,x_n^q)_D$, that is

$$\eta_{k,t}=\dim\left[\frac{\mathbb{F}_q[x_{k+1},\dots,x_n]}{(x_{k+1}^q,\dots,x_n^q)}\right]_{D-q+t}. \tag{3}$$

Since $x_{k+1}^q,\dots,x_n^q$ is a regular sequence in $\mathbb{F}_q[x_{k+1},\dots,x_n]$, a standard computation involving Hilbert series yields the explicit formula

$$\dim\left[\frac{\mathbb{F}_q[x_{k+1},\dots,x_n]}{(x_{k+1}^q,\dots,x_n^q)}\right]_{D-q+t}=\sum_{i=0}^{\lfloor\frac{D+t}{q}\rfloor-1}(-1)^i\binom{n-k}{i}\binom{n-k-1+D-(i+1)q+t}{n-k-1},$$

where we notice that $\eta_{k,t}\neq 0$ only if $D+t-q\leq (n-k)(q-1)$. This establishes the equality in (3). Similarly to the proof of Theorem 22, we notice that

$$\sigma_{k,t}=\sum_{i=1}^{k-1}\sum_{j=1}^{q-1}\eta_{i,j}+\sum_{j=1}^{t}\eta_{k,j}$$

is the number of monomials of degree D which do not belong to $(x_1^q,\dots,x_n^q)$ and are greater than or equal to $x_k^{q-t}x_n^{D-q+t}$, the smallest $(D,k,q-t)$-type monomial. Moreover, $\sigma_{k-1,q-1}=\sigma_{k,0}$ for $2\leq k\leq n-1$.

Let $I\subseteq\mathbb{F}_q[x_1,\dots,x_n]$ be the ideal generated by $\mathcal{F}^{\text{top}}\cup\{x_1^q,\dots,x_n^q\}$. Let $\mathcal{L}=(x_1^q,\dots,x_n^q)+L$ be the $(q,\dots,q;D)$-LPP ideal such that

$$\dim(I_D)=\dim(\mathcal{L}_D)=m+\dim(x_1^q,\dots,x_n^q)_D.$$

Hence $\mathcal{L}$ is minimally generated by $x_1^q,\dots,x_n^q$ and m monomials of degree D that do not belong to $(x_1^q,\dots,x_n^q)$. Recall that, by assumption, L is the largest lex-segment ideal generated in degree D such that $\mathcal{L}=(x_1^q,\dots,x_n^q)+L$, i.e., $L_D\supseteq(x_1^q,\dots,x_n^q)_D$.

Let u be the smallest degree D monomial in $\mathcal{L}/(x_1^q,\dots,x_n^q)$. Notice that u is a $(D,k,q-t)$-type monomial, since m belongs to the interval $(\sigma_{k,t-1},\sigma_{k,t}]$. Let v be the smallest degree D monomial in L. If u is the smallest monomial in $(R/(x_1^q,\dots,x_n^q))_D$, then $u=x_k^{q-t}x_{k+1}^{q-1}\cdots x_n^{q-1}$ and $D=q-t+(n-k)(q-1)$. Moreover, $\mathcal{L}_D=R_D$ and $v=x_n^D$ is a (D,n,D)-type monomial.

Assume now that u is not the smallest monomial in $(R/(x_1^q,\dots,x_n^q))_D$ and let w be the monomial in $(R/(x_1^q,\dots,x_n^q))_D$ which follows u in decreasing lexicographic order. Then $u\geq v\geq w$ and v is the degree D monomial next to w in increasing lexicographic order. If $m\neq\sigma_{k,t}$, then w has type $(D,k,q-t)$, hence so does v by Lemma 33. Suppose therefore that $m=\sigma_{k,t}$. Write $D-q+t=(n-\ell)(q-1)+r$, where $0\leq r<q-1$. Notice that $D-q+t<(n-k)(q-1)$, since u is not the smallest monomial in $(R/(x_1^q,\dots,x_n^q))_D$. Then $\ell>k$. In this situation, $u=x_k^{q-t}x_\ell^r x_{\ell+1}^{q-1}\cdots x_n^{q-1}$ and $w=x_k^{q-t-1}x_{k+1}^{q-1}\cdots x_{k+n-\ell}^{q-1}x_{k+n-\ell+1}^{r+1}$. Notice that w is a $(D,k,q-t-1)$-type monomial if $t\neq q-1$ and a $(D,k+1,q-1)$-type monomial if $t=q-1$. If w is not the smallest degree D monomial of its type in R_D, then v has the same type as w. If w is the smallest degree D monomial of its type in R_D, then $w=x_k^{q-t-1}x_{k+1}^{D-q+t+1}$ in the case $t\neq q-1$ and $w=x_{k+1}^{q-1}x_{k+2}^{D-q+1}$ in the case $t=q-1$. Since $w\notin(x_1^q,\dots,x_n^q)$, this is only possible if $D\leq 2q-t-2$ for $t\neq q-1$ and $D\leq 2q-2$ for $t=q-1$. In this situation, v has type $(D,k,q-t)$ in the case $t\neq q-1$ and type $(D,k+1,q)$ in the case $t=q-1$.

If Conjecture 19 holds, then

$$d_{\text{reg}}(\mathcal{F}\cup\{x_1^q-x_1,\dots,x_n^q-x_n\})\leq\text{reg}(\mathcal{L}). \tag{4}$$

Moreover, by Proposition 20

$$\text{reg}(\mathcal{L})=\begin{cases} q-t-1+(n-k)(q-1) & \text{if } m=\sigma_{k,t}\text{ and}\\ & \text{either } t\neq q-1\text{ and } D\geq 2q-t-1,\\ & \text{or } t=q-1\text{ and } D\geq 2q-1,\\ \\ q-t+(n-k)(q-1) & \text{else.}\end{cases} \tag{5}$$

The bound on the degree or regularity now follows from (4) and (5). The bounds on the solving degree follow from the bound on the degree of regularity, Theorem 3, and Theorem 4. □

5 Applications to the study of GeMSS and Rainbow

GeMSS and Rainbow were the only multivariate schemes in Round 3 of the NIST Post-Quantum Cryptography Standardization process. They are based on modifications of HFE (Hidden Field Equation) and UOV (Unbalanced Oil and Vinegar), respectively. They were subsequently revealed to be insecure and susceptible to a MinRank Attack, see e.g. [3, 8, 9, 1], therefore they were excluded from the NIST competition.

In this section, we demonstrate how to use our results to show that these systems are far from being algebraically random. We do so by computing their degree of regularity or solving degree for small instances and comparing it with the upper bounds that we obtained in Theorem 22 and Theorem 34. The degree of regularity and solving degree of the systems associated to GeMSS and Rainbow are much smaller than the corresponding invariants for an algebraically random system with the same parameters, which reveals the presence of a hidden structure that may be used to mount an ad-hoc attack, as it was done in practice with the MinRank Attacks mentioned above.

We used Magma to compute the solving degree and Singular to compute the degree of regularity. The values that we obtain (and that we indicate in the tables below) are almost always the same for systems with the same parameters. For each choice of the parameters in the table, we produce ten instances of the public key $\mathcal{PK} = \{p_1, \dots, p_m\} \subseteq \mathbb{F}_q[x_1, \dots, x_n]$ of the chosen scheme. For each one of them, we choose a random vector $s = (s_1, \dots, s_m) \in \mathbb{F}_q^m$ as a signature (in case the chosen vector is not a valid signature, we replace it with another randomly chosen vector). In order to forge the signature, an attacker may want to solve the system $\mathcal{PK}s = \{p_1 - s_1, \dots, p_m - s_m\}$. We make the system $\mathcal{PK}s$ square by assigning random values to the last $n - m$ variables. This yields a system $\mathcal{F} = \{f_1, \dots, f_m\} \subseteq \mathbb{F}_q[x_1, \dots, x_m]$.

Since in GeMSS we work over $\mathbb{F}_2$, we add to the system $\mathcal{F}$ the field equations $\mathcal{E} = \{x_1^2 + x_1, \dots, x_m^2 + x_m\}$. In the next table we compare the experimental results we obtained for GeMSS with the bounds from Theorem 34. The experiments show that both the solving degree and the degree of regularity of $\mathcal{F}$ can be more than twice the solving degree of $\mathcal{F} \cup \mathcal{E}$. This confirms the intuition that adding the field equations is a good strategy in order to solve the system $\mathcal{F}$ over $\mathbb{F}_2$.

Unfortunately we were able to compute the degree of regularity of $\mathcal{F} \cup \mathcal{E}$ only for small values of the parameters. In the next table the first three columns contain the parameters of the cryptosystem, and the fourth the number of polynomials and variables that appear in $\mathcal{F}$. The columns labelled '$d_{\text{reg}}(\mathcal{F} \cup \mathcal{E})$' and '$\text{solv. deg}(\mathcal{F} \cup \mathcal{E})$' contain the values computed with Magma and Singular. The columns labelled '$\max \text{solv. deg}(\mathcal{F} \cup \mathcal{E})$' and '$\max d_{\text{reg}}(\mathcal{F} \cup \mathcal{E})$' are the bounds given by the Theorem 34 for the chosen parameters and a standard Gröbner basis algorithm.

n, D	a	v	m	$d_{\text{reg}}(\mathcal{F} \cup \mathcal{E})$	$\max d_{\text{reg}}(\mathcal{F} \cup \mathcal{E})$	$\text{solv. deg}(\mathcal{F} \cup \mathcal{E})$	$\max \text{solv. deg}(\mathcal{F} \cup \mathcal{E})$
12, 4	1	1	11	5	10	3	18
8, 9	1	1	7	3	6	3	10
8, 9	1	2	7	3	6	3	10
8, 9	2	1	6	3	5	3	8
8, 9	2	2	6	3	5	3	8
24, 4	1	1	23		22	4	42
24, 4	1	2	23		22	4	42
24, 4	1	3	23		22	4	42
24, 4	2	1	22		21	4	40
24, 4	2	2	22		21	4	40
24, 4	3	1	21		20	4	38

While GeMSS is random according to Definition 12, the experimental results make it clear that both the degree or regularity and the solving degree of GeMSS are far from the largest values that one can find for a system of those parameters according to Theorem 34. This indicates that, for the same parameters, one should be able to find systems for which the complexity of computing a Gröbner basis is much larger. More importantly, it reveals the presence of a hidden algebraic structure, which may be exploited in ad-hoc attack (as it was in fact done in the MinRank attacks mentioned in the opening paragraph).

For Rainbow, we choose to work over $\mathbb{F}_4$ and $\mathbb{F}_9$. Since $\mathcal{F}$ is a square system, $\mathcal{F}$ is algebraically random if and only if $\mathcal{F}^{\text{top}}$ is a regular sequence. This turns out to be the case in most of the examples that we computed and in that case

$$d_{\text{reg}}(\mathcal{F}) = m + 1 \tag{6}$$

by Theorem 22. This is confirmed by our computations.

Since the systems coming from this scheme are quadratic, adding the field equations may increase the solving degree of the system. However, for the small values of q that we tried in our experiments, we find that in all cases but one the solving degree decreases when adding the field equations. This makes sense, as the degree q of the equations that we add is never larger than the solving degree of the system to which we add them. In the next table we summarize the results that we obtained in our computational experiments. Since in our experiments the solving degree of $\mathcal{F} \cup \mathcal{E}$ is almost always smaller than that of $\mathcal{F}$, in our examples $\mathcal{F} \cup \mathcal{E}$ is the relevant system to consider, that is, the system that one wants to try to solve. Therefore, we consider the degree of regularity and solving degree of $\mathcal{F} \cup \mathcal{E}$. In our table, we compare the degree of regularity and the solving degree of $\mathcal{F} \cup \mathcal{E}$ with the upper bounds from Theorem 22 and Corollary 28. We use Corollary 28 as the algorithm implemented in Magma is a standard one. The first three columns contain the chosen values for the parameters and the number of polynomials and variables that appear in $\mathcal{F}$. The columns labelled '$d_{\text{reg}}(\mathcal{F})$' and 'solv. deg$(\mathcal{F} \cup \mathcal{E})$' contain the values computed with Magma and Singular. The column labelled 'max solv. deg$(\mathcal{F} \cup \mathcal{E})$' contains the bounds from Corollary 28. We do not include the values of the degree of regularity and solving degree of $\mathcal{F}$ in the table, as experimentally we find that the system $\mathcal{F} \cup \mathcal{E}$ can always be solved more efficiently than the system $\mathcal{F}$.

q	$[v_1, o_1, o_2]$	m	$d_{\text{reg}}(\mathcal{F})$	$d_{\text{reg}}(\mathcal{F} \cup \mathcal{E})$	solv. deg$(\mathcal{F} \cup \mathcal{E})$	max solv. deg$(\mathcal{F} \cup \mathcal{E})$
4	$[3, 2, 2]$	4	4/5	4	4	8
4	$[3, 3, 3]$	6	6/7	5	5	12
4	$[7, 5, 5]$	10	10/11	6	6	20
9	$[3, 2, 2]$	4	5	5	9	–
9	$[7, 5, 5]$	10	11	9/10	10	20

Notice that Corollary 28 does not apply to the case $q = 9$ and $[v_1, o_1, o_2] = [3, 2, 2]$, since $d_{\text{reg}}(\mathcal{F}) = 5$. For these parameters, the bound from Corollary 28 would yield solv. deg$(\mathcal{F} \cup \mathcal{E}) \leq 8$. However, the bound does not hold in this case, as our experiments show.

As for GeMSS, we observe that the values that we computed for the solving degree of $\mathcal{F} \cup \mathcal{E}$ are far from the upper bounds predicted by Corollary 28. We conclude that, also in this case, one expects to find systems with the same parameters as these instances of Rainbow and for which the complexity of computing a Gröbner basis is larger. More importantly, once again it reveals the presence of a hidden algebraic structure, which may be exploited in ad-hoc attack (as it was in fact done in the attacks that we mentioned at the start of the section).

References

[1] John Baena, Pierre Briaud, Daniel Cabarcas, Ray Perlner, Daniel Smith-Tone, and Javier Verbel, *Improving Support-Minors Rank Attacks: Applications to GeMSS and Rainbow*, Advances in Cryptology – CRYPTO 2022, Springer (2022), 376–405.

[2] Magali Bardet, *Étude des systèmes algébriques surdéterminés. Applications aux codes correcteurs et à la cryptographie*, PhD thesis, Université Pierre et Marie Curie - Paris VI, 2004.

[3] Magali Bardet, Maxime Bros, Daniel Cabarcas, Philippe Gaborit, Ray Perlner, Daniel Smith-Tone, Jean-Pierre Tillich, and Javier Verbel, *Improvements of Algebraic Attacks for Solving the Rank Decoding and MinRank Problems*, Advances in Cryptology – ASIACRYPT 2020, Springer (2020), 507–536.

[4] Magali Bardet, Jean-Charles Faugere, and Bruno Salvy, *Complexity of Gröbner basis computation for Semi-regular Overdetermined sequences over $\mathbb{F}_2$ with solutions in $\mathbb{F}_2$*, PhD thesis, INRIA, 2003.

[5] Magali Bardet, Jean-Charles Faugère, and Bruno Salvy, *On the complexity of Gröbner basis computation of semi-regular overdetermined algebraic equations*, Proceedings of the International Conference on Polynomial System Solving 2004, 71-74.

[6] Magali Bardet, Jean-Charles Faugère, and Bruno Salvy, *On the complexity of the F5 Gröbner basis algorithm*, Journal of Symbolic Computation **70** (2015), 49-70.

[7] Magali Bardet, Jean-Charles Faugère, Bruno Salvy, and Bo-Yin Yang, *Asymptotic behaviour of the degree of regularity of semi-regular polynomial systems*, Proceedings of MEGA **5**, 2005.

[8] Ward Beullens, *Improved Cryptanalysis of UOV and Rainbow*, Advances in Cryptology – EUROCRYPT 2021, Springer (2021), 348–373.

[9] Ward Beullens, *Breaking rainbow takes a weekend on a laptop*, Annual International Cryptology Conference, Springer (2022), 464-479.

[10] Mina Bigdeli, Emanuela De Negri, Manuela Muzika Dizdarevic, Elisa Gorla, Romy Minko, and Sulamithe Tsakou, *Semi-regular sequences and other random systems of equations*, Women in Numbers Europe III, Springer (2021), 75-114.

[11] Bruno Buchberger, *Bruno Buchberger's PhD thesis 1965: An algorithm for finding the basis elements of the residue class ring of a zero dimensional polynomial ideal*, Journal of Symbolic Computation **41** (2006), 475–511.

[12] Johannes A. Buchmann, Jintai Ding, Mohamed Saied Emam Mohamed, and Wael Said Abd Elmageed Mohamed, *MutantXL: Solving multivariate polynomial equations for cryptanalysis*, Dagstuhl seminar proceedings (2009).

[13] Alessio Caminata and Elisa Gorla, *Solving multivariate polynomial systems and an invariant from commutative algebra*, 8th International Workshop, WAIFI 2020, Springer (2021), 3-36.

[14] Alessio Caminata and Elisa Gorla, *Solving degree, last fall degree, and related invariants*, Journal of Symbolic Computation **114** (2023), 322-335.

[15] Antoine Casanova, Jean-Charles Faugère, Gilles Macario-Rat, Jacques Patarin, Ludovic Perret, and Jocelyn Ryckeghem, *GeMSS: a great multivariate short signature*, UPMC-Paris 6 Sorbonne Universités; INRIA Paris Research Centre, 2017.

[16] Giulio Caviglia and Alessandro De Stefani, *Linearly presented modules and bounds on the Castelnuovo-Mumford regularity of ideals*, Proceedings of the American Mathematical Society **150** (2022), no.4, 1397-1404.

[17] Giulio Caviglia, Alessandro De Stefani, and Enrico Sbarra, *The Eisenbud-Green-Harris Conjecture*, Commutative algebra - Expository Papers Dedicated to David Eisenbud on the Occasion of his 75th Birthday, Springer (2021), 159-187.

[18] Nicolas Courtois, Alexander Klimov, Jacques Patarin, and Adi Shamir, *Efficient Algorithms for Solving Overdefined Systems of Multivariate Polynomial Equations*, Advances in Cryptology – EUROCRYPT 2000, Springer (2000), 392-407.

[19] David Eisenbud, Mark Green, and Joe Harris, *Higher Castelnuovo Theory*, Astérisque **218** (1993), 187-202.

[20] Jean-Charles Faugère, *A new efficient algorithm for computing Gröbner bases (F4)*, Journal of Pure and Applied Algebra **139** (1999), 61-88.

[21] Jean-Charles Faugère, *A new efficient algorithm for computing Gröbner bases without reduction to zero (F5)*, Proceedings of the 2002 International Symposium on Symbolic and Algebraic Computation (ISSAC '02), 75-83.

[22] Ralf Fröberg, *An inequality for Hilbert series of graded algebras*, Mathematica Scandinavica **56** (1985), 117–144.

[23] Timothy Hodges, Sergio Molina, and Jacob Schlather, *On the existence of semi-regular sequences*, Journal of Algebra **476** (2017), 519-547.

[24] Keith Pardue, *Generic polynomials*, preprint (1999).

[25] Keith Pardue, *Generic sequences of polynomials*, Journal of Algebra **324** (2010), no. 4, 579-590.

[26] Flavio Salizzoni, *An upper bound for the solving degree in terms of the degree of regularity*, preprint available at https://arxiv.org/abs/2304.13485.

[27] Igor Semaev and Andrea Tenti, *Probabilistic analysis on Macaulay matrices over finite fields and complexity of constructing Gröbner bases*, Journal of Algebra **565** (2021), 651-674.

[28] Andrea Tenti, *Sufficiently overdetermined random polynomial systems behave like semiregular ones*, PhD Thesis, University of Bergen (2019), available at https://hdl.handle.net/1956/21158.