Asymptotic Behaviour of the Degree of Regularity of Semi-Regular Polynomial Systems

M. Bardet* J.-C. Faugère* B. Salvy† B-Y. Yang‡

Abstract

We compute the asymptotic expansion of the degree of regularity for overdetermined semi-regular sequences of algebraic equations. This degree implies bounds for the generic complexity of Gröbner bases algorithms, in particular the F_5 [Fau02] algorithm. Bounds can also be derived for the XL [SPCK00] family of algorithms used by the cryptographic community.

1 Motivations and Results

The worst-case complexity of Gröbner bases has been the object of extensive studies. In the most general case, it is well known after work by Mayr and Meyer that the complexity is doubly exponential in the number of variables. For subclasses of polynomial systems, the complexity may be much smaller. Of particular importance is the class of regular sequences of polynomials. There, it is known that after a generic linear change of variables the complexity of the computation for the degree-reverse-lexicographic order is simply exponential in the number of variables. Moreover, in characteristic 0, these systems are generic. Our goal is to give similar complexity bounds for *overdetermined* systems, for a class of systems that we call *semi-regular*.

The interest in overdetermined systems is not purely academic: there are a number of applications, such as error correcting codes (decoding of cyclic codes), robotics, calibration, cryptography,.... The security of many cryptographical primitives depends on the difficulty of system-solving. Sometimes (in the case of "multivariate public-key cryptosystems") the public keys themselves become the system to be solved. Sometimes primitives can be cracked if one can find a solution to an associated overdetermined system of algebraic equations over a finite field. This is known as Algebraic Cryptanalysis and is currently one of the "hot" topics in cryptography.

*LIP6, 8 rue du Capitaine Scott, F-75015 PARIS, {Magali.Bardet,Jean-Charles.Faugere}@lip6.fr

†INRIA Rocquencourt Bat. 9, Domaine de Voluceau, BP 105, F-78153 Le Chesnay Cedex, Bruno.Salvy@inria.fr

‡Mathematics Department, Tamkang University, Tamsui, Taiwan 251, by@moscito.org; research also sponsored by the National Science Council under the Taiwan Information Security Center (TWISC) project.

In most cases, only the solutions over a finite field are required, rather than solutions in the algebraic closure. Often the finite field is $\mathbb{F}_2$, and we may then think of the problem as solving the original system of, say m, equations over $\mathbb{F}_2$ together with the field equations $x_i^2 = x_i$ $(i = 1, \ldots, n)$. We would then have an overdetermined system of $m + n$ equations. For larger fields, the n field equations are of higher degree and the solution process is then affected to a lesser extent.

Gröbner bases algorithms are rather little known by the cryptographers, who prefer to use algorithms like Algorithm XL [SPCK00] (rediscovered in 1999 as an adapted version over finite fields of Lazard's proposed method of 1983 [Laz83]) and its variants. Since XL can be seen as a particular case of Gröbner bases algorithms [AFI$^+$04], the bounds for XL are at least equal to the bounds derived in this paper (see also [YC04] for a specific study).

We now state more precisely our results. We consider polynomials $(f_1, \ldots, f_m)$ in $\mathbb{K}[\mathbf{x}] = \mathbb{K}[x_1, \ldots, x_n]$ where $\mathbb{K}$ is a field. We denote by d_i the total degree of f_i, and by $\langle f_1, \ldots, f_m \rangle$ the ideal generated by the f_i's.

The Hilbert series of this ideal is well known to be related to its Gröbner bases for orders that refine the degree. In the case of a regular system this series is

$$S_{m,n} = \frac{\prod_{i=1}^{m}(1 - z^{d_i})}{(1 - z)^n}.$$

The degree of regularity d_{reg} of the series is the smallest D such that the coefficient of z^i in the series $S_{m,n}$ is equal to the value of the Hilbert polynomial at i for all $i \geq D$. This is precisely the highest degree in elements of a Gröbner basis for an order that refines the degree, after a generic linear change of variables [Laz83, Giu84]. Easy manipulations on series give for D the value we call the *Macaulay bound*:

$$d_{\text{reg}} = \sum_{i=1}^{m}(d_i - 1) + 1. \tag{1}$$

When the number of polynomials m is larger than the number n of variables, the series $S_{m,n}$ has negative coefficients. It turns out that for the semi-regular systems we consider, the degree of regularity is then found to be the index of the first non-positive coefficient in $S_{m,n}(z)$. When working over $\mathbb{F}_2[\overline{x_1}, \ldots, \overline{x_n}] = \mathbb{F}_2[\mathbf{x}]/\langle x_1^2 - x_1, \ldots, x_n^2 - x_n \rangle$, we have to work with the modified generating series

$$T_{m,n}(z) = (1 + z)^n \Big/ \prod_{i=1}^{m}(1 + z^{d_i}).$$

Again, the degree of regularity is the index of the first non-positive coefficient and it is a bound for the highest degree of elements of a Gröbner basis for any order refining the degree. The generating series $S_{m,n}(z)$ and $T_{m,n}(z)$ and associated degrees of regularity have also appeared recently in cryptography to analyse the XL algorithm and its variant XL2 [YC04]. In [Fau02, BFS03, Bar04], we have shown that in all three cases (regular, semi-regular, semi-regular over $\mathbb{F}_2[\overline{x_1}, \ldots, \overline{x_n}]$), the Gröbner basis algorithm F_5 does not perform any reduction to 0 before degree d_{reg}. This leads to complexity estimates in terms of the complexity of linear algebra in dimension the number of monomials of degree at most d_{reg}.

In the case of regular sequences, using the Macaulay bound then gives a very precise complexity estimate in terms of the degrees d_i and the number n of variables. While we are not able to give such a simple formula in the overdetermined case, we give an asymptotic analysis of d_{reg}. For simplicity, we restrict to the quadratic case ($d_i = 2$), and refer to [BFS04, Bar04] for more general results and sketch of the proof. Our main results can now be stated.

Theorem 1. *For $m = n + k$ ($k > 1$ fixed) quadratic equations in n variables, the degree of regularity d_{reg} behaves asymptotically like*

$$d_{reg} = \frac{m}{2} - h_{k,1}\sqrt{\frac{m}{2}}(1 + o(1)), \tag{2}$$

where H_k denotes the Hermite polynomial of order k and $h_{k,1}$ is the largest zero of H_k.

For $m = \alpha n$ ($\alpha > 1$ fixed) quadratic equations in n variables, the degree of regularity d_{reg} behaves asymptotically like

$$d_{reg} = (\alpha - \frac{1}{2} - \sqrt{\alpha(\alpha-1)})n + \frac{-a_1}{2(\alpha(\alpha-1))^{\frac{1}{6}}}n^{\frac{1}{3}} - \left(2 - \frac{2\alpha - 1}{4(\alpha(\alpha-1))^{\frac{1}{2}}}\right) + O(\frac{1}{n^{1/3}}), \tag{3}$$

where $a_1 \approx -2.3381$ is the largest zero of the classical Airy function.

For $m = \alpha n$ ($\alpha > 1$ fixed) quadratic equations in n variables in $\mathbb{F}_2[\overline{x_1}, \ldots, \overline{x_n}]$, the degree of regularity d_{reg} behaves asymptotically like

$$d_{reg} \sim \left(-\alpha + \frac{1}{2} + \frac{1}{2}\sqrt{2\alpha^2 - 10\alpha - 1 + 2(\alpha+2)\sqrt{\alpha(\alpha+2)}}\right) n. \tag{4}$$

Intuitively, these results give a quantification of the gain obtained by adding more and more information in the form of new equations.

These asymptotics results show that the logarithm of the complexity in the semi-regular case is dominated by a linear term in n when $m \sim \alpha n$, hence is simply exponential (the number of monomials with n variables at degree $D = (c + o(1))n$ is simply exponential in n, even when considering the field equations). See also [Die04] about previous conjectures by cryptographers that XL may be able to solve the multivariate quadratic problem in sub-exponential time.

These results also allow to quantify the consequences of the Frobenius criterion: consider a sequence $(f_1, \ldots, f_n, x_1^2 - x_1, \ldots, x_n^2 - x_n) \subset \mathbb{K}[\mathbf{x}]$, the degree of regularity is given by

$$\begin{aligned} d_{\mathrm{reg}} &= 0.086\,n + 1.04\,n^{\frac{1}{3}} - 1.47 + O(n^{-\frac{1}{3}}) \text{ if } \mathbb{K} \text{ has characteristic } 0, \\ d_{\mathrm{reg}} &= 0.09n + 1.00n^{\frac{1}{3}} - 1.58 + O(n^{-\frac{1}{3}}) \text{ if } \mathbb{K} \text{ has characteristic } 2. \end{aligned}$$

This article is structured as follows. In Section 2 we recall the definitions and properties of regular, semi-regular sequences and semi-regular sequences in $\mathbb{F}_2[\overline{x_1}, \ldots, \overline{x_n}]$. Then in Section 3, we give the proofs of the asymptotic expansions of d_{reg} in the three cases presented above.

2 Regular and semi-regular systems

We consider polynomials $(f_1, \dots, f_m)$ in $\mathbb{K}[\mathbf{x}]$ where $\mathbb{K}$ is a field. We denote by d_i the total degree of f_i, by f_i^h the homogeneous part of highest degree of f_i and by $\mathcal{I} = \langle f_1, \dots, f_m \rangle$ the ideal generated by the f_i's.

2.1 Regular sequences

Algebraic properties of regular sequences [Mac16] are well known (Hilbert series, index of regularity, ... [CLO98, Lan02, Frö97]) and their behavior w.r.t. Gröbner bases computation is well understood [Giu84, Laz83]. Moreover, if the field $\mathbb{K}$ has characteristic zero, regular sequences are *generic* among all sequences (the integers n, m and d_i being fixed), that is in the space of all sequences, non-regular sequences form an algebraic set of codimension at least 1.

We recall definitions and properties of regular sequences. Geometrically, the system $(f_1, \dots, f_m)$ of homogeneous equations is regular when for each $i = 1, \dots, m$, the algebraic set defined by $(f_1, \dots, f_i)$ has codimension i. Algebraically, this is expressed by the fact that f_i is not a zero-divisor in the quotient $\mathbb{K}[\mathbf{x}]/\langle f_1, \dots, f_{i-1} \rangle$. Regular sequences can also be characterized by the set of relations between the f_i's: regular sequences can be viewed as sequences for which no relation but the trivial ones (generated by $f_i f_j = f_j f_i$) occurs.

We slightly restrict the usual definition of regular sequences in the affine case so that our complexity results can apply. This restriction is discussed in Section 2.2.

Definition 2. *A homogeneous sequence of polynomials $(f_1, \dots, f_m)$ is regular if for all $i = 1, \dots, m$ and g such that*

$$g f_i \in \langle f_1, \dots, f_{i-1} \rangle$$

then g is also in $\langle f_1, \dots, f_{i-1} \rangle$.

An affine sequence of polynomials $(f_1, \dots, f_m)$ is regular if the homogeneous sequence $(f_1^h, \dots, f_m^h)$ is, where f_i^h is the homogeneous part of f_i of highest degree.

Classical properties of *homogeneous* regular systems are:

Theorem 3. *(i) $(f_1, \dots, f_m)$ is regular if and only if its Hilbert series is given by*

$$\frac{\Pi_{i=1}^m (1 - z^{d_i})}{(1 - z)^n} \tag{5}$$

(ii) after a generic linear change of variables, the highest degree of elements of a Gröbner basis for the DRL order is bounded by the index of regularity

$$\sum_{i=1}^m (d_i - 1) + 1$$

(iii) $(f_1, \dots, f_m)$ is regular if and only if there are no reduction to 0 in Algorithm F_5,

Proof. The proof of the property (*ii*) can be found in [Laz83, Giu84]. Property (*iii*) is proved in [Fau02]. The property (*i*) follows directly from [Lan02, Theorem 6.6 p. 436]; see also [Frö97, p. 137]. □

2.2 Linear algebra, Gröbner basis algorithms and Algorithm XL

The link between polynomial system solving and linear algebra was described by Macaulay in [Mac16] where he generalized Sylvester's matrix (for the resultant of two univariate polynomials) to multivariate polynomials. The idea is to construct a matrix in degree d whose lines contain all multiples of the polynomials f_i $(i = 1, \ldots, n)$ in the original system by monomials t such that $\deg(tf_i) \leq d$, the columns representing a basis of monomials up to degree d. It was observed by Lazard [Laz83] that for a large enough degree d, ordering the columns according to a monomial ordering and performing row reduction without column pivoting on the matrix (a particular Gaussian Elimination) is equivalent to Buchberger's Gröbner basis algorithm.

The XL algorithm was designed to solve a system of multivariate polynomials that has only one solution over a finite field. It constructs the Macaulay matrix in a given degree and solves the resultant system using sparse matrix methods. There are several variants of this algorithm (e.g. XL2). It can be shown that at the degree of regularity d_{reg}, a semi-regular system (see definition in the next Section) will be solved using XL2 [YC04].

One of the main difficulties with this Macaulay matrix is that many rows are linearly dependent upon the previous ones and a lot of time is wasted to produce 0 during the Gaussian Elimination. Faugère's F_5 criterion [Fau02] can be used to avoid useless rows in the Macaulay matrix coming from the relations $f_i f_j = f_j f_i$. The matrix version of the F_5 algorithm [Bar04] constructs incrementally in the degree, then in the number of polynomials a submatrix of the Macaulay matrix in degree d that is full rank for regular sequences and for semi-regular sequences as $d < d_{\text{reg}}$. The algorithm stops when a large enough degree has been reached, which is d_{reg} for semi-regular homogeneous sequences.

For affine sequences, the F_5 criterion applies without any changes in a matrix version of F_5 as long as there is no *fall of degree*, which is equivalent to a reduction to 0 for the homogeneous part of highest degree of the polynomials. This justify our definition of regular (and semi-regular) sequences for affine systems. For an affine regular sequence, we can just run the F_5 matrix algorithm up to degree d_{reg}, and then end the computation by running another algorithm like F_4 [Fau99] for instance. The rate-determining step is the first part.

For sequences over $\mathbb{F}_2$ containing the field equations $x_i^2 = x_i$, the matrices constructed by F_5 are no longer full rank, because of the Frobenius morphism. Another criterion, called the *Frobenius criterion* [BFS03, Bar04], can be used to avoid useless rows in the Macaulay matrix coming from the relations $f_i f_i = f_i$. The F_5 algorithm in a matrix version with the Frobenius criterion constructs full rank matrices for semi-regular sequences over F_2.

2.3 Semi-regular sequences

Regular systems have at most as many polynomials as variables; we generalize this definition to overdetermined systems [Bar04, BFS04]:

Definition 4. *The degree of regularity of a homogeneous ideal $\mathcal{I} = \langle f_1, \ldots, f_m \rangle$ is defined by*

$$d_{reg} = \min \left\{ d \geq 0 \mid \dim_{\mathbb{K}}(\{f \in \mathcal{I}, \quad \deg(f) = d\}) = \binom{n+d-1}{d} = \# \begin{array}{c} \textit{monomials} \\ \textit{of degree } d \end{array} \right\}$$

This definition implies that for any monomial ordering refining the degree, all monomials in degree d_{reg} are leading terms for an element of the ideal. Thus d_{reg} is clearly an upper bound on the degree of the elements of a Gröbner basis for such a monomial ordering.

Definition 5. *A homogeneous sequence of polynomials $(f_1, \ldots, f_m)$ is semi-regular if for all $i = 1, \ldots, m$ and g such that*

$$g f_i \in \langle f_1, \ldots, f_{i-1} \rangle \textbf{ and } \deg(g f_i) < d_{reg}$$

then g is also in $\langle f_1, \ldots, f_{i-1} \rangle$.

An affine sequence of polynomials $(f_1, \ldots, f_m)$ is semi-regular if the sequence $(f_1^h, \ldots, f_m^h)$ is semi-regular, where f_i^h is the homogeneous part of f_i of highest degree.

Properties of semi-regular sequences are:

Proposition 6. *Let $(f_1, \ldots, f_m)$ be a sequence of m polynomials in n variables, f_i being of degree d_i. Then:*

(i) The sequence $(f_1, \ldots, f_m)$ is semi-regular if and only if the Hilbert series of the homogeneous sequence $(f_1^h, \ldots, f_m^h)$ is given by

$$\left[S_{m,n}(z) \right],$$

where $\left[\sum_{i \geq 0} a_i z^i\right] = \sum_{i \geq 0} b_i z^i$ with $b_i = a_i$ if $a_j > 0 \; \forall 0 \leq j \leq i$ and $b_i = 0$ otherwise.

(ii) For $m \leq n$, the sequence $(f_1, \ldots, f_m)$ is regular if and only if it is semi-regular. In other words, the notion of semi-regularity coincides with the notion of regularity.

(iii) The degree of regularity of a semi-regular sequence $(f_1, \ldots, f_m)$ is the index of the first non-positive coefficient in the series $S_{m,n}(z)$.

(iv) For a semi-regular system, there is no reduction to 0 in Algorithm F_5 for degrees smaller than d_{reg}. Moreover, the total number of arithmetic operations in $\mathbb{K}$ performed by F_5 (matrix version) is bounded by

$$O\left(\binom{n + d_{reg}}{n}^{\omega} \right).$$

Proof. We prove property (i) for homogeneous equations. Consider the exact sequence

$$0 \to (\mathbb{K}[\mathbf{x}]/\langle f_1,\ldots,f_{i-1}\rangle)_{d-d_i} \xrightarrow{f_i} (\mathbb{K}[\mathbf{x}]/\langle f_1,\ldots,f_{i-1}\rangle)_d \to (\mathbb{K}[\mathbf{x}]/\langle f_1,\ldots,f_i\rangle)_d \to 0$$

then as long as $d < d_{\text{reg}}$ the associated Hilbert functions verify the relation [CLO98]

$$HF_{\langle f_1,\ldots,f_{i-1}\rangle}(d-d_i) - HF_{\langle f_1,\ldots,f_{i-1}\rangle}(d) + HF_{\langle f_1,\ldots,f_i\rangle}(d) = 0$$

for all $d < d_{\text{reg}}$. Moreover, $HF_{\langle f_1,\ldots,f_i\rangle}(d) = 0$ for all i and d and $HF_{\langle 0\rangle}(d) = \binom{n+d-1}{d}$ which implies the following relations for the Hilbert series:

$$HS_{\langle f_1,\ldots,f_m\rangle}(z) = \sum_{d=0}^{\infty} HF_{\langle f_1,\ldots,f_m\rangle}(d) z^d = \left[\prod_{i=1}^{m}(1-z^{d_i})\Big/(1-z)^n\right].$$

Conversely, consider the exact sequence

$$0 \to K_{d-d_i} \to (\mathbb{K}[\mathbf{x}]/\langle f_1,\ldots,f_{i-1}\rangle)_{d-d_i} \xrightarrow{f_i} (\mathbb{K}[\mathbf{x}]/\langle f_1,\ldots,f_{i-1}\rangle)_d \to (\mathbb{K}[\mathbf{x}]/\langle f_1,\ldots,f_i\rangle)_d \to 0$$

where K is the kernel of the multiplication map by f_i. For all $d < d_{\text{reg}}$ the kernel is necessary $K_{d-d_i} = \{0\}$, hence by Definition 5 the sequence is semi-regular.

Property (ii) is a consequence of (i) and Theorem 3 (i). By definition the degree of regularity of a homogeneous sequence is the first d for which $HF_{\langle f_1,\ldots,f_m\rangle}(d) = 0$, which proves property (iii). For property (iv) see [Bar04]. □

Let us mention another definition that extends the notion of regular sequences to overdetermined systems. In [PR03], the authors define semi-regular sequences as follows:

Definition 7 (Semi-regular sequences [PR03]). *A sequence of forms* $(f_1,\ldots,f_m)$ *of degrees* $(d_1,\ldots,d_m)$ *in* $\mathbb{K}[\mathbf{x}]$ *is called a semi-regular sequence if for all* $i = 1,\ldots,m$*, the multiplication map* $(\mathbb{K}[\mathbf{x}]/\langle f_1,\ldots,f_{i-1}\rangle)_{a-d_i} \xrightarrow{f_i} (\mathbb{K}[\mathbf{x}]/\langle f_1,\ldots,f_{i-1}\rangle)_a$ *are linear maps of maximal rank for all* a*.*

Semi-regular sequences according to our definition are more general than semi-regular sequences according to Definition 7: the latter ones have the property that any sub-sequence $f_1,\ldots,f_i$ of polynomials is also semi-regular, which is not true for our semi-regular sequences (e.g. $\{f_1 = x_1^2, f_2 = x_1x_2, f_3 = x_2^2\}$). As a consequence, property 1 from Theorem 6 is false for semi-regular sequences according to Pardue-Richert, but our complexity bounds still apply to their sequences.

2.4 Semi-regular sequences over $\mathbb{F}_2$

Consider now the case of a system $(f_1,\ldots,f_m)$ of m equations in n variables with coefficients in $\mathbb{F}_2$, together with the field equations $x_i(x_i-1) = 0$. Hence the system to be solved contains $m+n$ equations in n variables over the field $\mathbb{F}_2$. An additional difficulty comes from the property that in the quotient ring $\mathbb{F}_2[\overline{x_1},\ldots,\overline{x_n}] = \mathbb{F}_2[\mathbf{x}]/\langle x_1^2+x_1,\ldots,x_n^2+x_n\rangle$, every polynomial f belonging to the ideal $\langle f_1, \ldots, f_m \rangle$ is fixed by the Frobenius morphism $p \to p^2$, i.e. is a solution of the equation $f^2 = f$.

Hence we must slightly modify the definition of semi-regular sequence to take the Frobenius morphism into account. First, let us consider only homogeneous polynomials: we keep only the homogeneous part of greatest degree of the field equations x_i^2, then every homogeneous polynomial of degree d satisfies the relation $f^2 = 0$ in the quotient ring $\mathbb{F}_2[\mathbf{x}]/\langle x_1^2, \ldots, x_n^2 \rangle$. The degree of regularity d_{reg} is defined as before:

Definition 8. *The degree of regularity of a homogeneous ideal $\mathcal{I} = \langle f_1, \ldots, f_m \rangle$ in the quotient ring $\mathbb{F}_2[\mathbf{x}]/\langle x_1^2, \ldots, x_n^2 \rangle$ is defined by*

$$d_{reg} = \min \left\{ d \geq 0 \mid \dim_{\mathbb{F}_2}(\{f \in \mathcal{I}, \quad \deg(f) = d\}) = \binom{n}{d} = \# \begin{array}{l} \textit{square free} \\ \textit{monomials of degree } d \end{array} \right\}$$

Definition 9. *A homogeneous sequence $(f_1, \ldots, f_m) \subset \mathbb{F}_2[\mathbf{x}]/\langle x_1^2, \ldots, x_n^2 \rangle$ is semi-regular over $\mathbb{F}_2$ if for all $i = 1, \ldots, m$ and g such that*

$$g f_i \in \langle f_1, \ldots, f_{i-1} \rangle \textbf{ and } \deg(g f_i) < d_{reg}$$

then g is also in $\langle f_1, \ldots, f_{i-1}, f_i \rangle$

An affine sequence of polynomials $(f_1, \ldots, f_m) \subset \mathbb{F}_2[\overline{x_1}, \ldots, \overline{x_n}]$ is semi-regular over $\mathbb{F}_2$ if the homogeneous sequence $(f_1^h, \ldots, f_m^h) \subset \mathbb{F}_2[\mathbf{x}]/\langle x_1^2, \ldots, x_n^2 \rangle$ is semi-regular over $\mathbb{F}_2$, where f_i^h is the homogeneous part of f_i of highest degree.

Remark: Definition 5 says that for semi-regular sequences, the only polynomials g such that $g f_i \in \langle f_1, \ldots, f_{i-1} \rangle$ are those belonging to $\langle f_1, \ldots, f_{i-1} \rangle$ (together with a condition on the degrees). But in $\mathbb{F}_2[\mathbf{x}]/\langle x_1^2, \ldots, x_n^2 \rangle$ every polynomial f_i verifies $f_i f_i = 0$. This explains the difference between Definitions 5 and 9.

A modified version of Algorithm F_5 so that useless relations are not computed is described in [Bar04]. With this definition and the new F_5 criterion, properties of semi-regular sequences are preserved:

Proposition 10. *Let $(f_1, \ldots, f_m) \subset \mathbb{F}_2[\overline{x_1}, \ldots, \overline{x_n}]$ be a sequence of m polynomials in n variables, f_i being of degree d_i. Then:*

(i) The sequence $(f_1, \ldots, f_m)$ is semi-regular over $\mathbb{F}_2$ if and only if the Hilbert series of the homogeneous sequence $(f_1^h, \ldots, f_m^h) \subset \mathbb{F}_2[\mathbf{x}]/\langle x_1^2, \ldots, x_n^2 \rangle$ is given by

$$\Big[T_{m,n}(z) \Big],$$

(ii) If $(f_1, \ldots, f_m)$ is a semi-regular sequence over $\mathbb{F}_2$ then its degree of regularity is the index of the first non-positive coefficient in the series $T_{m,n}(z)$.

(iii) For a semi-regular sequence over $\mathbb{F}_2$, there is no reduction to 0 in Algorithm F_5 for degrees smaller than d_{reg}. Moreover, the total number of arithmetic operations in $\mathbb{F}_2$ performed by F_5 (matrix version including the Frobenius criterion) is bounded by

$$O\left(\binom{n}{d_{reg}}^{\omega}\right)$$

Where the exponent $\omega < 2.39$ is the exponent in the complexity of matrix multiplication.

Proof. The proof of property (i) is almost the same as for Theorem 6. The exact sequence is now (where $\mathbb{F}_2[\overline{\mathbf{x}}] = \mathbb{F}_2[\mathbf{x}]/\langle x_1^2, \ldots, x_n^2\rangle$):

$$0 \to (\mathbb{F}_2[\overline{\mathbf{x}}]/\langle f_1, \ldots, f_i\rangle)_{d-d_i} \xrightarrow{f_i} (\mathbb{F}_2[\overline{\mathbf{x}}]/\langle f_1, \ldots, f_{i-1}\rangle)_d \to \left(\mathbb{F}_2[\overline{\mathbf{x}}]/\langle x_1^2, \ldots, x_n^2, f_1, \ldots, f_i\rangle\right)_d \to 0$$

then as long as $d < d_{\mathrm{reg}}$ the associated Hilbert functions verify the relation

$$HF_{\langle x_1^2,\ldots,x_n^2,f_1,\ldots,f_i\rangle}(d-d_i) - HF_{\langle x_1^2,\ldots,x_n^2,f_1,\ldots,f_{i-1}\rangle}(d) + HF_{\langle x_1^2,\ldots,x_n^2,f_1,\ldots,f_i\rangle}(d) = 0$$

for all $d < d_{\mathrm{reg}}$. Using the limit conditions, we get the Hilbert series:

$$HS_{\langle x_1^2,\ldots,x_n^2,f_1,\ldots,f_m\rangle}(z) = \sum_{d=0}^{\infty} HF_{\langle x_1^2,\ldots,x_n^2,f_1,\ldots,f_m\rangle}(d) z^d = \left[(1+z)^n \Big/ \prod_{i=1}^{m}(1+z^{d_i})\right].$$

The converse of property (i) is proved exactly as for Proposition 6. Property (ii) is a consequence of the definition of the degree of regularity. For property (iii) see [Bar04]. □

3 Asymptotic Analysis

This section is devoted to the proof of Theorem 1.

We are looking for the first index d for which the d-th coefficient of the series $S_{m,n}$ (resp. $T_{m,n}$) is non-positive. Our method consists in three steps:

- write the d-th coefficient of the series using the Cauchy integral representation, for instance:

$$\mathcal{I}_n(d) = s_{d,m}(n) = \frac{1}{2\imath\pi}\oint S_{m,n}(z)\frac{dz}{z^{d+1}} = \frac{1}{2\imath\pi}\oint e^{nf(z)}dz \tag{6}$$

  where the integration path enclose the origin and no other singularity of $S_{m,n}(z)$

- compute the dominant term in (6) in terms of d and n as $n \to \infty$, d being considered as a parameter,

- determine the asymptotic expansion of d that makes this behavior vanish asymptotically: this gives the first term of the asymptotic expansion of d_{reg}.

By repeatedly doing this process in the neighborhood of the already computed asymptotic expansion of d_{reg}, we get the whole asymptotic expansion of d_{reg}.

For the second step we use the saddle-point and the coalescent saddle points methods, which are standard tools from asymptotic analysis [Hwa97, CFU57, Won89]. The saddle points are the roots of $f'(z)$.

The saddle-point method consists in deforming the integration path to go through the saddle points (see Figure 1) and showing that asymptotically, a small portion of the integration path on both sides of each saddle point contributes most of the integral. A dominant saddle point is a saddle point such that its contribution is exponentially large compared to the contribution of the other saddle points. In our case we will get one dominant saddle point, and we prove that locally, the integrand can be approximated by a Gaussian function, the error term becoming exponentially small as $n \to \infty$.

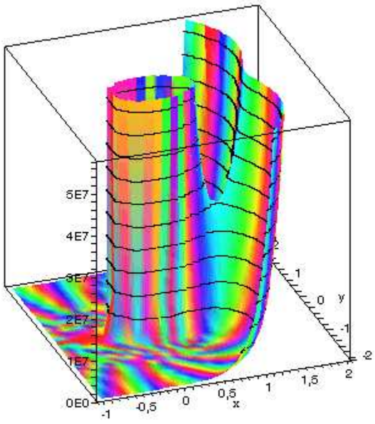

Figure 1: A simple saddle point.

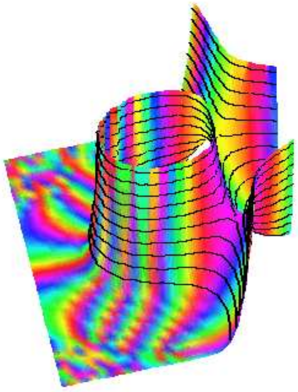

Figure 2: Two coalescing saddle points.

The other case we encounter is the case of two dominant saddle points that coalesce for a particular value d_0 of the parameter d (see Figure 2): we use here a more sophisticated analysis based on the coalescent saddle points method [CFU57]. This method gives an asymptotic expansion of the integral uniformly valid in a neighborhood of d_0, and approximates locally the integrand by a cubic function, thus revealing the connexion with the Airy function.

We write the d-th coefficient of the series using the Cauchy integral representation

$$\mathcal{I}_n(d) = \frac{1}{2\imath\pi} \oint (1-z)^{m-n}(1+z)^m z^{-d-1} dz \tag{7}$$

$$\mathcal{J}_n(d) = \frac{1}{2\imath\pi} \oint (1+z)^n (1+z^2)^{-m} z^{-d-1} dz \tag{8}$$

We distinguish two cases : the case $m = n + k$ for a fixed integer constant $k > 0$, and the case $m = \alpha n$ for a fixed constant $\alpha > 1$.

3.1 Few more equations than unknowns: the case $m = n + k$.

This case is only concerned with the integral $\mathcal{I}_n(d)$. It is convenient to write it as

$$\mathcal{I}_n(d) = \frac{1}{2\imath\pi} \oint \underbrace{(1-z)^{m-n}}_{g(z)} \underbrace{(1+z)^m z^{-d-1}}_{F(z)=e^{nf(z)}} dz = \oint g(z) e^{nf(z)} dz \tag{9}$$

There is only one single saddle point $z_0 = \frac{1}{\frac{m}{d+1}-1}$, root of $f'(z) = \frac{m}{1+z} - \frac{d+1}{z}$.

Proposition 11. *For $m = n + k$, the dominant term in (7) is*

$$\mathcal{I}_n(d) \sim \frac{(1+z_0)^{m+1}(1-z_0)^{m-n}}{\sqrt{2\pi} z_0^{d+1/2}\; m^{1/2}}$$

which vanishes only if $z_0 = 1$, i.e.

$$d_{reg} \sim \frac{m}{2} \tag{10}$$

For $z_0 = 1 - \Delta z$ with $\Delta z \to 0$ as $n \to \infty$ then

$$\mathcal{I}_n(d) \sim \frac{2^{n+\frac{3}{2}k+\frac{1}{2}}}{\sqrt{\pi}\sqrt{m}^{k+1}} H_k\left(\frac{\sqrt{m}}{2^{3/2}}\Delta z(1+o(1))\right)$$

where H_k denotes the Hermite polynomial of order k. This term cancels for $\Delta z = \frac{2^{3/2}}{\sqrt{m}} h_{k,1}$ where $h_{k,1}$ is the largest zero of H_k. Hence the degree of regularity behaves asymptotically like

$$d_{reg} = \frac{m}{2} - h_{k,1}\sqrt{\frac{m}{2}}(1+o(1)) \tag{11}$$

Proof. All computational details of the proof can be found in [Bar04]. A preliminary analysis reveals that the degree of regularity grows roughly linearly with n, that is to say we can restrict our asymptotic analysis of $\mathcal{I}_n(d)$ to the case $1 < \epsilon_1 \leq \frac{n+k}{d+1} \leq \epsilon_2 < \infty$.

The saddle point being real, we choose as integration path:

- A vertical segment L, having for middle z_0. Let us denote by z_1 and z_2 its endpoints, z_1 being of negative imaginary part, and by $2N$ its length, with $N = \frac{\theta_0}{1+z_0}\sqrt{\frac{z_0}{2}}$. θ_0 will be fixed later on.
- An arc of circle C centered at the origin, joining z_1 and z_2 and crossing the negative real axis.

Let us write $\mathcal{I}_n(d) = I_L + I_C$ and $\theta_0 = \frac{1}{n^\alpha}$, simple estimates show that

$$\begin{aligned} \left|\frac{F(z_2)}{F(z_0)}\right| &\leq 2\exp\left(-\frac{(\epsilon_1-1)n^{1-2\alpha}}{2\epsilon_1\epsilon_2}\right) \text{ for } n \text{ large enough and } \frac{1}{4} < \alpha < \frac{1}{2} \\ \text{and } \left|\frac{I_C}{F(z_2)}\right| &\leq \frac{4\pi}{\epsilon_1 - 1}\left(\frac{\epsilon_1+1}{\epsilon_1-1}\right)^k \end{aligned}$$

so that

$$\left|\frac{I_C}{F(z_0)}\right| = O(\frac{1}{n^M}) \text{ for all } M > 0 \text{ as } \theta_0 = \frac{1}{n^\alpha} \text{ and } \frac{1}{4} < \alpha < \frac{1}{2}$$

The dominating part of the integral is concentrated on the segment L around the saddle point. We make the change of variables $u = \frac{i}{(1+z_0)\sqrt{2z_0}}(z - z_0)$ in the integral I_L to get a real integral:

$$\frac{I_L}{F(z_0)} = \frac{(1+z_0)\sqrt{2z_0}}{2\pi} \int_{-N}^{N} g(z(u)) \exp\left[m\left(-u^2 + O(u^3)\right)\right] du$$

the $O(u^3)$ term being uniform in d and n. We apply the Laplace method as in [dB81] and get the dominant term

$$\frac{I_L}{F(z_0)} \sim \frac{(1+z_0)\sqrt{2z_0}}{2\pi} \int_{-\infty}^{\infty} g(z_0) e^{-mu^2} du = \frac{(1+z_0)\sqrt{z_0}}{\sqrt{2\pi}} g(z_0) \frac{1}{\sqrt{m}}$$

In the neighborhood of $z_0 = 1 - \Delta z$, applying again the Laplace method we find the dominant term of the integral to be

$$\begin{aligned} \frac{I_L}{F(z_0)} &\sim \frac{(1+z_0)\sqrt{2z_0}}{2\pi} \int_{-\infty}^{\infty} g(z(u)) e^{-mu^2} du \\ &= \frac{\left((1+z_0)\sqrt{2z_0}\right)^{k+1}}{2\pi\sqrt{m}^{k+1}} \int_{-\infty}^{\infty} (x + \imath u)^k e^{-u^2} du \text{ with } x = \frac{1-z_0}{(1+z_0)\sqrt{2z_0}}\sqrt{m} \\ &\sim \frac{2^{n+\frac{3}{2}k+\frac{1}{2}}}{\sqrt{\pi}\sqrt{m}^{k+1}} H_k\left(\frac{\sqrt{m}}{2^{3/2}}\Delta z(1 + o(1))\right) \end{aligned}$$

with $H_k(x) = \frac{2^k}{\sqrt{\pi}} \int_{-\infty}^{\infty} (x + \imath u)^k e^{-u^2} du$ the k-th Hermite polynomial. □

Indeed, tracing the errors carefully shows that $H_k\left(\frac{\sqrt{m}}{2^{3/2}}\Delta z(1 + o(1))\right)$ can be written as

$$H_k\left(\frac{\sqrt{m}}{2^{3/2}}\Delta z\right) + \frac{k}{\sqrt{8m}} H_{k+1}\left(\frac{\sqrt{m}}{2^{3/2}}\Delta z\right) + O(m^{-1}),$$

Since $H_{k+1}(z)/H_k(z) = z$ for large z, the asymptotics will be valid as long as the second term goes to zero, which works out to $k = o(m^{1/3})$. Since $h_{k,1} = \sqrt{2k+1} + O(k^{-1/6})$, the above is consistent with the uniform asymptotics of the next section (as it should be).

3.2 More equations: the case $m = \alpha n$.

A similar analysis can be done when $m = \alpha n$ ($\alpha > 1$ being fixed) for both generating series. In this case, the factor $(1 - z)^k$ is not a small perturbation any longer, and the integral are written as

$$\mathcal{I}_n(d) = \frac{1}{2\imath\pi}\oint \underbrace{(1-z)^{m-n}(1+z)^m z^{-d-1}}_{F(z)} dz = \oint e^{nf(z)}dz \qquad (12)$$

$$\mathcal{J}_n(d) = \frac{1}{2\imath\pi}\oint \underbrace{(1+z)^n(1+z^2)^{-m} z^{-d-1}}_{F_J(z)} dz \qquad (13)$$

Let us consider first the integral $\mathcal{I}_n(d)$. The behavior of the integrand changes qualitatively and the integral is then dominated by two conjugate saddle points $z_0^\pm = \frac{1\pm\sqrt{\Delta}}{2((2\alpha-1)-\frac{d+1}{n})}$ where $\Delta = 4\left(\frac{d+1}{n}\right)^2 + 4(1-2\alpha)\frac{d+1}{n} + 1$. It vanishes for $\frac{d+1}{n} = \lambda_0^\pm$ with $\lambda_0^\pm = \alpha - \frac{1}{2} \pm \sqrt{\alpha(\alpha-1)} > 0$. As $\frac{d+1}{n} \neq \lambda_0^\pm$, both saddle points are simple and for $\frac{d+1}{n} = \lambda_0^\pm$ there is a double real positive saddle point, denoted by z_0.

As long as $\frac{d+1}{n}$ does not belong to the neighborhood of $\lambda_0^\pm$, the contributions of these saddle points to the integral are conjugate values whose sum does not vanish. This qualitative analysis reveals that a new phenomenon must occur for the integral to vanish: the parameter d must be such that the saddle points coalesce, giving rise to a double saddle point. This happens when both F' and F'' vanish and these equations are sufficient to give the first order behavior of d_{reg}.

A more precise analysis (the coalescent saddle-points method [CFU57]) is achieved by capturing the coalescence of $z_0^\pm$ by means of a cubic change of variables $f(z) = P(u) = \frac{u^3}{3} - \zeta u + \eta$, where $\zeta^{\frac{3}{2}} = \frac{3}{4}(f(z_0^-) - f(z_0^+))$ and $\eta = \frac{1}{2}(f(z_0^-) + f(z_0^+))$ are chosen so that the values of P at its saddle points $-\sqrt{\zeta}$ and $\sqrt{\zeta}$ are the same as that of f at z_0^- and z_0^+. The integral is then renormalized, and leads to a proven [CFU57] full asymptotic expansion:

$$\mathcal{I}_n(d) = e^{n\eta}\left[\frac{\mathrm{Ai}(n^{\frac{2}{3}}\zeta)}{n^{\frac{1}{3}}}\sum_{m\geq 0}\frac{B_m}{n^m} + \frac{\mathrm{Ai}'(n^{\frac{2}{3}}\zeta)}{n^{\frac{2}{3}}}\sum_{m\geq 0}\frac{C_m}{n^m}\right](1+o(1))$$

where Ai is the classical Airy function (the A_m and B_m coefficients can be expressed in terms of f and its derivatives at $z_0^\pm$). By repeatedly canceling the dominant term in the asymptotic expansion of $\mathcal{I}_n(d)$, we get the asymptotic expansion of d_{reg} and the second part of the Theorem.

This asymptotic analysis applies to the integral $\mathcal{J}_n(d)$ to get the third part of the Theorem exactly in the same way: the only changes is that there are three saddle points, two are conjugate and the last one is real and its contribution to the integral is negligible.

4 Conclusion

We provide a definition of semi-regular sequences in the general case and over the finite field $\mathbb{F}_2$, for which we conjecture that almost all sequences are semi-regular: over any field of characteristic 0 it is another form of Fröberg conjecture [Frö85], and over a field of positive characteristic we conjecture that the proportion of semi-regular sequences tends to 1 as the number of variables tends to infinity.

For such systems, we provide sharp asymptotic complexity bounds for the degree of regularity as the number of variables $n \to \infty$, that imply complexity bounds for the Gröbner basis computation. Those asymptotics are very precise compared to the true value of the degree of regularity even for small values of n ($n \geq 3$).

From a cryptographical point of view, for $m = \alpha n$ equations, the global complexity of solving "random" systems is simply exponential in n, even for quadratic equations: "random" systems remain exponential, therefore out of reach as soon as $n \geq 80$ for instance, and are a good source of difficult problems for the design of cryptosystems.

References

[AFI$^+$04] G. Ars, J.-C. Faugère, H. Imai, M. Kawazoe, and M. Sugita. Comparison Between XL and Gröbner Basis Algorithms. In Pil Joong Lee, editor, *Advances in Cryptology - ASIACRYPT 2004, Jeju Island, Korea*, number 3329 in LNCS, p. 338 – 353. Springer Heidelberg, December 5-9 2004.

[Bar04] M. Bardet. *Étude des systèmes algébriques surdéterminés. Applications aux codes correcteurs et à la cryptographie.* PhD thesis, Université Paris VI, Décembre 2004.

[BFS03] Magali Bardet, Jean-Charles Faugère, and Bruno Salvy. Complexity of Gröbner basis computation for semi-regular overdetermined sequences over GF(2) with solutions in GF(2). Research Report RR-5049, INRIA, Décembre 2003. 19 p. .

[BFS04] M. Bardet, J.-C. Faugère, and B. Salvy. On the complexity of gröbner basis computation of semi-regular overdetermined algebraic equations. In *Proc. ICPSS International Conference on Polynomial System Solving Paris, November 24-25-26 2004 in honor of Daniel Lazard*, 2004.

[Buc65] B. Buchberger *Ein Algorithmus zum Auffinden der Basiselemente des Restklassenringes nach einem nulldimensionalen Polynomideal.* PhD thesis, Innsbruck, 1965.

[CFU57] C. Chester, B. Friedman, and F. Ursell. An extension of the method of steepest descents. *Proc. Camb. Philos. Soc.*, 53:599–611, 1957.

[CLO98] D. Cox, J. Little, and D. O'Shea. *Using Algebraic Geometry.* Springer Verlag, New York, 1998.

[dB81] N. G. de Bruijn. *Asymptotic methods in analysis.* Dover Publications Inc., New York, third edition, 1981.

[Die04] C. Diem. The XL-Algorithm and a Conjecture from Commutative Algebra. In Pil Joong Lee, editor, *Advances in Cryptology - ASIACRYPT 2004, Jeju Island, Korea*, number 3329 in LNCS, p. 323–337. Springer Heidelberg, December 5-9 2004.

[Fau99] Jean-Charles Faugère. A new efficient algorithm for computing Gröbner bases (F_4). *J. Pure Appl. Algebra*, 139(1-3):61–88, 1999. Effective methods in algebraic geometry (Saint-Malo, 1998).

[Fau02] J.-C. Faugère. A new efficient algorithm for computing Gröbner bases without reduction to zero F5. In T. Mora, editor, *Proceedings of ISSAC*, p. 75–83. ACM Press, July 2002.

[FJ03] J.-C. Faugère and A. Joux. Algebraic cryptanalysis of Hidden Field Equation (HFE) cryptosystems using Gröbner bases. In Dan Boneh, editor, *Advances in Cryptology - CRYPTO 2003*, volume 2729 of *LNCS*, p. 44–60. Springer, 2003.

[Frö85] Ralf Fröberg. An inequality for Hilbert series of graded algebras. *Math. Scand.*, 56(2):117–144, 1985.

[Frö97] R. Fröberg. *An introduction to Gröbner bases.* Pure and Applied Mathematics. John Wiley & Sons Ltd., Chichester, 1997.

[Giu84] M. Giusti. Some effectivity problems in polynomial ideal theory. In *Proc. Int. Symp. on Symbolic and Algebraic Computation EUROSAM 84, Cambridge (England)*, volume 174 of *LNCS*, p. 159–171. Springer, 1984.

[Hwa97] H.-K. Hwang. Asymptotic estimates of elementary probability distributions. *Stud. Appl. Math.*, 99(4):393–417, 1997.

[Lan02] S. Lang. *Algebra*, volume 211 of *Graduate Texts in Mathematics.* Springer, New York, third edition, 2002.

[Laz83] D. Lazard. Gaussian Elimination and Resolution of Systems of Algebraic Equations. In *Proc. EUROCAL 83*, volume 162 of *LNCS*, p. 146–157, 1983.

[Laz01] D. Lazard. Solving systems of algebraic equations. *ACM SIGSAM Bulletin*, 35(3):11–37, Septembre 2001.

[Mac16] F.S. Macaulay. *The algebraic theory of modular systems.*, volume xxxi of *Cambridge Mathematical Library.* Cambridge University Press, 1916.

[PR03] K. Pardue and B. Richert. Syzygies of semi-regular sequences. Preprint available at: `http://www.math.lsa.umich.edu/~brichert/publications/`, 2003.

[SPCK00] A. Shamir, J. Patarin, N. Courtois, and A. Klimov. Efficient algorithms for solving overdefined systems of multivariate polynomial equations. In *Advances in cryptology – EUROCRYPT '00*, volume 1807 of *LNCS*, p. 392–407, Heidelberg, 2000. Springer.

[Sza04] A. Szanto. Multivariate subresultants using Jouanoulou's resultant matrices. Accepted to Journal of Pure and Applied Algebra. Preprint Available at: `http://www.mathpreprints.com/math/Preprint/aszanto/20011204/2`, 2004.

[Won89] R. Wong. *Asymptotic approximations of integrals.* Computer Science and Scientific Computing. Academic Press Inc., Boston, MA, 1989.

[YC04] B.-Y. Yang and J.-M. Chen. All in the XL Family: Theory and Practice. In *Proc. 7th ICISC '04 (Dec. 2-3, 2004, Seoul, Korea)*, a revised version to appear in a volume of LNCS, Dec. 2-3 2004.