Probabilistic analysis on Macaulay matrices over finite fields and complexity of constructing Gröbner bases. Extended absract.

Igor Semaev and Andrea Tenti

Department of Informatics, University of Bergen, Norway
`igor.semaev@uib.no, andrea.tenti@uib.no`

Abstract. Gröbner basis methods are used to solve systems of polynomial equations over finite fields, but their complexity is poorly understood. In this work an upper bound on the time complexity of constructing a Gröbner basis is proved. A key parameter in this estimate is the degree of regularity of the leading forms of the polynomials. Therefore we provide an upper bound on the degree of regularity for a sufficiently overdetermined system of forms over any finite field. The bound holds with probability tending to 1 and depends only on the number of variables, the number of polynomials, and their degrees. Our results imply that sufficiently overdetermined systems of polynomial equations are solved in polynomial time with high probability.

Keywords: Algebraic cryptanalysis · Overdetermined systems of polynomial equations · Gröbner Basis · Macaulay matrices · Multisets

1 Introduction

Problems in cryptanalysis may commonly be reduced to solving a system of multivariate polynomial equations over a finite field $\mathbb{F}_q$:

$$P_1(x_1, \dots, x_n) = 0, \dots, P_m(x_1, \dots, x_n) = 0. \tag{1}$$

Finding solutions of the system in $\mathbb{F}_q$ is equivalent to breaking a cryptosystem. A particularly successful example is due to Faugère and Joux that broke HFE (Hidden Field Equations) with a Gröbner basis algorithm [FJ03]. The worst case time-complexity of Gröbner basis methods is known to be bounded by a double exponential function in the number of variables already for quadratic systems [MM82]. In some cryptographic applications the problem is reduced to overdetermined polynomial systems, where the number of equations m is larger than the number of variables n. For instance, AES (Advanced Encryption Standard) S-box may be represented by an overdetermined system of quadratic equations. So the whole cipher is reduced to solving an overdetermined quadratic equation system [CP02]. Such systems may generally be solved faster when using algorithms from Gröbner basis or XL families [BFS03; Cou+00]. Hence, time-complexity of those algorithms for overdetermined polynomial equation systems is interesting to study.

Let I be an ideal in $R = \mathbb{F}_q[x_1, \ldots, x_n]/(x_1^q, \ldots, x_n^q)$ generated by the leading forms $f_1, \ldots, f_m$ of the polynomials $P_1, \ldots, P_m$. By I_d we denote a vector space over $\mathbb{F}_q$ containing all forms in I of degree d. The degree of regularity of I is the smallest integer d for which $\dim_{\mathbb{F}_q} I_d = l_q(n, d)$, the number of monomials of total degree d.

In Theorem 2 we show that time-complexity of constructing a Gröbner basis for $P_1, \ldots, P_m$ (we need to add $x_i^q - x_i, i = 1, \ldots, n$ to avoid solutions in the extensions of the ground field) is polynomial in $L_q(n, d_{\text{reg}})$, where $L_q(n, d)$ is the number of monomials of total degree $\leq d$. One finds a solution to (1) with the same complexity.

The notion of a semiregular system of polynomials (forms) was introduced by Bardet, Faugère, and Salvy in [BFS03]. For semiregular polynomials over $\mathbb{F}_2$ it was there proved that the degree of regularity only depends on the number of variables n, the number of equations m, and their degrees. So the degree of regularity for a particular semiregular polynomial system may be computed by expanding a Hilbert series defined by those parameters. It was also conjectured that a random system of polynomials over $\mathbb{F}_2$ is semiregular with probability tending to 1 as n increases. The conjecture, in the way it was presented, was disproved in [HMS17]. Still it is believed that most systems behave like semiregular ones.

The present work gives an upper bound on the degree of regularity for an overdetermined system of forms $f_1, \ldots, f_m$ with coefficients in $\mathbb{F}_q$ taken uniformly at random and of the same degree D. The bound holds with probability tending to 1. We do not impose any other restrictions on the polynomials as semiregularity, etc. The following statement is proved.

Theorem 1. *Let $m \geq l_q(n, D + d)/l_q(n, d)$, where $D > d$. Then*

$$\mathbb{P}(d_{reg} \leq D + d) \geq 1 - q^{l_q(n,D+d) - m l_q(n,d)} + O(n^d q^{-n^D})$$

as $n \to \infty$.

Theorem 1 implies that for sufficiently large m almost all polynomial equation systems (1) are solved in polynomial time. For instance, let $q = 2$. Then for $m \geq \frac{(n-1)(n-2)}{6} + 1$ quadratic polynomials ($D = 2, d = 1$) a Gröbner basis may be computed at $d_{\text{reg}} \leq 3$ with probability tending to 1. Similarly, for $m \geq \frac{(n-2)(n-3)(n-4)}{60} + 1$ cubic polynomials ($D = 3, d = 2$) a Gröbner basis may be computed at $d_{\text{reg}} \leq 5$ with probability tending to 1.

Over $\mathbb{F}_2$ the bound on d_{reg} is as predicted in [BFS03] for a semiregular system with the same parameters (number of variables n, number of equations, and of degree D). Under a conjecture from commutative algebra a lower bound on the degree of regularity for homogeneous polynomial systems in $\mathbb{F}_q[x_1, \ldots, x_n]$ is proved in [Die04]. Our result complies with this bound as well.

The sketch of the proof of Theorem 1 is in Section 3, where we show that a Macaulay matrix of size $m\, l_q(n, d) \times l_q(n, D + d)$ constructed for the forms $f_1, \ldots, f_m$ has linearly independent columns with probability tending to 1.

Section 4 contains a combinatorial Theorem 3 used in the proof of the main Theorem 1. Each monomial $x_1^{a_1}\dots x_n^{a_n}$ of total degree d defines a d-multiset $(a_1,\dots,a_n)$, where $0 \le a_i \le q-1$ and $\sum_{i=1}^n a_i = d$. Theorem 3 implies that the minimum number of monomials of total degree D divisible by monomials of total degree d from a family of size v is achieved for a family of the first (largest) v monomials of total degree d taken in a lexicographic order.

Theorem 2 was proved by Semaev. The main idea of the proof of Theorem 1 belongs to Semaev too, who first proved the it for $\mathbb{F}_2$ and $D=2, d=1$. The generalisation to any $\mathbb{F}_q$ and $D > d$ is due to Tenti. Also Tenti conjectured the statement of Theorem 3 for $k = k_1 = \dots = k_n$ and proved it for $k=1, d=2$. With a different method presented in Section 4 the theorem in its generality was proved by Semaev.

2 Complexity of constructing Gröbner bases

We can assume that the polynomials $P_1,\dots,P_m$ from (1) are in

$$\mathbb{F}_q[x_1,\dots,x_n]/(x_1^q - x_1,\,\dots,x_n^q - x_n)$$

and all computations are performed in this ring. Time-complexity of constructing a Gröbner basis for $P_1,\dots,P_m$ is here estimated. Let $N = L_q(n, d_{\text{reg}})$. We have $d_{\text{reg}} \le (q-1)n$. To compute d_{reg} one gradually triangulates with elimination Macaulay matrices for $d \le d_{\text{reg}}$ with at most

$$\sum_{i=1}^{d} l_q(n,i) l_q(n,d-i) \le dN^2$$

rows and $L_q(n,d) \le N$ columns. The overall cost is $O(d_{\text{reg}}^2 N^4)$ operations in $\mathbb{F}_q$. The result is a system of linearly independent polynomials $B = \{Q_1,\dots,Q_r\}$ of degree $\le d_{\text{reg}}$. Exactly $l_q(n,d_{\text{reg}})$ polynomials are of degree d_{reg} and their leading forms are all possible monomials of degree.

Generally, that is not enough. For instance, the polynomial system $P_1 = x_1x_2 + 1, P_2 = x_1x_3, P_3 = x_2x_3 \in \mathbb{F}_2[x_1,x_2,x_3]$ has $d_{\text{reg}} = 2$ and that is not a Gröbner basis, as the ideal $< P_1, P_2, P_3 >$ contains the polynomial $x_3 = x_3P_1 + x_2P_2$ and its leading term is not divisible by the leading terms of P_1, P_2, P_3. In order to compute a Gröbner basis one has to work with polynomials of degree $> d_{\text{reg}}$ as well. So the argument in Section 2.2 of [BFS03], on the complexity of constructing a Gröbner basis is not valid.

The following theorem proves that one can construct a Gröbner basis at maximum degree $\le 2d_{\text{reg}}$. Assume a total degree monomial ordering.

Theorem 2. *Time-complexity of constructing a Gröbner basis for $P_1,\dots,P_m$ is polynomial in N.*

Proof. We will prove that the construction takes $O(N^6)$ operations in $\mathbb{F}_q$. Let's consider an application of the Buchberger algorithm, see [CLO13], to the polynomials B. For each $Q_1, Q_2 \in B$ the algorithm computes a residue R of the

S-polynomial $S(Q_1, Q_2)$ after division by the polynomials B. Each monomial of degree d_{reg} occurs as a leading monomial of some polynomial in B. Therefore $\deg R < d_{\mathrm{reg}}$. If $R \neq 0$, then B is augmented with R and the step repeats. If the residue is 0 for each pair, then B is a Gröbner basis. At each step of the algorithm the polynomials in B are linearly independent.

One has to examine $< N^2$ pairs before finding a non-zero residue or terminating. The number of possible linearly independent residues is $< N$, so the number of divisions is at most N^3. Each S-polynomial has degree $\leq 2d_{\mathrm{reg}}$, so it incorporates at most N^2 monomials. Computing its residue takes at most $O(N^3)$ field operations. Overall complexity is the one stated.

A more careful analysis shows that one can work with polynomials of degree $\leq 2d_{\mathrm{reg}} - 2$ and the time-complexity is

$$O(N^2\, L_q^2(n, d_{\mathrm{reg}} - 1)\, L_q(n, 2d_{\mathrm{reg}} - 2))$$

operations in $\mathbb{F}_q$.

3 Analysis of the probability

The sketch of Theorem 1 proof is here presented. We consider a system of m homogeneous polynomials $f_1, \ldots, f_m$ of degree D. Let d be a natural number. The degree d Macaulay matrix of the system is the matrix $M = M(n, m, d)$, whose rows are labelled by pairs (r, f_i), where r are all monomials of degree d, and columns are labelled by the monomials t of degree $D + d$. The entry of the matrix M in the row (r, f_i) and the column t is equal to the coefficient at the monomial t in rf_i computed in R. The size of the matrix M is $m\, l_q(n, d) \times l_q(n, D + d)$. If the columns of M are linearly independent, then $d_{reg} \leq D + d$.

Let the homogeneous polynomials $f_1, \ldots, f_m$ of degree D be taken uniformly at random. By p we denote the probability that the columns of M are linearly dependent. We prove that if $d < D$ and $m \geq l_q(n, D + d)/l_q(n, d)$, then

$$p \leq q^{l_q(n,D+d) - m l_q(n,d)} + O(n^d q^{-n^D})$$

as n tends to infinity. That will prove Theorem 1.

The matrix M can be divided into m blocks $M_1, \ldots, M_m$, each with $l_q(n, d)$ rows. The matrix M_j is the Macaulay matrix for the single polynomial f_j. Let u be a vector of length $l_q(n, D + d)$. Its entries can be indexed by the multisets $\mathcal{X}^{D+d}$, where $\mathcal{X}$ is defined in Section 4 with $k_1 = \ldots = k_n = q - 1$.

Let u be a column vector of length $l_q(n, D+d)$ over $\mathbb{F}_q$. Since the polynomials f_j are chosen independently, the probability $p_u = \mathbb{P}(Mu = 0)$ is equal to p_{1u}^m, where $p_{ju} = \mathbb{P}(M_j u = 0)$. We deduce $p \leq \sum_{u \neq 0} p_u = \sum_{u \neq 0} p_{1u}^m$, as the presence of linear dependencies of the columns in M is equivalent to a nontrivial kernel.

Let c denote a vector of length $l_q(n, D)$, whose entries c_L are indexed by the multisets in $\mathcal{X}^D$. They are the coefficients at the monomials of f_1. Let m_{JI} denote the entry of M_1 in the row $J \in \mathcal{X}^d$ and the column $I \in \mathcal{X}^{D+d}$. Then

$m_{JI} = c_{I\setminus J}$ if $J \subseteq I$ and $m_{JI} = 0$ otherwise. So $M_1 u = 0$ is equivalent to the following equalities which hold for every $J \in \mathcal{X}^d$:

$$\sum_{I\in\mathcal{X}^{D+d}} m_{JI}\, u_I = \sum_{J\subseteq I} c_{I\setminus J}\, u_I = \sum_{L+J\in\mathcal{X}^{D+d}} c_L\, u_{L+J} = 0,$$

where the second sum is over $I \in \mathcal{X}^{D+d}$ such that $J \subseteq I$, and the third sum is over $L \in \mathcal{X}^D$ such that $L + J \in \mathcal{X}^{D+d}$.

Let $Y^{(u)}$ be a matrix of size $l_q(n,d) \times l_q(n,D)$, whose rows and columns are labelled by the elements of $\mathcal{X}^d$ and $\mathcal{X}^D$ respectively. The entries of $Y^{(u)}$ are defined by

$$Y^{(u)}_{J,L} = \begin{cases} u_{J+L} & \text{if } J+L \in \mathcal{X}^{D+d}, \\ 0 & \text{otherwise.} \end{cases}$$

So $M_1 u = 0$ is equivalent to $Y^{(u)} c = 0$, so $p_{u1} = q^{-\operatorname{rk}(Y^{(u)})}$. Therefore,

$$p \le \sum_{u\neq 0} q^{-m\operatorname{rk}(Y^{(u)})} = \sum_{v=0}^{l_q(n,d)-1} N_v q^{-m(l_q(n,d)-v)}, \tag{2}$$

where N_v denotes the number of vectors u such that $\operatorname{rk}(Y^{(u)}) = l_q(n,d) - v$. The value N_v is upper bounded by the size of the set

$$S_v = \left\{ u \middle| \operatorname{rk}(Y^{(u)}) \le l_q(n,d) - v \right\}.$$

In particular, $u \in S_v$ if and only if there exists a subspace $V \subseteq \mathbb{F}_q^{l_q(n,d)}$ of dimension v in the kernel of $Y^{(u)}$. Let $B = (b_1, \ldots, b_v)$ be a basis for this space. We label the coordinates of b_i with elements J of $\mathcal{X}^d$ according to the lexicographic order from left to right. Then $b_i Y^{(u)} = 0$ is equivalent to the following equality which holds for every $L \in \mathcal{X}^D$:

$$\sum_{J+L\in\mathcal{X}^{D+d}} b_{i,J} u_{J+L} = 0, \tag{3}$$

where the sum is over $J \in \mathcal{X}^d$ such that $J + L \in \mathcal{X}^{D+d}$. The basis B may be represented as a matrix of size $v \times l_q(n,d)$ in a row echelon form, where every leading coefficient is 1.

$$B = \begin{pmatrix} 0 \ldots\ 0\ 1 * \ldots\ * 0 * \ldots \\ 0 \ldots\ 0\ 0\ 0 \ldots\ 0\ 1 * \ldots \\ \ldots \end{pmatrix}.$$

For every $0 < i \le v$ we define the matrix A_i by the following.

- A_i has $l_q(n, D+d)$ rows and $l_q(n,D)$ columns, labelled by $I \in \mathcal{X}^{D+d}$ and by $L \in \mathcal{X}^D$ respectively. The labels are ordered according to the lexicographic order from left to right and from top to bottom.

– The entry I, L of A_i is

$$A_{i,I,L} = \begin{cases} b_{i,I\setminus L} & \text{if } L \subseteq I, \\ 0 & \text{otherwise.} \end{cases}$$

Let A_V denote the horizontal concatenation of the matrices $A_1, \dots, A_v$, that is $A_V = A_1|A_2|\dots|A_v$. The equalities (3) are equivalent to $uA_V = 0$ and therefore

$$|S_v| \leq \sum_{\dim(V)=v} q^{l_q(n,D+d)-\mathrm{rk}(A_V)},$$

where the sum is over subspaces V of dimension v in $\mathbb{F}_q^{l_q(n,d)}$. Let the multiset $J_i \in \mathcal{X}^d$ label the first nonzero entry of the vector b_i.

Lemma 1. *For the subspace V with the basis B it holds that*

$$\mathrm{rk}(A_V) \geq \left| \bigcup_{i=1}^{v} \{ I \in \mathcal{X}^{D+d} | I \supseteq J_i \} \right|.$$

We omit the proof of the lemma in this abstract. By combining Lemma 1 and Theorem 3, one concludes that for every V of dimension v, $\mathrm{rk}(A_V) \geq \chi_v^{D+d}$ and so

$$N_v \leq \sum_{\dim(V)=v} q^{l_q(n,d+D)-\mathrm{rk}(A_V)} \leq s_v q^{l_q(n,d+D)-\chi_v^{D+d}}.$$

where s_v is the number of subspaces of dimension v in $\mathbb{F}_q^{l_q(n,d)}$. It is easy to see that $s_v \leq q^{(l_q(n,d)-v+1)v}$. By applying (2), one gets:

$$p \leq \sum_{v=0}^{l_q(n,d)-1} q^{(l_q(n,d)-v+1)v+l_q(n,D+d)-\chi_v^{D+d}-(l_q(n,d)-v)m} =$$

$$= q^{l_q(n,D+d)-ml_q(n,d)} + \sum_{v=1}^{l_q(n,d)-1} q^{(l_q(n,d)-v+1)v+l_q(n,D+d)-\chi_v^{D+d}-(l_q(n,d)-v)m}.$$

An analysis of the second term reveals that for n large enough it is $O(n^d q^{-n^D})$. That finishes the proof. □

Remark 1. We notice that if $m < l_q(n, D+d)/l_q(n,d)$, then the regularity degree for m homogeneous polynomials of degree D cannot be smaller than or equal to $D+d$, for the Macaulay matrix of degree d cannot have linearly independent columns.

4 Minimal covering family of multisets

A multiset over $\{1, \dots, n\}$ is a sequence $A = (a_1, \dots, a_n)$ with integer $a_i \geq 0$. Let $B = (b_1, \dots, b_n)$ be another multiset. We say that $A \subseteq B$ if $a_i \leq b_i$ for every i. One defines $A+B=(a_1+b_1,\dots,a_n+b_n)$ and if $A\subseteq B$, then $B\setminus A=(b_1-a_1,\dots,b_n-a_n)$. We say that $|A|=d$ if $\sum_{i=1}^n a_i=d$ and call A a d-multiset.

For integer $k_1,\dots,k_n\geq 0$ we define

$$\mathcal{X}=\{(a_1,\dots,a_n)|0\leq\ a_i\leq k_i, i=1,\dots,n\},$$

and $\mathcal{X}^d=\{A\in\mathcal{X}$ such that $|A|=d\}$.

Let $\mathcal{A}=\{A_1,\dots,A_v\}$ be a family of d-multisets and $D\geq d$. By $||\mathcal{A}||$ we denote the number of multisets from $\mathcal{X}^D$ which contain at least one from $\mathcal{A}$ (we say covered by $\mathcal{A}$). The ordering on $\{1,2,\dots,n\}$ induces a lexicographic order on the family $\mathcal{X}^d$. Let $\mathcal{X}_v=\{X_1,\dots,X_v\}$ denote the first(largest) v multisets according to that ordering and $\chi_v^D=||\mathcal{X}_v||$.

Theorem 3. *Let* $k_1\leq k_2\leq\ldots\leq k_n$*, then* $||\mathcal{A}||\geq\chi_v^D$.

Let $\mathcal{Y}_u$ be the family of the first(largest) u elements in $\mathcal{X}^D$ according to the lexicographic order on $\mathcal{X}^D$. Let X_v be a d-multiset at place v in the ordered family $\mathcal{X}$ and $Y_{\ell(v)}$ denote the smallest D-multiset such that $Y_{\ell(v)}\supseteq X_v$ (covered by X_v). So $\mathcal{Y}_{\ell(v)}=\{Y_1,\dots,Y_{\ell(v)}\}$ the ordered family of $Y\geq Y_{\ell(v)}$ in $\mathcal{X}^D$. We give a sketch of the theorem proof based on several lemmas.

Lemma 2. *The family of D-multisets covered by* $\mathcal{X}_v$ *is* $\mathcal{Y}_{\ell(v)}$. *In particular,* $\chi_v^D=\ell(v)$.

Proof. To simplify notation we write $Y=Y_{\ell(v)}$ and $X=X_v$. Let $X'\geq X$, we will prove that for any D-multiset Y' such that $Y'\supseteq X'$ we have $Y'\geq Y$. We denote

$$X=(x_1,\dots,x_{i-1},x_i,\dots,x_n),\quad X'=(x_1,\dots,x_{i-1},x'_i,\dots,x'_n),$$

where $x'_i>x_i$, and

$$Y=(y_1,\dots,y_{i-1},y_i,\dots,y_n),\quad Y'=(y'_1,\dots,y'_{i-1},y'_i,\dots,y'_n).$$

Then $y'_1\leq y_1$, otherwise $Y'>Y$ and it is nothing to prove. If $y'_1<y_1$, then we get a contradiction with the minimality of Y by constructing $Y''<Y$ and $X\subseteq Y''$. So $y'_1=y_1$. By the same argument we prove that $y'_j=y_j$, $1\leq j\leq i-1$. So $x_i<x'_i\leq y'_i\leq y_i$. If $i=n$, then $Y'=Y$ and nothing is to prove. If $i<n$, then one gets a similar contradiction. So $y'_i>y_i$ and $Y'>Y$. It is easy to see that for any D-multiset $Y'\geq Y$ there exists d-multiset $X'\geq X$ such that $X'\subseteq Y'$. That proves the statement.

Let s be a natural number and $f(v)=|\mathcal{Y}_{\ell(v+s)}\setminus\mathcal{Y}_{\ell(v)}|$ for $0\leq v\leq|\mathcal{X}^d|-s$.

Lemma 3. $f(|\mathcal{X}^d|-s)\leq f(v)\leq f(0)$.

We omit the proof of the lemma in this abstract.

Lemma 4. *It is enough to prove Theorem 3 for* $D=d+1$.

Proof. Let the theorem be true for $D = d + 1$ and any d. We prove it is true for $D = d + 2$. Let $\ell_{01}(s), \ell_{12}(s), \ell_{02}(s)$ be the above function for $d, d + 1$, and $d+1, d+2$, and $d, d+2$ respectively. The family $\mathcal{A}$ of d-multisets covers a family $\mathcal{B}$ of $(d + 1)$-multisets, and $\mathcal{B}$ covers a family $\mathcal{C}$ of $(d + 2)$-multisets. Then $\mathcal{C}$ contains all $(d + 2)$-multisets covered by $\mathcal{A}$. So, in particular, it is easy to see that $\ell_{12}(\ell_{01}(s)) = \ell_{02}(s)$. Let $|\mathcal{A}| = s, |\mathcal{B}| = r, |\mathcal{C}| = t$. Then

$$t \geq \ell_{12}(r), \quad r \geq \ell_{01}(s).$$

Therefore $t \geq \ell_{12}(r) \geq \ell_{12}(\ell_{01}(s)) = \ell_{02}(s)$ and the statement is true for $D = d + 2$. One uses the same argument to prove the lemma for $D > d + 2$.

Proof (sketch) of the Theorem. Let $\{1, \ldots, n\} = \{i_1, \ldots, i_r\} \cup \{j_1, \ldots, j_{n-r}\}$, where $1 \leq r < n$. One splits $\mathcal{A}$ into subfamilies $\mathcal{A}_Z$, where Z runs over t-multisets $(z_{i_1}, \ldots, z_{i_r})$, $0 \leq z_{i_l} \leq k_{i_l}$ and $0 \leq t \leq d$. Each d-multiset $(a_1, a_2, \ldots, a_n) \in \mathcal{A}_Z$ satisfies $(a_{i_1}, \ldots, a_{i_r}) = Z$ and $(a_{j_1}, \ldots, a_{j_{n-r}})$ is a $(d-t)$-multiset. Let $|\mathcal{A}_Z| = s_Z$ and $\mathcal{C}_Z$ be a family of d-multiset $(a_1, a_2, \ldots, a_n)$, where $(a_{i_1}, \ldots, a_{i_r}) = Z$ and $(a_{j_1}, \ldots, a_{j_{n-r}})$ are the first(largest) s_Z of $(d - t)$-multisets in the lexicographic ordering. We put $\mathcal{C} = \bigcup_Z \mathcal{C}_Z$ and say $\mathcal{C}$ satisfies the condition $(i_1, \ldots, i_r)$. Obviously, $|\mathcal{C}| = |\mathcal{A}|$. By using induction we will now prove that $||\mathcal{C}|| \leq ||\mathcal{A}||$.

Let $\mathcal{B}_Z$ be a family of D-multisets $(b_1, b_2, \ldots, b_n)$ covered by $\mathcal{A}_Z$. One splits $\mathcal{B}_Z$ into subfamilies $\mathcal{B}_{Z,U}$, where U runs over T-multisets $(u_{i_1}, \ldots, u_{i_r})$ for $Z \subseteq U$ and $t \leq T \leq D$. Each D-multiset $(b_1, b_2, \ldots, b_n) \in \mathcal{B}_{Z,U}$ satisfies $(b_{i_1}, \ldots, b_{i_r}) = U$, where $(b_{j_1}, \ldots, b_{j_{n-r}})$ is a $(D - T)$-multiset.

We now consider multisets $(a_{j_1}, \ldots, a_{j_{n-r}})$. Let $\ell_{Z,U}(s)$ be the number of such $(D-T)$-multisets covered by the first s such $(d-t)$-multisets in the lexicographic order. As $n - r < n$, by induction, $|\mathcal{B}_{Z,U}| \geq \ell_{Z,U}(s_Z)$ and therefore $|\bigcup_Z \mathcal{B}_{Z,U}| \geq \max_Z \ell_{Z,U}(s_Z)$. Then

$$||\mathcal{A}|| = |\bigcup_{Z,U} \mathcal{B}_{Z,U}| = \sum_U |\bigcup_Z \mathcal{B}_{Z,U}| \geq \sum_U \max_Z \ell_{Z,U}(s_Z) = ||\mathcal{C}||.$$

If $\mathcal{A}$ does not satisfy the condition $(i_1, \ldots, i_r)$, one transforms $\mathcal{A}$ into a family for which this condition is satisfied. As the members of $\mathcal{A}$ are becoming larger in the lexicographic order, this process stops at some point. So we assume $\mathcal{A}$ satisfies all the conditions $(i_1, \ldots, i_r)$ for $1 \leq r < n$.

One now combines the conditions (1) and $(3, \ldots, n)$, and the inequalities in Lemma 3 to finish the proof. □

Acknowledgements

We are grateful to two anonymous reviewers for providing insightful comments.

References

[BFS03] Magali Bardet, Jean-Charles Faugère, and Bruno Salvy, "Complexity of Gröbner basis computation of Semi-Regular Overdetermined sequences over F_2 with solutions in F_2", *INRIA Research Report 5049*, 2003.

[CLO13] David Cox, John Little, and Donal O'Shea, *Ideals, Varieties, and Algorithms: An introduction to Computational Algebraic Geometry and Commutative Algebra*, 4th ed., Springer, Heidelberg, 2013.

[Cou+00] Nicolas Courtois et al., "Efficient algorithms for solving overdefined systems of multivariate polynomial equations", *EUROCRYPT 2000. LNCS 1807*, Springer, Heidelberg, 2000, pp. 392–407.

[CP02] Nicolas Courtois and Josef Pieprzyk, "Cryptanalysis of block ciphers with overdefined systems of equations", *ASIACRYPT 2002. LNCS 2501*, Springer, Heidelberg, 2002, pp. 267–287.

[Die04] Claus Diem, "The XL-algorithm and a conjecture from commutative algebra", *ASIACRYPT 2004. LNCS 3329*, Springer, Heidelberg, 2004, pp. 323–337.

[FJ03] Jean-Charles Faugère and Antoine Joux, "Algebraic cryptanalysis of hidden field equation (HFE) cryptosystems using Gröbner bases", *CRYPTO 2003. LCNS 2729*, Springer, Heidelberg, 2003, pp. 44–60.

[HMS17] Timothy J. Hodges, Sergio D. Molina, and Jacob Schlather, "On the existence of homogeneous semi-regular sequences in F2[X1,...,Xn]/-(X12,...,Xn2)", *Journal of Algebra* 476 (2017), pp. 519–547.

[MM82] Ernst W. Mayr and Albert R. Meyer, "The complexity of the word problems for commutative semigroups and polynomial ideals", *Advances in Mathematics* 46.3 (1982), pp. 305–329.