# Admissible Parameter Sets and Complexity Estimation of Crossbred Algorithm

Shuhei Nakamura

Department of Computer and Information Sciences, Ibaraki University
shuhei.nakamura.fs71@vc.ibaraki.ac.jp

**Abstract.** The Crossbred algorithm is one of the algorithms for solving a system of polynomial equations, proposed by Joux and Vitse in 2017. It has been implemented in Fukuoka MQ challenge, which is related to the security of multivariate crytography, and holds several records. A framework for estimating the complexity has already been provided by Chen et al. in 2017. However, it is generally unknown which parameters are actually available. This paper investigates how to select available parameters for the Crossbred algorithm. As a result, we provide formulae that give an available parameter set and estimate the complexity of the Crossbred algorithm.

## 1 Introduction

The problem of solving systems of polynomial equations has been applied in various fields such as cryptography, coding theory, statistics, and robotics. For instance, in cryptography, the security of current public key cryptography relies on integer factorization and discrete logarithm problems, which can be solved using a large-scale quantum computer. As a result, multivariate polynomial cryptography, whose security is based on the hardness of solving the system of polynomial equations, is expected to be resistant to quantum computers.

Several methods for solving a system of polynomial equations have been proposed, such as utilizing Gröbner bases or employing the kernel search of Macaulay matrices. Some of these methods have been widely implemented and utilized in cryptography. Moreover, in cryptography, it is not always necessary to find all solutions to the polynomial system. Instead, a combination of exhaustive search for some variables and solving algorithms is often employed. BooleanSolve [1], FXL [6], and these variants are examples of such approaches.

Crossbred algorithm is an algorithm proposed by Joux and Vitse in 2017 [13]. While BooleanSolve and FXL perform linear algebra using a Macaulay matrix after fixing some values during exhaustive search, Crossbred algorithm performs the exhaustive search after linear algebra on the Macaulay matrix. This algorithm achieved several records at Fukuoka MQ Challenge [20], which aimed to solve polynomial equations appearing in cryptography. The Crossbred algorithm has certain parameters that control its operation, and Chen et al. [4] already provide a framework for estimating the complexity with a parameter set. It is expected that Crossbred has a parameter set which the complexity is lower than that of BooleanSolve/FXL for some input polynomial systems. However, it is unknown which parameter sets actually operate the algorithm.

### 1.1 Our contribution

For quadratic polynomials $f_1, \dots, f_m$ in $n$ variables and positive integers $D, d$, and $k$, the Crossbred algorithm generates new polynomials $p_1, \dots, p_r$ from $Mac_D(\{f_1, \dots, f_m\})$ whose degree in first $k$ variables is less than or equal to $d$, and performs linear algebra on $Mac_d(\{f_1^*, \dots, f_m^*\}) \cup Mac(\{p_1^*, \dots, p_r^*\})$ where $f_i^*$ and $p_i^*$ are polynomials $f_i$ and $p_i$ specialized at $n-k$ variables, respectively. If the following inequality holds, then one can determine whether the system is consistent and the parameter set $(k, D, d)$ is available:

$$\mathrm{Rank}(Mac_d(\{f_1^*, \dots, f_m^*\}) \cup Mac(\{p_1^*, \dots, p_r^*\})) \geq \binom{k+d}{d} - 1.$$

In this paper, we introduce two inequalities (see Equation (4) and (8)) based on the above as a necessary condition for the admissibility and study the availability of a parameter set under the converse assumption. Note that the assumption is already introduced in [4] and [13]. We compute each inequality under some regularity conditions that a polynomial system with randomly chosen coefficients satisfies, and provide two formulae for determining the availability (see Theorem 1 and 2). Moreover, we present a formal power series description for each formula (see Theorem 7 and 8). For example, the second formula implies that under the above assumptions, the parameter set $(k, D, d)$ is admissible if the coefficient of $t_1^d t_2^D$ is less than or equal to one in the two-variable power series

$$\frac{(1-t_2^2)^m(1-t_2^q)^n}{(1-t_1)(1-t_2)^{n+1}} + \frac{(1-t_1^2)^m(1-t_1^q)^k}{(1-t_1)^{k+1}(1-t_2)} - \frac{(1-t_1^2t_2^2)^m(1-t_1^qt_2^q)^k}{(1-t_1t_2)^k(1-t_2)^{n-k+1}(1-t_1)}.$$

Furthermore, using these inequalities for finding an admissible parameter set, we provide the complexity estimation of the Crossbred algorithm.

### 1.2 Related work

Finding an admissible parameter set of Crossbred using a two-variable power series was studied by the paper [7,8] and the power series in [7,8] is similar to the above one. However, the paper uses undefined terms and a questionable claim that could not be confirmed by our experiments in the proof for deriving the power series (see Remark 9). In the NIST PQC standardization process for additional digital signatures, several submissions utilize the software `MQEstimator` [3] provided by Bellini et al. for security analysis. Bellini et al. [2] estimate the complexity of the Crossbred algorithm based on the paper [7] and `MQEstimator` employs the above two-variable power series to determine the admissibility of a parameter set. Our results support the complexity estimation for the Crossbred algorithm in `MQEstimator`.

## 2 Preliminaries

In this section, we introduce some notations used in this paper.

Let $n, m$ be two positive numbers, $q$ be a power of a prime number, and $\mathbb{F}$ be the finite field of order $q$. For a set $S$, denote by $\sharp S$ the cardinality of $S$. Denote by $Mat_{a\times b}(\mathbb{F})$ the set of $a \times b$ matrices over the field $\mathbb{F}$. For a matrix $M \in Mat_{a\times b}(\mathbb{F})$, we denote by $rows(M)$ the set of rows in $M$, and by $cols(M)$ the set of columns in $M$. Moreover, in this paper, we define the "corank" of $M$ as $\mathrm{Corank}(M) = \sharp cols(M) - \mathrm{Rank}(M)$.

Let $\mathbb{F}[x_1, \dots, x_n]$ be the polynomial ring in $n$ variables over a finite field $\mathbb{F}$. For $f \in \mathbb{F}[x_1, \dots, x_n]$, denote by $\deg f$ the total degree of $f$, and by $\deg_k f$ the degree of $f$ in the first $k$ variables $x_1, \dots, x_k$. Moreover, for a set $G \subset \mathbb{F}[x_1, \dots, x_n]$, the vector space $\{c_1 g_{i_1} + \cdots + c_r g_{i_r} \mid c_i \in \mathbb{F}, g_{i_j} \in G\}$ generated by $G$ is denoted by $\langle G\rangle_{\mathbb{F}}$ or simply $\langle G\rangle$. For a set $X = \{x_1, \dots, x_n\}$ of variables and a non-negative integer $d$, we define $Mon(X) = \{x_1^{e_1} \cdots x_n^{e_n} \mid e_i \in \mathbb{Z}_{\geq 0}\}, Mon_d(X) = \{x_1^{e_1} \cdots x_n^{e_n} \in Mon(X) \mid e_1 + \cdots + e_n = d\}$, and $Mon_{\leq d}(X) = \cup_{i=0}^{d} Mon_i(X)$. We denote by $\mathbb{F}[x_1, \dots, x_n]_d$ and $\mathbb{F}[x_1, \dots, x_n]_{\leq d}$ the vector spaces $\langle Mon_d(X)\rangle$ and $\langle Mon_{\leq d}(X)\rangle$, and these dimensions are coincide with the numbers $\binom{n+d-1}{d}$ and $\binom{n+d}{d}$ of monomials, respectively. Namely,

$$\mathbb{F}[x_1, \dots, x_n]_d \cong \mathbb{F}^{\binom{n+d-1}{d}} \text{ and } \mathbb{F}[x_1, \dots, x_n]_{\leq d} \cong \mathbb{F}^{\binom{n+d}{d}}.$$

For $f \in \mathbb{F}[x_1, \dots, x_n]$ with $\deg f \leq d$, we see $f \in \mathbb{F}[x_1, \dots, x_n]_{\leq d}$ and so denote by $vect_d(f)$ its corresponding vector in $\mathbb{F}^{\binom{n+d}{d}}$. Here, when we denote by $\mathrm{Coeff}(f, m_0)$ the coefficient of $m_0 \in Mon_{\leq d}(X)$ in $f$,

$$vect_d(f) = (\dots, \mathrm{Coeff}(f, m_0), \dots) \in \mathbb{F}^{\binom{n+d}{d}}.$$

Then, for $F = (f_1, \dots, f_m) \in \mathbb{F}[x_1, \dots, x_n]^m$, we define the *Macaulay* matrix $Mac(F)$ whose rows consist of $vect_d(f_i)$ where $1 \leq i \leq m$ and $d = \max\{\deg f_i\}_i$. Namely,

$$Mac(F) = \begin{pmatrix} vect_d(f_1) \\ \vdots \\ vect_d(f_m) \end{pmatrix} \in Mat_{m\times\binom{n+d}{d}}(\mathbb{F}).$$

Furthermore, for an integer $D$ and $F = (f_1, \dots, f_m)$ with $\deg f_1 = \cdots = \deg f_m = 2$, we can obtain a set $XM_D(F) := \{uf_i \mid 1 \leq i \leq m, u \in Mon_{\leq D-2}(X)\}$ and we call the operation obtaining $XM_D(F)$ from $F$ the *eXtension by multiplying with Monomials* (XM). For a positive integer $D$ and $F = (f_1, \dots, f_m)$ with $\deg f_1 = \cdots = \deg f_m = 2$, we define the *Macaulay matrix $Mac_D(F)$ of degree $D$* as

$$Mac_D(F) = Mac(XM_D(F)) \in Mat_{R\times\binom{n+D}{D}}(\mathbb{F}),$$

where $R = m\binom{n+D-2}{D-2}$.

## 3 Crossbred algorithm

In this section, we provide a brief overview of research on the MQ problem and describe the Crossbred algorithm.

### 3.1 Fukuoka MQ challenge

The following problem is called the *MQ problem*:

*Problem 1.* Given $f_1, \dots, f_m \in \mathbb{F}[x_1, \dots, x_n]$ such that $\deg f_1 = \cdots = \deg f_m = 2$. Find $(a_1, \dots, a_n) \in \mathbb{F}^n$ such that $f_1(a_1, \dots, a_n) = \cdots = f_m(a_1, \dots, a_n) = 0$.

If the number of equations in an instance is lower than that of variables, then the instance is said to be *underdetermined*. Otherwise, the instance is said to be *overdetermined*.

The MQ problem is known to be NP-complete [12], and it has various applications. In particular, in cryptography, multivariate polynomial cryptosystems based on the MQ problem are actively researched due to their expected resistance against attacks using a large-scale quantum computer. Against this backdrop, Fukuoka MQ Challenge [20] was initiated in 2015. In this challenge, two models, i.e. overdetermined ($m \approx 2n$) and underdetermined ($n \approx 1.5m$), are considered for both encryption and signature schemes of the cryptosystem. Each model sets different orders of a finite field, such as $\mathbb{F}_2, \mathbb{F}_{31}$, and $\mathbb{F}_{256}$, resulting in a total of six parameter sets (see Table 3.1). The Crossbred algorithm has been implemented by Kai-Chun Ning and Ruben Niederhagen [16] and Yao Sun and Ting Li [17], achieving record-breaking results in this challenge [20].

| Type | Model | Size | Field |
|---|---|---|---|
| I | Encryption | $m \approx 2n$ | $\mathbb{F}_2$ |
| II | Encryption | $m \approx 2n$ | $\mathbb{F}_{256}$ |
| III | Encryption | $m \approx 2n$ | $\mathbb{F}_{31}$ |
| IV | Signature | $n \approx 1.5m$ | $\mathbb{F}_2$ |
| V | Signature | $n \approx 1.5m$ | $\mathbb{F}_{256}$ |
| VI | Signature | $n \approx 1.5m$ | $\mathbb{F}_{31}$ |

**Table 1.** Fukuoka MQ challenge categories [20]

In the case of overdetermined instances in the challenge, it is expected that there are only a few solutions, while for underdetermined instances, there are at least $q^{n-m}$ solutions.

For solving an overdetermined MQ instance, various methods are known, including the Crossbred algorithm [13] and algorithms based on a Gröbner basis algorithm such as F4 [9] and F5 [10], and XL families [19]. These methods are often combined with the exhaustive search, as described in the following section.

On the other hand, for an underdetermined instance, it is possible to fix $n - m$ variables and we can reduce the instance to an overdetermined one.

Additionally, for an underdetermined instance, there is an approach by Kipnis, Patarin, and Goubin [14], which was further developed by Thomae and Wolf [18], to take advantage of sufficiently large variables. Thomae and Wolf's method allows us to reduce an underdetermined MQ instance to an overdetermined one with $m - \alpha$ equations and variables, where $\alpha = \lfloor m/n \rfloor - 1$. Moreover, an improved version of their method has been proposed by Furue et al. [11], which introduces an integer $k \geq 0$ and sets $\alpha_k = \lfloor (m-k)/(n-k) \rfloor - 1$, transforming an underdetermined instance into an overdetermined one with $m - k - \alpha_k$ variables and $m - \alpha_k$ equations.

### 3.2 Overview of the algorithm

In this section, we consider to solve an overdetermined quadratic system $F = \{f_1, \ldots, f_m\}$ where $f_i \in \mathbb{F}[x_1, \ldots, x_n]$, $\deg f_i = 2$, and $m \geq n$. Typically, we can use the Macaulay matrix of degree $D_{reg}(F)$ to solve the system $F$. The BooleanSolve algorithm, FXL algorithm, or their variants combines it with the exhaustive search. Indeed, after fixing $n - k$ variables in the input system, these algorithms determine if a resulting system has a solution by using its Macaulay matrix and solve the system if so. More precisely, for each $(a_{k+1}, \ldots, a_n) \in \mathbb{F}^{n-k}$, the following steps are performed:

1. Compute a polynomial system $F^* = \{f_1^*, \ldots, f_m^*\}$ in $k$ variables where $f_i^* = f_i(x_1, \ldots, x_k, a_{k+1}, \ldots, a_n)$.
2. Compute $Mac_d(F^*)$ at $d = D_{reg}(F^*)$ and determine if $F^*$ has a solution.

The most costly part of this process is the linear algebra on the Macaulay matrix $Mac_d(F^*)$, which needs to be performed $q^{n-k}$ times. To avoid this, Joux and Vitse proposed the Crossbred algorithm in [13], which performs a specialization step of $n - k$ variables after the linear algebra on a Macaulay matrix. More precisely, for parameters $1 \leq d < D$, the following steps are executed:

1. By the linear algebra on $Mac_D(F)$, obtain a polynomial system $P = \{p_1, \ldots, p_r\}$ whose degree in the first $k$ variables $x_1, \ldots, x_k$ is lower than or equal to $d$, namely $\deg_k p_i \leq d$.
2. For $(a_{k+1}, \ldots, a_n) \in \mathbb{F}^{n-k}$, perform the linear algebra on $Mac_d(F^*) \cup Mac(P^*)$ to determine if a solution exists. Here, $P^* = \{p_1^*, \ldots, p_r^*\}$ and $p_i^* = p_i(x_1, \ldots, x_k, a_{k+1}, \ldots, a_n)$.

By combining $Mac_d(F^*)$ with $Mac(P^*)$, the algorithm determines if the resulting system has a solution, whereas the previous algorithm only used $Mac_d(F^*)$. Therefore, the Crossbred algorithm allows to take $d < D_{reg}(F^*)$ if there exist a sufficiently number of independent polynomials $p_i^*$.

### 3.3 Description of the algorithm

First, the Crossbred algorithm generates the following matrices:

- $Mac_D(F)$: the Macaulay matrix of degree $D$ for the input system $F$
- $Mac_{D,d}^{(k)}(F)$: the submatrix of $Mac_D(F)$ whose each row is that of $Mac_D(F)$ corresponding to a polynomial $uf_i$ such that $\deg_k u \geq d-1$.
- $M_{D,d}^{(k)}(F)$: the submatrix of $Mac_{D,d}^{(k)}(F)$ whose each column is that of $Mac_{D,d}^{(k)}(F)$ corresponding to a monomial $m_0$ such that $\deg_k m_0 \geq d+1$.

Next, the Crossbred algorithm performs the following steps:

1. Search vectors $v_1, \dots, v_r$ in $\mathrm{Ker}(M_{D,d}^{(k)}(F))$.
2. Compute a polynomial $p_i$ corresponding to $v_i Mac_{D,d}^{(k)}(F)$ for each $1 \leq i \leq r$ and obtain $P = \{p_1, \dots, p_r\}$ such that $\deg_k p_i \leq d$.
3. For $(a_{k+1}, \dots, a_n) \in \mathbb{F}^{n-k}$,
   (a) Compute $F^* = (f_1^*, \dots, f_\ell^*)$ where $f_i^* = f_i(x_1, \dots, x_k, a_{k+1}, \dots, a_n)$ and obtain the Macaulay matrix $Mac_d(F^*)$.
   (b) Compute $P^* = (p_1^*, \dots, p_r^*)$ where $p_i^* = p_i(x_1, \dots, x_k, a_{k+1}, \dots, a_n)$ and obtain the Macaulay matrix $Mac(P^*)$.
   (c) By the linear algebra on $Mac_d(F^*) \cup Mac(P^*)$, determine if the resulting system in $k$ variables has a solution.

In the next section, we further consider the following matrices.

**Definition 1.** *Denote by $\widehat{Mac}_{D,d}^{(k)}(F)$ the submatrix disjoint with $Mac_{D,d}^{(k)}(F)$ in $Mac_D(F)$.*

*Remark 1.* Note that the polynomials corresponding to the rows in $\widehat{Mac}_{D,d}^{(k)}(F)$ do not contribute to give a new independent row outside $Mac_d(F^*)$. Namely, if $vect(g) \in \langle rows(\widehat{Mac}_{D,d}^{(k)}(F))\rangle$, then $vect(g^*) \in \langle rows(Mac_d(F^*))\rangle$. Indeed, $vect(g) \in \langle rows(\widehat{Mac}_{D,d}^{(k)}(F))\rangle$ implies there is a linear relation $g = c_1u_1f_{i_1} + \cdots + c_su_sf_{i_s}$ such that $\deg_k u_j \leq d-2$ and $\deg u_j \leq D-2$. Thus $\deg u_j^* \leq d-2$ and $g^* = c_1u_1^*f_{i_1}^* + \cdots + c_su_s^*f_{i_s}^* \in \langle XM_d(F^*)\rangle$. Namely, $vect(g^*) \in \langle rows(Mac_d(F^*))\rangle$.

## 4 Admissible parameter sets

In this section, we investigate an admissible parameter set for the Crossbred algorithm and provide two formulae for obtaining admissible parameter sets.

### 4.1 Strategy on our first formula

Let $X_k = \{x_1, \dots, x_k\}$. We consider a necessary condition for a parameter set to be admissible.

By Remark 1, a polynomial $p$ such that $vect(p) \in \langle \widehat{Mac}^{(k)}_{D,d}(F) \rangle$ do not contribute to give a new independent row outside $Mac_d(F^*)$. Thus we may assume the system $P$ satisfies that $vect_D(p_i) \notin \langle rows(\widehat{Mac}^{(k)}_{D,d}(F)) \rangle$ for any $p_i \in P$. Then $\mathrm{Rank}(P)$ consists of the dimension $L$ of the subspace $V$ such that

$$\langle f \in \langle XM_D(F) \rangle \mid \deg_k f \leq d \rangle = V \oplus \langle f \mid vect_D(f) \in rows(\widehat{Mac}^{(k)}_{D,d}(F) \rangle .$$

When the Crossbred algorithm works (see Section 3.3), the rank of the matrix $Mac_d(F^*) \cup Mac(P^*)$ needs at least $\sharp Mon_{\leq d}(X_k) - 1$ and it requires $\mathrm{Rank}(Mac(P^*)) \geq \sharp Mon_{\leq d}(X_k) - \mathrm{Rank}(Mac_d(F^*)) - 1$.

**Lemma 1.** *Let $(a_{k+1}, \dots, a_n) \in \mathbb{F}^{n-k}$, $P = (p_1, \dots, p_s) \in \mathbb{F}[x_1, \dots, x_n]^s$, and $P^* = (p_1^*, \dots, p_s^*) \in \mathbb{F}[x_1, \dots, x_k]^s$ where $p_i^* = p_i(x_1, \dots, x_k, a_{k+1}, \dots, a_n)$. Then* $\mathrm{Rank}(Mac(P)) \geq \mathrm{Rank}(Mac(P^*))$

*Proof.* Let $p_1, \dots, p_s \in \mathbb{F}[x_1, \dots, x_n]$ and $(a_{k+1}, \dots, a_n) \in \mathbb{F}^{n-k}$. If $p_1, \dots, p_s$ are linearly dependent, so are $p_1^*, \dots, p_s^*$. $\square$

Hence, by Lemma 1, we have

$$L \geq \sharp Mon_{\leq d}(X_k) - \mathrm{Rank}(Mac_d(F^*)) - 1 \tag{1}$$

Thus, if the parameter set $(k, D, d)$ is admissible, then the above inequality holds. We introduce the following assumption and will investigate a lower bound for $L$ for standing the above inequality (1).

**Assumption 1** *A parameter set $(k, D, d)$ is admissible if and only if the inequality* (1) *holds.*

Note that this assumption is considered in Joux and Vitse's work [13] (see also Remark 2) and Chen et al.'s work [4] (see Equation (2.5) in Section 2.2.4 of [4]) for $d = 1$.

We have

$$L = \dim\langle f \in \langle XM_D(F) \rangle \mid \deg_k f \leq d \rangle - \dim\langle f \mid vect_D(f) \in rows(\widehat{Mac}^{(k)}_{D,d}(F) \rangle . \tag{2}$$

When we consider the vector space $V_0$ such that $\langle XM_D(F) \rangle = V_0 \oplus \langle f \in \langle XM_D(F) \rangle \mid \deg_k f \leq d \rangle$, the leading monomials $m_0$ of $V_0$ are $\deg m_0 \leq D$ and $\deg_k m \geq d+1$. Namely, $\dim V_0 \leq \sharp\{m_0 \in Mon_{\leq D}(X) \mid \deg_k m_0 \geq d+1\} = \sharp cols(M^{(k)}_{D,d}(F))$. Since $\dim\langle XM_D(F) \rangle = \mathrm{Rank}(Mac_D(F))$, we have

$$\begin{aligned} \dim\langle f \in \langle XM_D(F) \rangle \mid \deg_k f \leq d \rangle &= \dim\langle XM_D(F) \rangle - \dim V_0 \\ &\geq \mathrm{Rank}(Mac_D(F)) - \sharp cols(M^{(k)}_{D,d}(F)). \end{aligned} \tag{3}$$

Meanwhile, $\dim\langle f \mid vect_D(f) \in rows(\widehat{Mac}^{(k)}_{D,d}(F)\rangle = \mathrm{Rank}(\widehat{Mac}^{(k)}_{D,d}(F)) \leq \sharp rows(\widehat{Mac}^{(k)}_{D,d}(F))$. Thus, by Equation (2), we have

$$L \geq Mac_D(F) - \sharp cols(M^{(k)}_{D,d}(F)) - \sharp rows(\widehat{Mac}^{(k)}_{D,d}(F)).$$

By summarizing the above discussion, the inequality (1) holds if

$$\begin{aligned}\mathrm{Rank}(Mac_D(F)) - \sharp rows(\widehat{Mac}^{(k)}_{D,d}(F)) - \sharp cols(M^{(k)}_{D,d}(F)) \\ \geq \sharp Mon_{\leq d}(X_k) - \mathrm{Rank}(Mac_d(F^*)) - 1.\end{aligned} \tag{4}$$

Therefore, we have the following:

**Proposition 1.** *Under Assumption 1, a parameter set* $(k, D, d)$ *is admissible if the inequality* (4) *holds.*

When $d = 1$, the matrices $\widehat{Mac}^{(k)}_{D,d}(F)$ and $Mac_d(F^*)$ are often empty and the inequality (4) implies

$$\mathrm{Rank}(Mac_D(F)) - \sharp cols(M^{(k)}_{D,1}) \geq \sharp Mon_{\leq 1}(X_k) - 1 = k.$$

It coincides with the inequality provided Joux and Vitse in [13] (see Remark 2).

*Remark 2.* In [13], Joux and Vitse investigate an admissible parameter $k$ of Crossbred algorithm with parameters $d = 1$ and $D \in \{3, 4\}$. As an admissible parameter, they take a parameter $k$ such that

$$\mathrm{Rank}(Mac_D(F)) - \sharp cols(M^{(k)}_{D,d}(F)) \geq k.$$ [1]

In particular, they implicitly assume Assumption 1.

### 4.2 Our first formula

In this subsection, we compute (4) under regularity assumptions for systems $F$ and $F^*$. The numbers of rows and columns of a Macaulay matrix are computed by the following remark:

*Remark 3.* The following assertions hold:

1. $cols(M^{(k)}_{D,d}(F))$ corresponds to monomials of degree $\leq D$ in $n$ variables whose degree in the first $k$ variables is larger than or equal to $d+1$, namely,

$$\sharp cols(M^{(k)}_{D,d}(F)) = \sum_{i=d+1}^{D} \sum_{j=0}^{D-i} \binom{k+i-1}{i}\binom{n-k+j-1}{j}.$$

[1] [13] actually wrote $\geq k+1$ but it does not need "$+1$".

2. $rows(\widehat{Mac}^{(k)}_{D,d}(F))$ corresponds to the set of a multiplication $uf_i$ of each polynomial $f_i$ and a monomial $u$ such that $\deg_k u \le d-2$ and $\deg u \le D-2$, i.e. $\{uf_i \mid u \in Mon_{\le D-2}(X), \deg_k u \le d-2, 1 \le i \le m\}$, namely,

$$\sharp rows(\widehat{Mac}^{(k)}_{D,d}(F)) = m \cdot \sum_{i=0}^{d-2} \sum_{j=0}^{D-2-i} \binom{k+i-1}{i} \binom{n-k+j-1}{j}.$$

Moreover, ranks of $Mac_D(F)$ and $Mac_d(F^*)$ are computed under a regularity assumption as follows:

**Assumption 2 (Regularity assumption)** *Let $F$ be a system of $m$ quadratic polynomials in $n$ variables and $I$ be the ideal generated by $F$. The dimension of $\mathbb{F}[x_1, \ldots, x_n]_{\le i}/I_i$ coincides with the coefficient of $t^i$ in*

$$\left[\frac{(1-t^2)^m}{(1-t)^{n+1}}\right]$$

*where $[\sum_{i \ge 0} a_i t^i] = \sum_{i \ge 0} \max\{a_i, 0\} t^i$.*

**Lemma 2.** *Let $S_{m,n} = 1/(1-t)^{n+1} - [(1-t^2)^m/(1-t)^{n+1}] \in \mathbb{Z}[[t]]$. If the quadratic system $G$ satisfies the regularity assumption, then we have*

$$\mathrm{Rank}(Mac_i(G)) = \mathrm{Coeff}\left(S_{m,n}, t^i\right).$$

Hence, we can compute the inequality (4) as follows:

**Proposition 2.** *Under the notations in Lemma 2, if $F$ and $F^*$ satisfy the regularity assumption, then the inequality* (4) *is*

$$\mathrm{Coeff}\left(S_{m,n}, t^D\right) - m \cdot \sum_{i=0}^{d-2} \sum_{j=0}^{D-2-i} \binom{k+i-1}{i} \binom{n-k+j-1}{j}$$
$$- \sum_{i=d+1}^{D} \sum_{j=0}^{D-i} \binom{k+i-1}{i} \binom{n-k+j-1}{j} \ge \binom{k+d}{d} - \mathrm{Coeff}\left(S_{m,k}, t^d\right) - 1 \quad (5)$$

Therefore, we have the following theorem:

**Theorem 1.** *Let $F$ be a quadratic system such that $F$ and $F^*$ satisfy the regularity assumption. Under Assumption 1, the parameter set $(D, d)$ is admissible if* (5) *holds.*

### 4.3 Strategy for our second formula

Define $XM_{\le(d,D)}(G) = \{ug \mid g \in G, \deg u_k \le d-2, u \in Mon_{\le D-2}(X)\}$ for a quadratic system $G$. Then all elements of $\langle XM_{\le(d,D)}(F^*)\rangle$ are contained in $XM_d(F^*)$ after the specialization on the last $n-k$ variables (see Remark 4). In other words, the elements of $\langle XM_{\le(d,D)}(F^*)\rangle$ do not contribute to increasing the rank of $Mac_d(F^*) \cup Mac(P^*)$. Hence, as shown in the beginning of Section 4.1, we consider the dimension $L'$ of the subspace $V'$ such that

$$\langle g \in \langle XM_{\leq D}(F)\rangle \mid \deg_k g \leq d\rangle = V' \oplus (\langle g \in \langle XM_{\leq D}(F)\rangle \mid \deg_k g \leq d\rangle \cap \langle XM_{\leq (d,D)}(F^*)\rangle).$$

For an admissible parameter set, it requires that the following inequality holds with the same reason as the inequality (1) (see Section 4.1):

$$\sharp L' \geq Mon_{\leq d}(X_k) - \mathrm{Rank}(Mac_d(F^*)) - 1. \tag{6}$$

Hence we consider a lower bound for $L'$ under the following assumption:

**Assumption 3** *A parameter set $(k, D, d)$ is admissible if and only if the inequality* (6) *holds.*

By definition,

$$L' = \dim\langle g \in \langle XM_D(X)\rangle \mid \deg_k g \leq d\rangle - \dim(\langle g \in \langle XM_D(X)\rangle \mid \deg_k g \leq d\rangle \cap \langle XM_{\leq (d,D)}(F^*)\rangle).$$

Since $\langle g \in \langle XM_D(X)\rangle \mid \deg_k g \leq d\rangle \cap \langle XM_{\leq (d,D)}(F^*)\rangle \subseteq \langle XM_{\leq (d,D)}(F^*)\rangle$ and the inequality (3), we have

$$L' \geq \mathrm{Rank}(Mac_D(F)) - \sharp cols(M_{D,d}^{(k)}(F)) - \dim\langle XM_{\leq (d,D)}(F^*)\rangle. \tag{7}$$

Therefore, the inequality (6) holds if

$$\begin{aligned} \mathrm{Rank}(Mac_D(F)) - \sharp cols(M_{D,d}^{(k)}(F)) - \dim\langle XM_{\leq (d,D)}(F^*)\rangle \\ \geq \sharp Mon_{\leq d}(X_k) - \mathrm{Rank}(Mac_d(F^*)) - 1. \end{aligned} \tag{8}$$

**Proposition 3.** *Under Assumption 3, a parameter set $(k, D, d)$ is admissible if the inequality* (8) *holds.*

*Remark 4.* Let $g \in \langle XM_{\leq (d,D)}(F^*)\rangle$. Then $g = u_1 f_{i_1}^* + \cdots + u_s f_{i_s}^*$ ($\deg u_i \leq d-2, \deg u_i \leq D-2$). Hence, $g^* = u_1^* f_{i_1}^* + \cdots + u_s^* f_{i_s}^* \in \langle XM_d(F^*)\rangle$ and $vect(g^*) \in \langle rows(Mac_d(F^*)\rangle$.

### 4.4 Our second formula

In this subsection, we consider computing the inequality (8), and by Remark 3 and Lemma 2 (see Section 4.2), it is sufficient to estimate $\dim\langle XM_{\leq (d,D)}(F^*)\rangle$.

We introduce the bi-degree $\deg_{\mathbb{Z}^2}$ on $\mathbb{F}[x_1, \ldots, x_n]$ defined by

$$\deg_{\mathbb{Z}^2} = (\deg_k, \deg).$$

Define $Mon_{(i,j)}(X) = \{u \in Mon(X) \mid \deg_{\mathbb{Z}} u = (i,j)\}$, $\mathbb{F}[x_1, \ldots, x_n]_{(i,j)} = \langle Mon_{(i,j)}(X)\rangle$, and $\mathbb{F}[x_1, \ldots, x_n]_{\leq (i,j)} = \oplus_{0 \leq e \leq i} \mathbb{F}[x_1, \ldots, x_n]_{(i-e, j-e)}$. Note that

$$f_1^*, \ldots, f_m^* \in \mathbb{F}[x_1, \ldots, x_n]_{\leq (2,2)}.$$

Meanwhile, we have the homogeneous system $(f_1^*)^h, \dots, (f_m^*)^h \in \mathbb{F}[x_0, \dots, x_n]$ by homogenizing $F^*$ with $x_0$, i.e. $(f_i^*)^h = x_0^2 f_i^*(x_1/x_0, \dots, x_k/x_0)$. The polynomial ring $\mathbb{F}[x_0, \dots, x_n]$ has the bi-degree $\deg_{\mathbb{Z}^2} = (\deg_k', \deg)$ where $\deg_k'$ is the degree in $k+1$ variables $x_0, \dots, x_k$. Then $\deg_{\mathbb{Z}^2} x_i$ is $(1,1)$ if $0 \le i \le k$, $(0,1)$ otherwise, and

$$(f_1^*)^h, \dots, (f_m^*)^h \in \mathbb{F}[x_0, \dots, x_n]_{(2,2)}.$$

In general, there is a correspondence

$$\begin{array}{rcl} \mathbb{F}[x_1, \dots, x_n]_{\le(i,j)} & \xrightarrow{\sim} & \mathbb{F}[x_0, \dots, x_n]_{(i,j)}, \\ f(x_1, \dots, x_n) & \mapsto & x_0^j f(x_1/x_0, \dots, x_n/x_0), \\ g(1, x_1, \dots, x_n) & \leftarrow\!\shortmid & g(x_0, x_1, \dots, x_n). \end{array} \tag{9}$$

Note that

$$\langle u \in Mon_{\le D} \mid \deg_k u \le d \rangle = \bigoplus_{0 \le i \le d, i \le j \le D} \mathbb{F}[x_1, \dots, x_n]_{(i,j)}$$

$$= \left( \bigoplus_{i=0}^{d} \mathbb{F}[x_1, \dots, x_n]_{\le(i,D)} \right) \oplus \left( \bigoplus_{j=d}^{D-1} \mathbb{F}[x_1, \dots, x_n]_{\le(d,j)} \right).$$

The correspondence (9) induces that between $I_{\le(i,j)} := \langle u_1 f_1^* + \dots + u_m f_m^* \mid \deg_k u_\ell f_\ell^* \le i, \deg u_\ell f_\ell^* \le j \rangle$ and $I^h_{(i,j)} := \langle u_1 (f_1^*)^h + \dots + u_m (f_m^*)^h \mid \deg_k u_m = i-2, \deg u_i (f_i^*)^h = j \rangle$. Since $\langle MX_{\le(d,D)}(F^*) \rangle \subseteq \sum_{i=0}^{d} I_{\le(i,D)} + \sum_{j=d}^{D-1} I_{\le(d,j)}$, we obtain

$$\dim \langle MX_{\le(d,D)}(F^*) \rangle \le \sum_{i=0}^{d} \dim I_{\le(i,D)} + \sum_{j=d}^{D-1} \dim I_{\le(d,j)} = \sum_{i=0}^{d} \dim I^h_{(i,D)} + \sum_{j=d}^{D-1} \dim I^h_{(d,j)}$$

We can compute the right hand side of this inequality under the following regularity assumption:

**Assumption 4 (Bi-graded regularity assumption)** *For $i \le D_{reg}(m,k)$ and $j \le D_{reg}(m,n)$, $\dim \mathbb{F}[x_0, \dots, x_n]_{(i,j)} / I^h_{(i,j)}$ coincides with the coefficient of $t_1^i t_2^j$ in*

$$\left[ \frac{(1 - t_1^2 t_2^2)^m}{(1 - t_1 t_2)^{k+1} (1 - t_2)^{n-k}} \right]$$

*where $[\sum_{(d_1,d_2) \in \mathbb{Z}^2} \alpha_{(d_1,d_2)} t_1^{d_1} t_2^{d_2}] = \sum_{(d_1,d_2) \in \mathbb{Z}^2} \max\{\alpha_{(d_1,d_2)}, 0\} t_1^{d_1} t_2^{d_2}$.*

Note that for polynomial systems with randomly chosen coefficients, we confirmed in our experiments that the above assumption holds.

**Lemma 3.** *We have that $\dim \mathbb{F}[x_0, \dots, x_n]_{(i,j)}$ is $\binom{k+i}{i} \binom{n-k+j-i-1}{j-i}$ and coincides with the coefficient of $t_1^i t_2^j$ in*

$$\frac{1}{(1 - t_1 t_2)^{k+1} (1 - t_2)^{n-k}}.$$

**Lemma 4.** *If $(F^*)^h$ satisfies the bi-graded regularity assumption, we have*

$$\dim\langle XM_{\leq(d,D)}(F^*)\rangle \leq \sum_{i=0}^{d} \mathrm{Coeff}(S', t_1^i t_2^D) + \sum_{j=d}^{D-1} \mathrm{Coeff}(S', t_1^d t_2^j) \tag{10}$$

*where*

$$S' = \frac{1}{(1-t_1t_2)^{k+1}(1-t_2)^{n-k}} - \left[\frac{(1-t_1^2t_2^2)^m}{(1-t_1t_2)^{k+1}(1-t_2)^{n-k}}\right].$$

Then, by Remark 3 and Lemma 2, the inequality (8) holds if

$$\mathrm{Coeff}(S_{m,n}, t^D) - N_{d,D} - \sum_{i=0}^{d} \mathrm{Coeff}(S', t_1^i t_2^D) - \sum_{j=d}^{D-1} \mathrm{Coeff}(S', t_1^d t_2^j)$$

$$\geq \binom{k+d}{d} - \mathrm{Coeff}(S_{m,k}, t^d) - 1. \tag{11}$$

Namely, then the inequality (6) holds. Therefore, we have the following theorem:

**Theorem 2.** *Let $F$ be a quadratic system such that $F$ and $F^*$ satisfy the regularity assumption and $(F^*)^h$ satisfies the bi-graded regularity assumption. Then, under Assumption 3, a parameter set $(k, D, d)$ is admissible if the inequality* (11) *holds.*

*Remark 5.* When $q = 16$, $m = 9$, and $n = 8$, our experiments show that the parameter sets $(k, D, d) = (3, 2, 1)$ and $(4, 5, 4)$ are available in the Crossbred algorithm. The first formula (5) can detect the availability of $(k, D, d) = (3, 2, 1)$, but the third formula (11) cannot. Conversely, the second formula recognizes the availability of $(k, D, d) = (4, 5, 4)$ but the first formula fails.

### 4.5 Description using the binomial coefficients and the formal power series

In this subsection, for two inequalities (5) in Theorem 1 and (11) in Theorem 2, we present these binomial coefficient description and formal power series description. Theorem 3 and 4 are binomial-coefficient descriptions of Theorem 1 and 2, respectively. Theorem 5 and 6 are formal power series descriptions for Theorem 1 and 2, respectively. Note that these formal-power-series description use a slightly stronger assumption than each theorem, but they are sufficient (see Remark 8).

*Binomial coefficient description:* First, for two inequalities (5) and (11), we give these binomial-coefficient description.

*Remark 6.* Let $H_{m,n} = (1-t^2)^m(1-t)^{-n-1}$. Then

$$\mathrm{Coeff}(H_{m,n}, t^i) = \sum_{\ell=0}^{m}(-1)^\ell \binom{m}{\ell}\binom{n+i-2\ell-1}{i-2\ell},$$

$$\mathrm{Coeff}([H_{m,n}], t^i) = \max\left\{\sum_{\ell=0}^{m}(-1)^\ell \binom{m}{j}\binom{n+i-2\ell-1}{i-2\ell}, 0\right\}.$$

**Definition 2.** *Define* $D_{reg}(m,n) = \min\{i \mid \mathrm{Coeff}(H_{m,n}, t^i) \le 0\}$.

**Theorem 3.** *Let $F$ be a quadratic system such that $F$ and $F^*$ satisfy the regularity assumption. Under Assumption 1, the parameter set $(D,d)$ is admissible if*

$$\sum_{i=0}^{d}\sum_{j=0}^{D-i}\binom{k+i-1}{i}\binom{n-k+j-1}{j} - m\cdot\sum_{i=0}^{d-2}\sum_{j=0}^{D-2-i}\binom{k+i-1}{i}\binom{n-k+j-1}{j}$$

$$\ge \max\left\{\sum_{\ell=0}^{m}(-1)^\ell\binom{m}{\ell}\binom{n+D-2\ell-1}{D-2\ell}, 0\right\} + \max\left\{\sum_{\ell=0}^{m}(-1)^\ell\binom{m}{\ell}\binom{k+d-2\ell-1}{d-2\ell}, 0\right\} - 1$$

The following remark also follows from a straightforward calculation:

*Remark 7.* Let $H' = (1-t_1^2t_2^2)^m(1-t_1t_2)^{-k-1}(1-t_2)^{-n+k}$. Then

$$\mathrm{Coeff}(H', t_1^i t_2^j) = \sum_{\ell=0}^{m}(-1)^\ell\binom{m}{\ell}\binom{k+i-2\ell}{i-2\ell}\binom{n-k+j-i-1}{j-i},$$

$$\mathrm{Coeff}([H'], t_1^i t_2^j) = \max\left\{\sum_{\ell=0}^{m}(-1)^\ell\binom{m}{\ell}\binom{k+i-2\ell}{i-2\ell}\binom{n-k+j-i-1}{j-i}, 0\right\}.$$

**Theorem 4.** *Let $F$ be a quadratic system such that $F$ and $F^*$ satisfy the regularity assumption and $(F^*)^h$ satisfies the bi-graded regularity assumption. Then, under Assumption 3, a parameter set $(k,D,d)$ is admissible if*

$$\sum_{i=0}^{d}\sum_{j=i}^{D}\max\left\{\sum_{\ell=0}^{m}(-1)^\ell\binom{m}{\ell}\binom{k+i-2\ell}{i-2\ell}\binom{n-k+j-i-1}{j-i}, 0\right\} - \sum_{i=0}^{d}\binom{k+i-1}{i-1}\binom{n-k+D-i}{D-i}$$

$$\ge \max\left\{\sum_{j=0}^{m}(-1)^j\binom{m}{j}\binom{n+D-2j-1}{D-2j}, 0\right\} + \max\left\{\sum_{j=0}^{m}(-1)^j\binom{m}{j}\binom{k+d-2j-1}{d-2j}, 0\right\} - 1.$$

*Formal power series description:* We present power series descriptions for two inequalities (5) and (11). Determining the admissibility of a parameter set is essential when $d < D_{reg}(m,k)$ and $D < D_{reg}(m,n)$ (see Remark 8).

**Lemma 5.** *The number of monomials of degree $\leq D$ in $n$ variables whose degree in the first $k$ variables is lower than or equal to $d$, i.e. $\sum_{i=0}^{d}\sum_{j=0}^{D-i}\binom{k+i-1}{i}\binom{n-k+j-1}{j}$, is the coefficient of $t_1^d t_2^D$ in the following power series:*

$$\frac{1}{(1-t_1t_2)^k(1-t_2)^{n-k+1}(1-t_1)}$$

**Theorem 5.** *Let $F$ be a quadratic system such that $F$ and $F^*$ satisfy the regularity assumption. Set $d < D_{reg}(m,k)$ and $D < D_{reg}(m,n)$. Under the assumptions in Proposition 2, the parameter set $(D,d)$ is admissible if the coefficient of $t_1^d t_2^D$ in the following power series is less than or equal to 1:*

$$\frac{mt_1^2t_2^2-1}{(1-t_1t_2)^k(1-t_2)^{n-k+1}(1-t_1)}+\frac{(1-t_2^2)^m}{(1-t_1)(1-t_2)^{n+1}}+\frac{(1-t_1^2)^m}{(1-t_1)^{k+1}(1-t_2)}. \tag{12}$$

*Proof.* See Appendix A. □

Next, we consider a formal power series description for (11). Define

$$S'' = \frac{1}{(1-t_1t_2)^{k+1}(1-t_2)^{n-k}} - H'.$$

Since $\mathrm{Coeff}(H', t_1^i t_2^j) \leq \mathrm{Coeff}([H'], t_1^i t_2^j)$, we obtain

$$\mathrm{Coeff}(S'', t_1^i t_2^j) \geq \mathrm{Coeff}(S', t_1^i t_2^j). \tag{13}$$

Thus, if we take $S''$ as $S'$, then the inequality (10) still holds and it gives a lower bound for $L'$, i.e. the inequality (7). As result, we obtain a slightly strong version of Theorem 2:

**Theorem 6.** *Let $F$ be a quadratic system such that $F$ and $F^*$ satisfy the regularity assumption and $(F^*)^h$ satisfies the bi-graded regularity assumption. Set $d < D_{reg}(m,k)$ and $D < D_{reg}(m,n)$. Then, under Assumption 3, a parameter set $(k,D,d)$ is admissible if the coefficient of $t_1^d t_2^D$ in the following power series is less than or equal to 1:*

$$\frac{(1-t_1^2)^m}{(1-t_1)^{k+1}(1-t_2)}+\frac{(1-t_2^2)^m}{(1-t_1)(1-t_2)^{n+1}}-\frac{(1-t_1^2t_2^2)^m}{(1-t_1t_2)^k(1-t_1)(1-t_2)^{n-k+1}} \tag{14}$$

*Proof.* See Appendix B. □

Here, we can provide finite field versions of Theorem 5 and 6 as applications. We define $D_{reg}(m,n,q) = \min\{i \mid \mathrm{Coeff}(H^q_{m,n}, t^i) \leq 0\}$ where $H^q_{m,n} = (1-t^q)^n(1-t^2)^m(1-t)^{-n}$. Then, Assumption 2 and 4 are slightly modified as follows.

**Assumption 5 (Regularity assumption)** *Let $F$ be a system of $m$ quadratic polynomials in $n$ variables and $I$ be the ideal generated by $F$. The dimension of $\mathbb{F}[x_1, \ldots, x_n]_{\le i}/I_i$ coincides with the coefficient of $t^i$ in*

$$\left[\frac{(1-t^2)^m(1-t^q)^n}{(1-t)^{n+1}}\right].$$

**Assumption 6 (Bi-graded regularity assumption)** *For $i \le D_{reg}(m,k,q)$ and $j \le D_{reg}(m,n,q)$, $\dim \mathbb{F}[x_0,\ldots,x_n]_{(i,j)}/I^h_{(i,j)}$ coincides with the coefficient of $t_1^i t_2^j$ in*

$$\left[\frac{(1-t_1^2t_2^2)^m(1-t_1^qt_2^q)^k}{(1-t_1t_2)^{k+1}(1-t_2)^{n-k}}\right].$$

Then, we obtain the following theorem for Theorem 5:

**Theorem 7.** *Let $F$ be a system such that contains the field equations and $F$ and $F^*$ satisfy the regularity assumption. Set $d < D_{reg}(m,k,q)$ and $D < D_{reg}(m,n,q)$. Under the assumptions in Proposition 2, the parameter set $(D,d)$ is admissible if the coefficient of $t_1^d t_2^D$ in the following power series is less than or equal to 1:*

$$\frac{-1+m\cdot t_1^2t_2^2+k\cdot t_1^qt_2^q+(n-k)\cdot t_2^q}{(1-t_1t_2)^k(1-t_2)^{n-k+1}(1-t_1)}+\frac{(1-t_2^2)^m(1-t_2^q)^n}{(1-t_1)(1-t_2)^{n+1}}+\frac{(1-t_1^2)^m(1-t_1^q)^k}{(1-t_1)^{k+1}(1-t_2)} \tag{15}$$

For Theorem 6, we obtain the following theorem:

**Theorem 8.** *Let $F$ be a system such that contains the field equations, $F$ and $F^*$ satisfy the regularity assumption, and $(F^*)^h$ satisfies the bi-graded regularity assumption. Set $d < D_{reg}(m,k,q)$ and $D < D_{reg}(m,n,q)$. Then, under Assumption 3, a parameter set $(k,D,d)$ is admissible if the coefficient of $t_1^d t_2^D$ in the following power series is less than or equal to 1:*

$$\frac{(1-t_2^2)^m(1-t_2^q)^n}{(1-t_1)(1-t_2)^{n+1}}+\frac{(1-t_1^2)^m(1-t_1^q)^k}{(1-t_1)^{k+1}(1-t_2)}-\frac{(1-t_1^2t_2^2)^m(1-t_1^qt_2^q)^k}{(1-t_1t_2)^k(1-t_2)^{n-k+1}(1-t_1)} \tag{16}$$

Note that their proofs are exactly the same as Theorem 5 and 6.

*Remark 8.* Assume that $F$ and $F^*$ satisfies the regularity assumption. We recall Section 3.2. In the case $D \ge D_{reg}(m,n)$, the vector space $\langle XM_D(F)\rangle$ has all monomials of degree $\le D$ as a leading monomial. Hence we can obtain a sufficiently number of linearly independent polynomials after the specialization on the last $n-k$ variables. In particular, then the parameter set is admissible. In the case $d \ge D_{reg}(m,k)$, the Macaulay matrix $Mac_d(F^*)$ has a sufficiently large rank and we can determine if there exists a solution. Then the parameter set is admissible. Moreover, we do not need to perform the linear algebra on $M^{(k)}_{D,d}(F)$. In particular, Crossbred algorithm becomes BooleanSolve/FXL actually.

*Remark 9.* The paper [7,8] also considered the admissibility of a parameter set but omitted certain assumptions and crucial discussions. For example, they claim that $\sum_{i,j} \mathrm{Corank}(M_{j,i}^{(k)}(F))t_1^i t_2^j = H'$ (see Equation (17) in [7] and Equation (15) in [8]) but our experiment did not support this assertion. Indeed, the columns of $M_{D,d}^{(k)}(F)$ corresponds to the monomial $\{m_0 \in Mon_{\leq D}(X) \mid \deg_k m_0 \geq d+1\}$, and the coefficient of $t_1^d t_2^D$ in the denominator part of $H'$, i.e. $1/(1-t_1t_2)^k(1-t_2)^{n-k}$, implies $\sharp\{m_0 \in Mon_D(X) \mid \deg_k m_0 = d\}$. Hence, even if one introduce some regularity assumption, one cannot conclude that the equality holds.

## 5 Complexity estimation

In this section, we provide the complexity estimation for the Crossbred algorithm using our formulae presented in the previous section.

The framework of a complexity estimation for the Crossbred algorithm is firstly given by Chen et al. [4]. In this section, we use a slightly detailed version by Bellini et al. [2] for a parameter set which given by our formulae in the previous section.

For a system of $m$ quadratic equations in $n$ variables, if a parameter set $(k, D, d)$ is admissible, the (time) complexity of the Crossbred algorithm is estimated by

$$\mathrm{Cost}_{(k,D,d)}(q,m,n) = \min\left\{\mathcal{O}\left(N_{d,D}^{\omega}\right), \mathcal{O}\left(3\binom{n+2}{2}N_{d,D}^2 \cdot N_{itr}\right)\right\}$$
$$+ \mathcal{O}\left(m \cdot q^{n-k} \cdot \binom{k+d}{d}^{\omega}\right), \tag{17}$$

where $2 \leq \omega \leq 3$ is a linear algebra constant,

$$N_{d,D} = \sum_{i=d+1}^{D}\sum_{j=0}^{D-i}\binom{k+i-1}{i}\binom{n-k+j-1}{j}, \text{ and}$$

$$N_{itr} = \max\left\{\sum_{\ell=0}^{m}(-1)^{\ell}\binom{m}{\ell}\binom{n+d-2\ell-1}{d-2\ell}, 0\right\} - 1 \;(= \mathrm{Coeff}([H_{m,k}], t^d) - 1).$$

Moreover, the memory complexity is estimated by

$$\mathcal{O}\left(N_{d,D}^2 + \binom{k+d}{d}^2\right).$$

In (17), the first term is the complexity estimation for searching left kernel vectors of $M_{D,d}^{(k)}$ in the Crossbred algorithm (see Section 3.3), and given by the minimum value of the Gaussian elimination and a kernel search using the block Wiedemann algorithm $N_{itr}$ times [5]. The second term is the complexity estimation for the dense linear algebra on $Mac(P^*) \cup Mac_d(F^*)$ in the algorithm (see Remark 10).

We obtain the complexity estimation of the Crossbred algorithm as the minimum value of (17) for a parameter set which is given by our formula (5) or (11). Using the Thomae-Wolf (TW) method [18], we give the complexity estimation for solving an MQ instance as

$$\min_{(k,D,d)\text{: admissibe}} \mathrm{Cost}_{(k,D,d)}(q, m-\alpha, m-\alpha) \tag{18}$$

where $\alpha = \lfloor n/m \rfloor - 1$.

*Remark 10.* Since $Mac_D(F)$ and $M^{(k)}_{D,d}(F)$ become sparse matrices, we can utilize the block Lanczos algorithm [15], the block Wiedemann algorithm [5], and these modifications. However, due to the density of $Mac(P^*)$, the efficiency of these algorithms is not guaranteed [4].

## References


1. Bardet, M., Faugère, J.C., Salvy, B., Spaenlehauer, P.J.: On the complexity of solving quadratic Boolean systems. Journal of Complexity 29(1), 53–75 (2013). `https://doi.org/https://doi.org/10.1016/j.jco.2012.07.001`
2. Bellini, E., Makarim, R.H., Sanna, C., Verbel, J.: An Estimator for the Hardness of the MQ Problem. Progress in Cryptology - AFRICACRYPT 2022. LNCS, vol. 13503. Springer, Cham. `https://doi.org/10.1007/978-3-031-17433-9_14`
3. Bellini, E., Makarim, R., Verbel, J.: An estimator for the complexity of the MQ problem. (2021), `https://github.com/Crypto-TII/multivariate_quadratic_estimator`
4. Chen, M.S., Hulsing, A., Rijneveld, J., Samardjiska, S., Schwabe, P.: MQDSS specifications version 1.0 (November 2017), `http://mqdss.org/specification.html`
5. Coppersmith, D.: Solving homogeneous linear equations over GF(2) via block Wiedemann algorithm. Mathematics of Computation 62, 333–350 (1994)
6. Courtois, N., Klimov, A., Patarin, J., Shamir, A.: Efficient algorithms for solving overdefined systems of multivariate polynomial equations. EUROCRYPT 2000, LNCS vol. 1807, 392–407 (2000)
7. Duarte, J.D.: On the complexity of the crossbred algorithm. Cryptology ePrint Archive, Report 2020/1058 (2020), https://eprint.iacr.org/2020/1058 (Withdrawn at 2022/04/18)
8. Duarte, J.D.: On the Complexity and Admissible Parameters of the Crossbred Algorithm in $\mathbb{F}_{q\geq 2}$. Cryptology ePrint Archive, Report 2023/1664 (2023), https://eprint.iacr.org/2023/1664
9. Faugère, J.C.: A new efficient algorithm for computing Gröbner bases (F4). Journal of Pure and Applied Algebra 139(1), 61–88 (1999)
10. Faugère, J.C.: A new efficient algorithm for computing Gröbner bases without reduction to zero (F5). In: Proceedings of the 2002 International Symposium on Symbolic and Algebraic Computation. p. 75–83. ISSAC '02, Association for Computing Machinery, New York, NY, USA (2002)

11. Furue, H., Nakamura, S., Takagi, T.: Improving Thomae-Wolf algorithm for solving underdetermined multivariate quadratic polynomial problem. In: Cheon, J.H., Tillich, J.P. (eds.) PQcrypto 2021. LNCS. pp. 65–78. Springer International Publishing, Cham (2021)
12. Garey, M.-R., Johnson, D.-S.: Computers and intractability: a guide to the theory of NP-completeness. W.H. Freeman (1979).
13. Joux, A., Vitse, V.: A crossbred algorithm for solving Boolean polynomial systems. In: Kaczorowski, J., Pieprzyk, J., Pomykala, J. (eds.) Number-Theoretic Methods in Cryptology. pp. 3–21. Springer International Publishing, Cham (2018)
14. Kipnis, A., Patarin, J., Goubin, L.: Unbalanced oil and vinegar signature schemes. In: Stern, J. (ed.) Advances in Cryptology – EUROCRYPT '99. pp. 206–222. Springer Berlin Heidelberg, Berlin, Heidelberg (1999)
15. Montgomery, P.L.: A block lanczos algorithm for finding dependencies over gf(2). In: Guillou, L.C., Quisquater, J.J. (eds.) Advances in Cryptology – EUROCRYPT '95: International Conference on the Theory and Application of Cryptographic Techniques Saint-Malo, France, May 21–25, 1995 Proceedings. pp. 106–120. Springer Berlin Heidelberg, Berlin, Heidelberg (1995), https://doi.org/10.1007/3-540-49264-X_9
16. Niederhagen, R., Ning, K.-C., Yang, B.-Y.: Implementing Joux-Vitse's Crossbred Algorithm for Solving MQ Systems over GF(2) on GPUs. PQCrypto 2018, pp. 121–141 (2018).
17. Sun, Y., Li, T.: Implementations in Type IV of Fukuoka MQ Challenge, https://www.mqchallenge.org/details/details_IV_20200608.html
18. Thomae, E., Wolf, C.: Solving underdetermined systems of multivariate quadratic equations revisited. In: Fischlin, M., Buchmann, J., Manulis, M. (eds.) Public Key Cryptography – PKC 2012. pp. 156–171. Springer Berlin Heidelberg (2012)
19. Yang, B.Y., Chen, J.M.: Theoretical analysis of XL over small fields. In: Wang, H., Pieprzyk, J., Varadharajan, V. (eds.) Information Security and Privacy. pp. 277–288. Springer Berlin Heidelberg, Berlin, Heidelberg (2004)
20. Yasuda, T., Dahan, X., Huang, Y.J., Takagi, T., Sakurai, K.: MQ challenge: Hardness evaluation of solving multivariate quadratic problems. In: NIST Workshop on Cybersecurity in a Post-Quantum World (2015), washington, D.C. https://www.mqchallenge.org

# Appendix

## Appendix A

The inequality (5) is rewritten as

$$-\mathrm{Coeff}\left(S_{m,n}, t^D\right) - \mathrm{Coeff}\left(S_{m,k}, t^d\right) + m \cdot \sum_{i=0}^{d-2} \sum_{j=0}^{D-2-i} \binom{k+i-1}{i}\binom{n-k+j-1}{j} + \sum_{i=d+1}^{D} \sum_{j=0}^{D-i} \binom{k+i-1}{i}\binom{n-k+j-1}{j} + \binom{k+d}{d} \leq 1.$$

By Lemma 2 and 5, the left hand side is the coefficient of $t_1^d t_2^D$ in

$$-\frac{1-(1-t_2^2)^m}{(1-t_1)(1-t_2)^{n+1}} - \frac{1-(1-t_1^2)^m}{(1-t_1)^{k+1}(1-t_2)} + \frac{m t_1^2 t_2^2}{(1-t_1 t_2)^k (1-t_2)^{n-k+1}(1-t_1)}$$

$$+\frac{1}{(1-t_1)(1-t_2)^{n+1}}-\frac{1}{(1-t_1t_2)^k(1-t_2)^{n-k+1}(1-t_1)}+\frac{1}{(1-t_1)^{k+1}(1-t_2)}.$$

Arranging this power series, we have the asserted series.

**Appendix B**

We have

$$\sum_{i=0}^{d}\dim I^h_{(i,D)}+\sum_{j=d}^{D-1}\dim I^h_{(d,j)}=\sum_{0\le i\le d, i\le j\le D}(\dim I^h_{(i,j)}-\dim I^h_{(i-1,j-1)}),\ \text{and}$$

$$\sum_{i=0}^{d}\mathrm{Coeff}(S'',t_1^it_2^D)+\sum_{j=d}^{D-1}\mathrm{Coeff}(S'',t_1^dt_2^j)=\sum_{0\le i\le d, i\le j\le D}\{\mathrm{Coeff}(S'',t_1^it_2^j)-\mathrm{Coeff}(t_1t_2\cdot S'',t_1^it_2^j)\}$$

$$=\sum_{0\le i\le d, i\le j\le D}\mathrm{Coeff}(S''-t_1t_2S'',t_1^it_2^j)=\sum_{0\le i\le d, 0\le j\le D}\mathrm{Coeff}(S''-t_1t_2S'',t_1^it_2^j).$$

Note that the last equality follows from that $S''-t_1t_2S''$ does not have monomials $t_1^it_2^j$ $(i>j)$ as follows:

$$\begin{aligned}S''-t_1t_2S''&=\frac{1-(1-t_1^2t_2^2)^m}{(1-t_1t_2)^{k+1}(1-t_2)^{n-k}}-\frac{t_1t_2\{1-(1-t_1^2t_2^2)^m\}}{(1-t_1t_2)^{k+1}(1-t_2)^{n-k}}\\&=\frac{1-t_1t_2}{(1-t_1t_2)^{k+1}(1-t_2)^{n-k}}-\frac{(1-t_1t_2)(1-t_1^2t_2^2)^m}{(1-t_1t_2)^{k+1}(1-t_2)^{n-k}}\\&=\frac{1}{(1-t_1t_2)^{k}(1-t_2)^{n-k}}-\frac{(1-t_1^2t_2^2)^m}{(1-t_1t_2)^{k}(1-t_2)^{n-k}}.\end{aligned}\qquad(19)$$

By Lemma 4 and the inequality (13), we obtain

$$\dim\langle XM_{\le(d,D)}(F^*)\rangle\le\sum_{0\le i\le d, 0\le j\le D}\mathrm{Coeff}(S''-t_1t_2S'',t_1^it_2^i).$$

Then the inequality (11) is rewritten as

$$-\mathrm{Coeff}(S_{m,n},t^D)+N_{d,D}+\sum_{0\le i\le d, 0\le j\le D}\mathrm{Coeff}(S''-t_1t_2S'',t_1^it_2^i)$$
$$+\binom{k+d}{d}-\mathrm{Coeff}(S_{m,k},t^d)\le 1.$$

By Equation (19), $\sum_{0\le i\le d,0\le j\le D}\mathrm{Coeff}(S''-t_1t_2S'',t_1^it_2^i)$ coincides with the coefficient of $t_1^dt_2^D$ in

$$\frac{1}{(1-t_1t_2)^k(1-t_1)(1-t_2)^{n-k+1}}-\frac{(1-t_1^2t_2^2)^m}{(1-t_1t_2)^k(1-t_1)(1-t_2)^{n-k+1}}.$$

Then, as the same discussion shown in Appendix A, the inequality (6) holds if the coefficient of $t_1^dt_2^D$ in the following power series is less than or equal to 1:

$$\frac{(1-t_1^2)^m}{(1-t_1)^{k+1}(1-t_2)}+\frac{(1-t_2^2)^m}{(1-t_1)(1-t_2)^{n+1}}-\frac{(1-t_1^2t_2^2)^m}{(1-t_1t_2)^k(1-t_1)(1-t_2)^{n-k+1}}.$$