MAXMINMAX PROBLEM AND SPARSE EQUATIONS OVER FINITE FIELDS

IGOR SEMAEV

ABSTRACT. Asymptotical complexity of sparse equation systems over finite field F_q is studied. Let the variable sets belong to a fixed family $\mathcal{X} = \{X_1, \ldots, X_m\}$ while the polynomials $f_i(X_i)$ are taken independently and uniformly at random from the set of all polynomials of degree $\leq q-1$ in each of the variables in X_i. In particular, for $|X_i| \leq 3$, $m = n$, we prove the average complexity of finding all solutions to $f_i(X_i) = 0, i = 1, \ldots, m$ by Gluing algorithm (Semaev, Des. Codes Cryptogr., vol. 49 (2008), pp.47–60) is at most $q^{\frac{n}{5.7883}+O(\log n)}$ for arbitrary $\mathcal{X}$ and q. The proof results from a detailed analysis of 3-MaxMinMax problem, a novel problem for hyper-graphs.

1. INTRODUCTION

1.1. **Sparse Equations over Finite Fields.** Let (q, l, n, m) be a quadruple of natural numbers, where q is a prime power. Let F_q denote the finite field with q elements and let $\{x_1, x_2, \ldots, x_n\}$ be a set of variables over F_q. We consider systems of equations of type

$$f_1(X_1) = 0, \ldots, f_m(X_m) = 0 \tag{1}$$

where each X_i $(1 \leq i \leq m)$ is a subset of $\{x_1, x_2, \ldots, x_n\}$ and the f_i are polynomials over F_q, which only depend on the variables in X_i. Such a system of equations is called l-sparse if each X_i has size at most l. We study the complexity (average complexity) of solving (1) in F_q. Therefore it suffices to consider only polynomials which have degree at most $q-1$ in each of the variables.

Sparse equations have a natural application in inverting a discrete function $G(x) = y$. The problem is fundamental in cryptography. For instance, G may be a block cipher encryption and its inverting recovers the encryption key; or discrete exponentiation and its inverting is the discrete logarithm problem. The values of G must be computed efficiently to provide fast block cipher encryption or key generation according to Diffie-Hellman protocol.

Therefore y is computed by a circuit with a low number of small gates, that is functions $g_i(x_i) = y_i$ with bounded number of e.g. Boolean variables. To invert G one writes bits of x_i, y_i as new variables and gets a system of sparse equations. Sparse equation systems in this relation were for the first time studied in [22] and later in [16], where a similar guess-and-determine solving algorithm was independently suggested. By experiments, it was there realised that the expansion of a guess $x_i = a$ over (1) is enhanced if the number of

Date: January 2, 2014.

This work was supported by SPIRE program at the University of Bergen.

the solutions to each particular equation in (1) taken separately is low. That in turn reduces the algorithm running time. However no complexity bounds based on this observation were given.

This is an area of algebraic cryptanalysis. In contrast with linear and differential cryptanalysis the former requires only few plain-text/cipher-text pairs to construct a system of equations the solving of which yields the encryption key. Several solving algorithms are known. They differ in how the equations are represented. The representation should be sparse in one or other way to keep the equations in computer memory even before solving.

For polynomial representation Gröbner basis algorithms, and similar as extended linearization(XL), and their extensions are applicable, see [3, 14, 10, 11, 8, 7]. For instance, F4 was introduced in [10] and implemented in MAGMA. Available average complexity estimates are conjectural and heavily depend on the equations total degree. In general Boolean equations Gröbner basis algorithms are faster than brute force only for quadratic polynomials as it follows from [1, 21], though the method is likely efficient for some highly structural equations even if they are not quadratic.

One can write the equations from ciphers as a system of systems of low rank linear equations with multiple right hand sides(MRHS) [17]. That provides with a more general definition of sparse equations than above. The system is then solved with a guess-and-determine algorithm. Asymptotical complexity of the method is unknown. By experiments, MRHS approach is significantly faster for quadratic Boolean equations from the Advanced Encryption Standard(AES) in comparison with Gröbner basis type algorithm F4.

The equations from ciphers may be represented by CNF formulas [2, 6] and solved with MiniSat [9] or any other modern SAT-solver. The latter implement a DPLL-type searching algorithm [4, 5] and are rather efficient though their asymptotical complexity is unknown.

1.2. **Complexity Definitions.** The equation system (1) can be encoded by an $l\lceil \log_2 q\rceil$-CNF formula and solved with a SAT-solving algorithm for any polynomials f_i and for any sets X_i of variables. The currently known best upper bound on 3-SAT solving is at most 1.3212^n in [13], where n stands for the number of variables in the CNF. This provides a worst-case bound for (1) when $q = 2$ and $l = 3$.

Another approach is average time complexity; the result depends on the distribution of instances (1). If no particular information on the equations is known beforehand, uniform distribution(the variable sets and the polynomials are taken independently and uniformly at random) is the most fair probabilistic model to compute average complexity. This approach was studied in [18, 19, 20]. For instance, in Boolean case ($q = 2$) when $l = 3$ and $m = n$, the average complexity of computing all solutions is at most 1.029^n for n large enough [20].

In this article a different model is studied. The sets of variables belong to a fixed family $\mathcal{X} = \{X_1, \ldots, X_m\}$ while the polynomials $f_i(X_i)$ are taken independently and uniformly at random from the set of all polynomials of degree $\leq q-1$ in each of the variables in X_i. The average complexity is then a function of $\mathcal{X}$. We prove it is upper bounded by

$$q^{\frac{n}{5.7883}+O(\log n)} \tag{2}$$

for $l = 3$, $m = n$, and for arbitrary $\mathcal{X}$ and q. For $q = 2$ and sufficiently large n this bound is 1.1273^n.

1.3. **Gluing Algorithm and a New Combinatorial Problem.** Gluing Algorithm has been designed in [18] to solve sparse equation systems. Let V_k denote the set of solutions to the first k equations in relevant variables $\bigcup_{j=1}^{k} X_j$. The algorithm constructs V_{k+1}, given V_k and the next equation $f_{k+1} = 0$. Let l, q be fixed while n, m are allowed to grow. The complexity of the algorithm is $\sum_{k=1}^{m} |V_k|$ up to a multiplicative constant. The average complexity is then $\sum_{k=1}^{m} E(|V_k|)$, where by [18], $E(|V_k|) = q^{|\bigcup_{j=1}^{k} X_j|-k}$. So the average complexity of finding all solutions to (1) by Gluing Algorithm is proportional to

$$\sum_{k=1}^{m} q^{|\bigcup_{j=1}^{k} X_j|-k} \leq m\, q^{\max_k |\bigcup_{j=1}^{k} X_j|-k}.$$

By permuting the equations with a permutation π on m symbols, one possibly reduces the maximal of the differences $|\bigcup_{j=1}^{k} X_{\pi(j)}| - k$ and the algorithm's running time. Estimating

$$\max_{\mathcal{X}} \min_{\pi} \max_{k} |\bigcup_{j=1}^{k} X_{\pi(j)}| - k \tag{3}$$

is a new problem in hypergraphs called l-MaxMinMax problem.

In May 2012 I communicated the problem to Peter Horak. In 2013 he and Zsolt Tuza proved that for $n = m$ and $l = 3$, the value (3) is between $n/12.214$ and $\lceil n/4 \rceil + 2$, [12]. Their lower bound shows there exist sparse equations, where the above variation of Gluing algorithm can only achieve exponential average time complexity at its best. The upper bound, though interesting by its method, is not constructive and therefore does not seem to have implications in sparse equations complexity.

In present paper it is proved that for any $\mathcal{X} = \{X_1, \ldots, X_n\}$, where $|X_i| \leq 3$ and $|\bigcup_{i=1}^{n} X_i| \leq n$, there is a permutation π on n symbols such that

$$\max_{k} |\bigcup_{j=1}^{k} X_{\pi(j)}| - k \leq \frac{n}{5.7883} + 1 + 2\log_2 n \tag{4}$$

and π is constructed in polynomial time. This is a corollary to a more general statement, see Theorem 9.1, Corollary 1 and Theorem 9.2 below. (4) implies an asymptotically better upper bound on (3) than in [12] and the complexity estimate (2).

I am grateful to Peter Horak for a number of suggestions on improving the presentation of an earlier variant of this work. The competition with [12] has stimulated my research.

2. General MaxMinMax Problem and Example

Let (S, μ) be a measurable space and $\mathcal{X} = \{X_1, \ldots, X_m\}$ a family of measurable subsets of S, where $\mu(X_i) \leq l$. Let $f(k)$ be a positive valued increasing function. The differences

$\mu\left(\bigcup_{j=1}^{k} X_j\right) - f(k), (1 \le k \le m)$ mark the growth of the set measure in the subsequent covering by $X_1, \ldots, X_m$ in comparison with $f(k)$. Estimating

$$\max_{\mathcal{X}} \min_{\pi} \max_{k} \mu\left(\bigcup_{j=1}^{k} X_{\pi(j)}\right) - f(k)$$

over all above families $\mathcal{X}$ and all permutations π on m symbols is called l-MaxMinMax problem. For $|S| = n$, and $\mu(X) = |X|, X \subseteq S$, and $f(k) = k$, this is the previous section problem. As an example, let $S = \{1, 2, \ldots, 99\}$. The family consists of 3-subsets of S. The first are $\{3i+1, 3i+2, 3i+3\}$ for $i = 0, \ldots, 32$ and the rest 66 are randomly generated, 99 sets in all.

$$\begin{aligned}\mathcal{X} = \{&\{1,2,3\},\{4,5,6\},\{7,8,9\},\{10,11,12\},\{13,14,15\},\{16,17,18\},\{19,20,21\},\\ &\{22,23,24\},\{25,26,27\},\{28,29,30\},\{31,32,33\},\{34,35,36\},\{37,38,39\},\\ &\{40,41,42\},\{43,44,45\},\{46,47,48\},\{49,50,51\},\{52,53,54\},\{55,56,57\},\\ &\{58,59,60\},\{61,62,63\},\{64,65,66\},\{67,68,69\},\{70,71,72\},\{73,74,75\},\\ &\{76,77,78\},\{79,80,81\},\{82,83,84\},\{85,86,87\},\{88,89,90\},\{91,92,93\},\\ &\{94,95,96\},\{97,98,99\},\{6,21,81\},\{36,42,90\},\{69,73,91\},\{25,37,44\},\\ &\{37,61,66\},\{44,45,75\},\{81,87,99\},\{41,52,91\},\{48,75,85\},\{74,92,93\},\\ &\{7,64,75\},\{22,25,33\},\{60,85,95\},\{3,64,99\},\{27,41,58\},\{27,82,98\},\\ &\{51,58,77\},\{3,8,47\},\{17,45,99\},\{6,7,74\},\{8,10,86\},\{43,65,68\},\\ &\{15,54,74\},\{10,72,73\},\{51,55,82\},\{23,44,52\},\{23,80,96\},\{34,85,95\},\\ &\{29,69,70\},\{11,20,49\},\{32,65,95\},\{20,60,90\},\{39,60,76\},\{18,31,41\},\\ &\{14,63,89\},\{20,49,79\},\{8,28,43\},\{26,47,56\},\{22,37,91\},\{55,81,82\},\\ &\{45,63,70\},\{20,55,85\},\{32,36,60\},\{39,52,67\},\{54,55,86\},\{49,66,69\},\\ &\{24,51,68\},\{63,66,96\},\{35,57,88\},\{50,66,80\},\{2,14,99\},\{1,19,73\},\\ &\{2,58,79\},\{23,73,91\},\{1,65,73\},\{2,35,50\},\{4,33,60\},\{15,22,45\},\\ &\{25,36,62\},\{20,63,79\},\{8,14,69\},\{20,60,88\},\{12,25,43\},\{16,29,53\},\\ &\{34,35,76\},\{12,68,83\}\}.\end{aligned}$$

The permutation

$$\begin{aligned}&27,69,93,73,58,63,75,17,83,79,81,60,32,46,61,65,95,36,87,30,68,94,4,54,57,78,\\ &29,40,84,86,89,98,12,20,35,50,66,76,82,19,64,88,22,85,7,34,1,26,47,48,49,51,\\ &71,9,14,33,41,59,77,92,96,15,18,21,23,37,38,39,42,44,45,52,53,55,56,62,70,72,\\ &74,80,90,91,97,99,2,3,5,6,8,10,11,13,16,24,25,28,31,43,67\end{aligned}$$

produces the permutation on the sets

$$\{79,80,81\},\{20,49,79\},\{20,63,79\},\{55,81,82\},\ldots,\{18,31,41\},$$

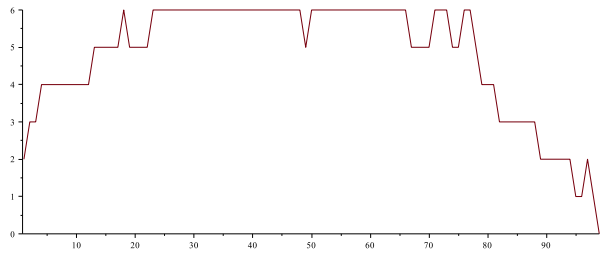

FIGURE 1. The difference profile

and the differences

$$2, 3^2, 4^9, 5^5, 6, 5^4, 6^{26}, 5, 6^{17}, 5^4, 6^3, 5^2, 6^2, 5, 4^3, 3^7, 2^6, 1^2, 2, 1, 0,$$

where by a^k a sequence of a repeated k times is denoted. That profile is shown in Fig. 1. The min max on $\mathcal{X}$ is at most 6. The average complexity of finding all solutions to related equations in 99 variables over a finite field F_q is proportional to

$$2q^2 + 9q^3 + 12q^4 + 17q^5 + 49q^6$$

operations by Gluing algorithm. Brute force takes q^{99} trials.

3. LEMMAS

Let $n, m, l \in \mathbb{N}$. Let

$$\mathcal{X} = \{X_1, \dots, X_m\} \tag{5}$$

be a family of sets such that $|X_i| \le l$ and $\left|\bigcup_{j=1}^m X_j\right| \le n$. We say X_i is a $(\le l)$-set. Also we write $\mathcal{X} = [X_1, \dots, X_m]$ to stress an ordering on $\mathcal{X}$. Let

$$\delta(X_1, \dots, X_m) = \max_k |\bigcup_{i=1}^k X_i| - k,$$

where $|\bigcup_{i=1}^k X_i| - k$ is called the k-th difference for $[X_1, \dots, X_m]$. Let

$$\Delta(\mathcal{X}) = \Delta(X_1, \dots, X_m) = \min_{i_1, \dots, i_m} \delta(X_{i_1}, \dots, X_{i_m})$$

over all permutations $i_1, \dots, i_m$ on $1, \dots, m$. Define

$$\mathbf{f}_l(n, m) = \max_{\mathcal{X}} \Delta(\mathcal{X})$$

over all families (5). We only study $l = 3$ in what follows, so $f_3(n, m) = \mathbf{f}(n, m)$. Also let $\mathbf{f}(n, m, s) = \max_{\mathcal{X}} \Delta(\mathcal{X})$ over all families (5) with precisely s (≤ 2)-sets.

Lemma 1. *The following statements hold*

(1) *Let* $U = \bigcup_{i=1}^{u} X_i$ *and* $\bar{X}_i = X_i \setminus U$, *then*

$$\begin{aligned}\delta(X_1,\dots,X_m) &= \max\{\delta(X_1,\dots,X_u), |U| - u + \delta(\bar{X}_{u+1},\dots,\bar{X}_m)\},\\ \Delta(X_1,\dots,X_m) &\le \max\{\Delta(X_1,\dots,X_u), |U| - u + \Delta(\bar{X}_{u+1},\dots,\bar{X}_m)\}.\end{aligned}$$

(2) *Let* $U \subseteq V$ *and* $\tilde{X}_i = X_i \setminus V$, *then*

$$\begin{aligned}\delta(X_1,\dots,X_m) &\le \max\{\delta(X_1,\dots,X_u), |V| - u + \delta(\tilde{X}_{u+1},\dots,\tilde{X}_m)\},\\ \Delta(X_1,\dots,X_m) &\le \max\{\Delta(X_1,\dots,X_u), |V| - u + \Delta(\tilde{X}_{u+1},\dots,\tilde{X}_m)\}.\end{aligned}$$

Proof. The first statement follows from $\left|\bigcup_{i=1}^{u+k} X_i\right| - (u+k) = |U| - u + \left|\bigcup_{i=1}^{k} \bar{X}_{u+i}\right| - k$. The second one from $\delta(\bar{X}_{u+1},\dots,\bar{X}_m) \le |V| - |U| + \delta(\tilde{X}_{u+1},\dots,\tilde{X}_m)$. □

Lemma 2. *Let* $m \ge n$ *and* $s \ge n-2$, *then* $\mathbf{f}(n,m,s) = 1$.

Proof. Let $\mathcal{X}$ be a family (5) with s (≤ 2)-sets. We prove by induction in n that $\Delta(\mathcal{X}) \le 1$ and $\Delta(\mathcal{X}) = 1$ for some such $\mathcal{X}$. That implies the statement.

The set covered by (≤ 2)-sets is split into connected components. Let C be one of them, and $X_1,\dots,X_u$ its cover by (≤ 2)-sets, then $\Delta(X_1,\dots,X_u) \le 1$. If $|\bigcup_{i=1}^{m} X_i| < u$, then $\Delta(\mathcal{X}) \le 1$. Let $|\bigcup_{i=1}^{m} X_i| \ge u$. We have $|C| \le u+1$. If $|C| \le u$, then by Lemma 1,

$$\Delta(\mathcal{X}) \le \max\{\Delta(X_1,\dots,X_u), |U| - u + \Delta(\bar{X}_{u+1},\dots,\bar{X}_m)\},$$

where $C \subseteq U \subseteq \bigcup_{i=1}^{m} X_i$, $|U| = u$, $\bar{X}_j = X_j \setminus U$. By induction, $\Delta(\bar{X}_{u+1},\dots,\bar{X}_m) \le 1$ so $\Delta(\mathcal{X}) \le 1$. We can assume $|C| = u+1$ for any such component. As $s \ge n-2$, there are at most 2 connected components. As $m \ge n$, it is easy to see $\Delta(\mathcal{X}) = 1$ in that case.

□

Lemma 3. $\mathbf{f}(n,m,s+1) \le \mathbf{f}(n,m,s)$.

Proof. It follows from $\Delta(X'_1,\dots,X'_m) \le \Delta(X_1,\dots,X_m)$, once $X'_i \subseteq X_i$. □

Lemma 4. *Let* a *be any non-negative integer number, then*

$$\mathbf{f}(n,m+a,s) \le \mathbf{f}(n,m,s) \le \mathbf{f}(n+a,m,s),$$

where the leftmost inequality is true for $m \ge n-2$.

Proof. The rightmost inequality is obvious. We'll prove the leftmost one. Let

$$\mathbf{f}(n,m+1,s) = \Delta(X_1,\dots,X_m,X_{m+1}).$$

As $m \ge s$, we assume X_{m+1} is a 3-set. Also we can assume there are no (≤ 1)-sets among $X_1,\dots,X_m$. Denote $U = \bigcup_{i=1}^{m} X_i$, so $|U| + |X_{m+1} \setminus U| \le n$. The leftmost inequality follows from

$$\begin{aligned}\mathbf{f}(n,m+1,s) &\le \max\{\Delta(X_1,\dots,X_m), |U| - m + |X_{m+1} \setminus U| - 1\}\\ &= \Delta(X_1,\dots,X_m) \le \mathbf{f}(n,m,s)\end{aligned}$$

by Lemma 1 and as $|U| - m + |X_{m+1} \setminus U| - 1 \le n - m - 1 \le 1 \le \Delta(X_1,\dots,X_m)$. □

Lemma 5. *Let a be any non-negative integer number, then*

$$\mathbf{f}(n,m,s) \le \mathbf{f}(n+a,m+a,s+a) \le \mathbf{f}(n+a,m+a,s).$$

Proof. It is enough to prove the statement for $a=1$. Let

$$\Delta(\mathcal{X}) = \mathbf{f}(n,m,s).$$

For some $z \notin \bigcup_{i=1}^{m} X_i$, let $X_{m+1} = \{z\}$. Then

$$\mathbf{f}(n,m,s) = \Delta(X_1,\dots,X_m,X_{m+1}) \le \mathbf{f}(n+1,m+1,s+1).$$

The rightmost inequality is true by Lemma 3.

□

Lemma 6. *Let $0 \le a \le n-1$. Then*

$$-a + \mathbf{f}(n,m,s) \le \mathbf{f}(n-a,m,s).$$

Proof. It is enough to prove the statement for $a=1$. Let

$$\Delta(\mathcal{X}) = \mathbf{f}(n,m,s)$$

for a family (5) with s (≤ 2)-sets. Denote $\bar{\mathcal{X}} = \{\bar{X}_1,\dots,\bar{X}_m\}$, where $\bar{X}_i = X_i \setminus \{x\}$ for some $x \in \bigcup_{i=1}^{m} X_i$. Then

$$\mathbf{f}(n,m,s) - 1 \le \Delta(\bar{\mathcal{X}}) \le \mathbf{f}(n-1,m,s).$$

The rightmost inequality is obvious. We'll prove the leftmost one. If $\Delta(\bar{\mathcal{X}}) < \mathbf{f}(n,m,s)-1$, then for a permutation $i_1,\dots,i_m$,

$$\delta(\bar{X}_{i_1},\dots,\bar{X}_{i_m}) = \Delta(\bar{\mathcal{X}}) < \mathbf{f}(n,m,s) - 1.$$

So

$$\delta(X_{i_1},\dots,X_{i_m}) \le \delta(\bar{X}_{i_1},\dots,\bar{X}_{i_m}) + 1 < \mathbf{f}(n,m,s),$$

a contradiction. That implies the statement. □

4. Definitions and Terminology

Let $\mathcal{Y} = \{X_1,\dots,X_u\}$ be a subfamily in $\mathcal{X} = \{X_1,\dots,X_m\}$. Any maximal(with respect to set inclusion) subfamily $\mathcal{C} = \{X_1,\dots,X_u,X_{u+1},\dots,X_{u+k}\}$ in $\mathcal{X}$ such that

$$|\bigcup_{i=1}^{u+r} X_i| - (u+r) \le |\bigcup_{i=1}^{u} X_i| - u, \quad r = 1,\dots,k,$$

is called a closure of $\mathcal{Y}$ in $\mathcal{X}$. We also say $\bigcup_{X\in\mathcal{C}} X$ is a closure of $\bigcup_{X\in\mathcal{Y}} X$ in $\mathcal{X}$ for $\mathcal{Y} \subseteq \mathcal{C}$. For instance,

$$\begin{aligned}
\mathcal{Z} &= \{\{1,2\},\{2,3\}\},\\
\mathcal{Y} &= \{\{1,2,3\},\{1,2\},\{2,3\}\},\\
\mathcal{X} &= \{\{1,2,3\},\{1,2\},\{2,3\},\{1,3,4\},\{1,5,6\},\{4,5,6\}\},
\end{aligned}$$

then $\mathcal{C} = \{\{1,2,3\},\{1,2\},\{2,3\},\{1,3,4\}\}$ is a closure of $\mathcal{Y}$ in $\mathcal{X}$. So $\{1,2,3,4\}$ is a closure of $\{1,2,3\}$ in $\mathcal{X}$ for $\mathcal{Y} \subseteq \mathcal{C}$. Similarly, $\mathcal{X}$ is a closure of $\mathcal{Z}$, so $\{1,2,3,4,5,6\}$ is a closure of $\{1,2,3\}$ in $\mathcal{X}$ for $\mathcal{Z} \subseteq \mathcal{X}$. The family $\mathcal{Y}$ is closed in $\mathcal{X}$ if

$$|\bigcup_{i=1}^{u} X_i \cup X| - (u+1) > |\bigcup_{i=1}^{u} Y_i| - u$$

for any $X \in \mathcal{X} \setminus \mathcal{Y}$. So $\bigcup_{X \in \mathcal{Y}} X$ is closed in $\mathcal{X}$ for $\mathcal{Y}$. Let α, β, γ be real numbers, where $\alpha, \gamma \leq \beta$. We say $\mathcal{Y} = \{X_1, \ldots, X_u\}$ is a $[\alpha,\beta,\gamma]$-family if there exists an ordering $X_1, \ldots, X_u$ on $\mathcal{Y}$ such that

$$|X_1| - 1 \leq \alpha, \quad |\bigcup_{i=1}^{k} X_i| - k \leq \beta,\ k = 1, \ldots, u, \quad |\bigcup_{i=1}^{u} X_i| - u \leq \gamma.$$

We also say the set $\bigcup_{i=1}^{u} X_i$ admits a $[\alpha,\beta,\gamma]$-covering by $\mathcal{Y}$. In the example above, $\mathcal{X}$ is a $[1,1,0]$-family. A $[\alpha,\beta,\gamma]$-subfamily $\mathcal{Y} = \{X_1, \ldots, X_u\} \subseteq \mathcal{X}$ is called a maximal $[\alpha,\beta,\gamma]$-subfamily in $\mathcal{X}$ if $|\bigcup_{i=1}^{u} X_i \cup X| - (u+1) > \gamma$ for any $X \in \mathcal{X} \setminus \mathcal{Y}$. Any $[\alpha,\beta,\gamma]$-subfamily may be extended to a maximal $[\alpha,\beta,\gamma]$-subfamily. Any maximal $[\alpha,\beta,\gamma]$-subfamily is closed.

Lemma 7. *Let $\mathcal{Y} = \{X_1, \ldots, X_u\}$ be a maximal $[\alpha,\beta,0]$-subfamily in $\mathcal{X}$. Then $|\bigcup_{i=1}^{u} X_i| = u$ or $\mathcal{Y}$ contains all (≤ 2)-sets and $\mathcal{Y}$ is a component in $\mathcal{X}$.*

Proof. By the definition of $[\alpha,\beta,0]$-subfamily, $|\bigcup_{i=1}^{u} X_i| \leq u$. If $|\bigcup_{i=1}^{u} X_i| = u$, the statement is true. Let $|\bigcup_{i=1}^{u} X_i| < u$. As $\mathcal{Y}$ is maximal, it contains all (≤ 2)-sets, otherwise any such set is added to $\mathcal{Y}$ to get a larger $[\alpha,\beta,0]$-subfamily. Similarly, for any 3-set $X \in \mathcal{X} \setminus \mathcal{Y}$ we have $X \bigcap \bigcup_{i=1}^{u} X_i = \emptyset$. So $\mathcal{Y}$ is a component in $\mathcal{X}$. □

The subfamily $\mathcal{Y} = \{X_1, \ldots, X_u\} \subseteq \mathcal{X}$ is called a connected component in $\mathcal{X}$ if it is a connected family and $(\bigcup_{i=1}^{u} X_i) \bigcap X = \emptyset$ for any $X \in \mathcal{X} \setminus \mathcal{Y}$.

Let (5) contains precisely s (≤ 2)-sets. We then say $\mathcal{X}$ has parameters n, m, s. In the example above, $\mathcal{X}$ has parameters $6, 6, 2$.

5. Constructing Closed Components from 2-Sets

Denote $L_x = 1 + \log_2 x$.

Theorem 5.1. *Let $\mathcal{X}$ have parameters n, m and $s > 0$. There exists a non-empty maximal $[1, L_m, 0]$-subfamily in $\mathcal{X}$ or*

1. *there are disjoint $E_i \subseteq \bigcup_{j=1}^{m} X_j$, $1 \leq i \leq v$ such that*
2. *E_i contains $T_i > 0$ 2-sets from $\mathcal{X}$ and $\sum_{i=1}^{v} T_i = s$,*
3. *E_i admits a $[1, L_{T_i}, 1]$-covering by $\mathcal{X}$, the number of sets in this covering is $|E_i| - 1$,*
4. *E_i is closed in $\mathcal{X}$.*

Proof. All (≤ 1)-sets in $\mathcal{X}$ is a $[1, L_m, 0]$-subfamily. One extends it to a maximal $[1, L_m, 0]$-subfamily and the theorem is true. So we assume there are no (≤ 1)-sets in $\mathcal{X}$ and prove the theorem by recursive construction.

Initial step. The set covered by 2-sets from $\mathcal{X}$ is split into r connected components $C_1, \dots, C_r$. Each C_i is covered by s_i 2-sets, where $s = s_1 + \dots + s_r$. The 2-sets in the covering of all C_i, such that $|C_i| \le s_i$, is a $[1, L_m, 0]$-subfamily in $\mathcal{X}$, it may be extended to a maximal one and the theorem is true. If there are no such components, then $|C_i| = s_i + 1$ and C_i admits a $[1, 1, 1]$-covering for all $i = 1, \dots, r$. The conditions (1)-(3) of the theorem are satisfied with $C_1, \dots, C_r$.

Recursive step. Let the conditions (1)-(3) of the theorem be satisfied with some $C_1, \dots, C_r$. Each C_i admits a $[1, g_i, 1]$-covering by subsets in $\mathcal{X}$, where $g_i \le 1 + \log_2 s_i$ and s_i is the number of 2-sets in the covering of C_i. Let $\bar{C}_i$ be a closure of C_i. Adding a new set introduces exactly one new element in $\bar{C}_i$. Otherwise, a maximal $[1, L_m, 0]$-subfamily is constructed. Otherwise, $\bar{C}_i$ has a $[1, g_i, 1]$-covering which is not a $[1, g_i, 0]$-covering.

$\bar{C}_1, \dots, \bar{C}_r$ are split into v connected components $E_1, \dots, E_v$. Let $\bar{C}_1, \dots, \bar{C}_t$, where $\bar{C}_i \nsubseteq \bar{C}_j$ and $t \ge 2$, compose one of the above components E and $g = \max_{i=1\dots t} g_i$. Let T be the number of 2-sets in the coverings of $\bar{C}_1, \dots, \bar{C}_t$, so $T \ge \sum_{i=1}^{t} s_i$ as $C_1, \dots, C_t$ are disjoint.

Firstly, assume there is at most one $\bar{C}_i$ with $[1, g, 1]$-covering and all other $\bar{C}_j$ have $[1, g-1, 1]$-coverings. Then E admits $[1, g, 1]$ covering as well. Really, let $\bar{C}_1, \dots, \bar{C}_r$ be ordered such that $\bar{C}_1$ has a $[1, g, 1]$-covering, and $\bar{C}_2, \dots, \bar{C}_r$ have $[1, g-1, 1]$-coverings, and that is a connected ordering. By Lemma 8 below, the concatenation of the coverings for $\bar{C}_1, \dots, \bar{C}_r$ after dropping repetitions, is a $[1, g, 1]$-covering for E or a $[1, g, 0]$-subfamily in $\mathcal{X}$ is constructed.

Secondly, let some $\bar{C}_i$ and $\bar{C}_j$ both have $[1, g, 1]$ coverings. Then $g \le 1 + \log_2 s_i$ and $g \le 1 + \log_2 s_j$. So $g + 1 \le 1 + \log_2(s_i + s_j) \le 1 + \log_2 T$. By Lemma 8, E admits $[1, g+1, 1]$ covering. Therefore, in both cases, E admits a $[1, L_T, 1]$ covering. If this is a $[1, L_T, 0]$-covering, one then constructs a maximal $[1, L_m, 0]$-subfamily in $\mathcal{X}$. Otherwise, the number of sets in the covering of E is exactly $|E| - 1$.

The conditions (1)-(3) of the theorem are satisfied with $E_1, \dots, E_v$. We apply the step recursively until the components E_i can not be further extended by taking closures in $\mathcal{X}$, we call them closed components.

Lemma 8. *Let $C_i, 1 \le i \le t$ be components with $[1, g_i, 1]$-coverings and their closures define a component $E = \bigcup_{i=1}^{t} \bar{C}_i$. Let $\bar{C}_1, \dots, \bar{C}_t$ be a connected ordering and $g = \max_{2 \le i \le t}\{g_1, 1 + g_i\}$. Then the concatenation of $\bar{C}_i$-coverings after dropping the repetitions is a $[1, g, 1]$-covering for E or some its truncation is a $[1, g, 0]$-subfamily.*

Proof. We prove by induction that for $1 \le a \le t$

(1) $\bigcup_{i=1}^{a} \bar{C}_i$ admits a $[1, \max_{2 \le i \le a}\{g_1, 1 + g_i\}, 1]$-covering,
(2) for any sets $Z_1, \dots, Z_v$, $v \ge 1$ in the covering of $\bigcup_{i=1}^{a} \bar{C}_i$ such that non of Z_i is a subset of $C_1, \dots, C_a$, holds $|\bigcup_{i=1}^{v} Z_v| > v$,

or one constructs a $[1, g, 0]$-subfamily. Let $a = 1$. The first claim is trivial. Each Z_i adds exactly one new element to the closure $\bar{C}_1$, otherwise its covering is a $[1, g, 0]$-subfamily. As $|Z_i| \ge 2$, we have $|\bigcup_{i=1}^{v} Z_i| > v$ and the second claim is true.

Let the claims be true for $a < t$, we prove them for $a = t$. Let the coverings of $\bigcup_{i=1}^{t-1} \bar{C}_i$ and $\bar{C}_t$ do not contain the same sets from $\mathcal{X}$. Then the first claim is obviously true. Let's prove the second claim. If $|(\bigcup_{i=1}^{t-1} \bar{C}_i) \bigcap \bar{C}_t| \geq 2$, then the concatenation of coverings for $\bigcup_{i=1}^{t-1} \bar{C}_i$ and $\bar{C}_t$ is a $[1, L_m, 0]$-subfamily. We can assume $|(\bigcup_{i=1}^{t-1} \bar{C}_i) \bigcap \bar{C}_t| = 1$. Let $Z_1, \ldots, Z_c$ be in the covering of $\bigcup_{i=1}^{t-1} \bar{C}_i$ and $Z_{c+1}, \ldots, Z_v$ in the covering of $\bar{C}_t$. Then $|\bigcup_{i=1}^{c} Z_i| > c$ by induction and $|\bigcup_{i=c+1}^{v} Z_i| > v - c$ as each of $Z_{c+1}, \ldots, Z_v$ introduces exactly one new element in $\bar{C}_t$. The sets $\bigcup_{i=1}^{c} Z_i$ and $\bigcup_{i=c+1}^{v} Z_i$ have at most one element in common, so $|\bigcup_{i=1}^{v} Z_i| > v$.

Let there be common sets in both coverings now. In particular, let some $Y_1, \ldots, Y_l$, $l \geq 1$ appear in the coverings of $\bigcup_{i=1}^{t-1} \bar{C}_i$ and $\bar{C}_t$. We will prove $|\bigcup_{i=1}^{l} Y_i| > l$. Firstly, let no sets Y_i be a subset of C_t. Then each Y_i adds exactly one new element to $\bar{C}_t$. So as $|Y_i| \geq 2$, we have $|\bigcup_{i=1}^{l} Y_i| > l$. Secondly, let non of $Y_1, \ldots, Y_c$ be a subset of C_t and let $Y_{c+1}, \ldots, Y_l$ be subsets of C_t. Therefore non of $Y_{c+1}, \ldots, Y_l$ is a subset of $C_1, \ldots, C_{t-1}$ and by induction $|\bigcup_{i=c+1}^{l} Y_i| > l - c$. Let A consist of elements introduced by $Y_1, \ldots, Y_c$ in $\bar{C}_t$. Then $|A| = c$, $A \cap \bigcup_{i=c+1}^{l} Y_i = \emptyset$ and so $|\bigcup_{i=1}^{l} Y_i| \geq |A| + |\bigcup_{i=c+1}^{l} Y_i| > l$.

Let $[X_1, \ldots, X_w]$ and $[Y_1, \ldots, Y_u]$ be constructed coverings for $\bigcup_{i=1}^{t-1} \bar{C}_i$ and $\bar{C}_t$ accordingly. One constructs a covering for $\bigcup_{i=1}^{t} \bar{C}_i$ as a concatenation $[X_1, \ldots, X_w, Y_1, \ldots, Y_u]$ after dropping those $Y_1, \ldots, Y_u$, which are already in $\{X_1, \ldots, X_w\}$. Let's estimate the difference with index $w + v$ in the concatenation. Let $Y_{i_1}, \ldots, Y_{i_l}$ be all dropped set such that $i_1, \ldots, i_l < v + l$. If $l = 0$ then

$$\left|\bigcup_{i=1}^{w} X_i \cup \bigcup_{i=1}^{v} Y_i\right| - (w + v) \leq 1 + \left|\bigcup_{i=1}^{v} \bar{Y}_i\right| - v \leq 1 + \left|\bigcup_{i=1}^{v} Y_i\right| - v,$$

where $\bar{Y}_i = Y_i \setminus \bigcup_{i=1}^{w} X_i$ and because $|\bigcup_{i=1}^{w} X_i| - w = 1$. Let $l \geq 1$, then

$$\begin{aligned} & \left|\bigcup_{i=1}^{w} X_i \cup \bigcup_{i \neq i_j}^{v+l} Y_i\right| - (w + v) = 1 + \left|\bigcup_{i \neq i_j}^{v+l} \bar{Y}_i\right| - v \qquad (6) \\ \leq \; & 1 + \left|\bigcup_{i=1}^{v+l} Y_i \setminus \bigcup_{j=1}^{l} Y_{i_j}\right| - v \\ = \; & 1 - \Big(\left|\bigcup_{j=1}^{l} Y_{i_j}\right| - l\Big) + \left|\bigcup_{i=1}^{v+l} Y_i\right| - (v + l) \leq \left|\bigcup_{i=1}^{v+l} Y_i\right| - (v + l) \end{aligned}$$

as $\left|\bigcup_{j=1}^{l} Y_{i_j}\right| > l$. That implies $[X_1, \ldots, X_w, Y_1, \ldots, Y_u]$ after dropping the repetitions is a $[1, g, 1]$ subfamily and the first claim is true. Let's prove the second claim. By (6), if $\left|\bigcup_{i=1}^{v+l} Y_i\right| - (v + l) = 1$, then $\left|\bigcup_{i=1}^{w} X_i \cup \bigcup_{i \neq i_j}^{v+l} Y_i\right| - (w + v) = 1$ or $[X_1, \ldots, X_w, Y_1, \ldots, Y_v]$ after dropping the repetitions is a $[1, g, 0]$ subfamily.

Let $Y_{i_1},\dots,Y_{i_k}$ be all sets which appear in both coverings. Let $Z_1,\dots,Z_c$ be in the covering of $\bigcup_{i=1}^{t-1}\bar{C}_i$, and $Z_{c+1},\dots,Z_v$ in the covering of $\bar{C}_t$ and non of them is a subset of $C_1,\dots,C_t$. We can assume $Z_{c+1},\dots,Z_v$ do not belong to the repetitions $\{Y_{i_1},\dots,Y_{i_k}\}$.

Let $Z_{c+1},\dots,Z_h$ appear in the covering of $\bar{C}_t$ before any of $Y_{i_1},\dots,Y_{i_k}$. As $Z_{h+1},\dots,Z_v$ appear in the covering of $\bar{C}_t$ and not in the covering of C_t, by (6) and the argument above, $Z_{h+1},\dots,Z_v$ introduce exactly $v-h$ new elements to $\bigcup_{i=1}^{t}\bar{C}_i$, or one gets a $[1,g,0]$-subfamily. We denote this set by A, where $|A|=v-h$. If $h=c$, then $|\bigcup_{i=1}^{h}Z_i|>h$ by induction. If $h>c$, then $Y_{i_1},\dots,Y_{i_k}$ are not in the covering of C_t but in the covering of $\bar{C}_t$ only. One can remove from the covering of $\bar{C}_t$ all sets starting from the first of Y_{i_j} and get some $\bar{C}'_t$. Then $\bar{C}'_t$ and $\bigcup_{i=1}^{t-1}\bar{C}_i$ are without common $\mathcal{X}$-members and $\left|\bigcup_{i=1}^{t-1}\bar{C}_i\bigcap\bar{C}'_t\right|\le 1$, otherwise one gets a $[1,g,0]$-subfamily. So by the above argument, $|\bigcup_{i=1}^{h}Z_i|>h$. In both cases, as $|A|=v-h$ and $A\cap\bigcup_{i=1}^{h}Z_i=\emptyset$, we get $|\bigcup_{i=1}^{v}Z_i|>v$. The lemma is proved. □

The theorem is proved. □

5.1. **Example.** Let $\mathcal{X}=\{\{1,2\},\{2,3\},\{5,6\},\{8,9\},\{1,3,4\},\{4,5,6\},\{1,6,7\},\{7,8,9\}\}$. The sequence of sets in that order has the differences: $1,1,2,3,3,2,2,1$. We now apply the recursive construction. The subfamilies

$$[\{1,2\},\{2,3\}],\quad [\{5,6\}],\quad [\{8,9\}],$$

cover components $C_1=\{1,2,3\}$, $C_2=\{5,6\}$, $C_3=\{8,9\}$. By taking closures,

$$[\{1,2\},\{2,3\},\{1,3,4\}],\quad [\{5,6\},\{4,5,6\}],\quad [\{8,9\},\{7,8,9\}]$$

cover $\bar{C}_1=\{1,2,3,4\},\bar{C}_2=\{4,5,6\},\bar{C}_3=\{7,8,9\}$. By concatenation,

$$[\{1,2\},\{2,3\},\{1,3,4\},\{5,6\},\{4,5,6\}],\quad [\{8,9\},\{7,8,9\}].$$

cover new components $D_1=\{1,2,3,4,5,6\},D_2=\{7,8,9\}$. By taking closures,

$$[\{1,2\},\{2,3\},\{1,3,4\},\{5,6\},\{4,5,6\},\{1,6,7\}],\quad [\{8,9\},\{7,8,9\}]$$

cover $\bar{D}_1=\{1,2,3,4,5,6,7\},\bar{D}_2=\{7,8,9\}$. By concatenation,

$$[\{1,2\},\{2,3\},\{1,3,4\},\{5,6\},\{4,5,6\},\{1,6,7\},\{8,9\},\{7,8,9\}]$$

covers $E=\{1,2,3,4,5,6,7,8,9\}$ which is a closed component in $\mathcal{X}$. The covering has differences: $1,1,1,2,1,1,2,1$ and is a $[1,2,1]$-subfamily.

6. Procedure ORDER

Let E be a closed component constructed by the procedure in Section 5. Let T be the number of 2-sets in the covering of E. If $|E|=T+1$, then E is called a small component. If $|E|\ge T+2$, then E is called a large component.

In this section a procedure ORDER is defined. Input: a family $\mathcal{X}=\{X_1,\dots,X_m\}$ with parameters n,m,s. Output: an ordered list of sets $[X_{i_1},\dots,X_{i_m}]$, that is $[X_{i_1},\dots,X_{i_m}]=$ ORDER$(X_1,\dots,X_m)$.

While $m>0$:

(1) if $s = 0$, let x appear in the maximal number of $\mathcal{X}$-members. Then recursively $[\bar{X}_{i_1}, \ldots, \bar{X}_{i_m}] = \mathrm{ORDER}(\bar{X}_1, \ldots, \bar{X}_m)$, where $\bar{X}_i = X_i \backslash \{x\}$. Return $[X_{i_1}, \ldots, X_{i_m}]$.
(2) if $s \geq 1$, apply the procedure in Theorem 5.1 and get a non-empty maximal $[1, L_m, 0]$-subfamily $\{X_1, \ldots, X_u\}$ or a closed component:
   (a) Assume a $[1, L_m, 0]$-covering $[X_1, \ldots, X_u]$. Put $\bar{X}_i = X_i \backslash \bigcup_{i=1}^{u} X_i$ and compute recursively $[\bar{X}_{i_{u+1}}, \ldots, \bar{X}_{i_m}] = \mathrm{ORDER}(\bar{X}_{u+1}, \ldots, \bar{X}_m)$. Return
   $$[X_1, \ldots, X_u, X_{i_{u+1}}, \ldots, X_{i_m}].$$
   (b) Let E be a closed component with a $[1, L_m, 1]$-covering $[X_1, \ldots, X_v]$:
      (i) E is a small closed component. For $z \notin \bigcup_{i=1}^{m} X_i$, put
      $$\bar{X}_j = \begin{cases} X_j, & \text{if } X_j \cap E = \emptyset; \\ (X_j \setminus E) \cup \{z\}, & \text{if } X_j \cap E \neq \emptyset. \end{cases}$$
      Then $[\bar{X}_{i_{v+1}}, \ldots, \bar{X}_{i_m}] = \mathrm{ORDER}(\bar{X}_{v+1}, \ldots, \bar{X}_m)$. Let $k, v+1 \leq k \leq m$ be the smallest index such that $z \in \bar{X}_{i_k}$, and $[X_{j_1}, \ldots, X_{j_v}]$ is a covering of E such that $X_{i_k} \cup X_{j_1} \neq \emptyset$ or if $\{X_1, \ldots, X_v\}$ is a connected component in $\mathcal{X}$ then $k = m$. Return
      $$[X_{i_{v+1}}, \ldots, X_{i_k}, X_{j_1}, \ldots, X_{j_v}, X_{i_{k+1}}, \ldots, X_{i_m}].$$
      (ii) E is a large closed component, put $t = \lceil \frac{3(m-s)}{n-s} \rceil$ and $X_{v+1}, \ldots, X_{v+r}$ be all X_j such that $|X_j \bigcap E| = 1$:
         (A) $r + 2 \leq t$. Then $[X_{i_{v+r+1}}, \ldots, X_{i_m}] = \mathrm{ORDER}(X_{v+r+1}, \ldots, X_m)$. Return $[X_{i_{v+r+1}}, \ldots, X_{i_m}, X_1, \ldots, X_{v+r}]$.
         (B) $r + 1 \geq t$. Then $[\bar{X}_{i_{v+1}}, \ldots, \bar{X}_{i_m}] = \mathrm{ORDER}(\bar{X}_{v+1}, \ldots, \bar{X}_m)$, where $\bar{X}_i = X_i \setminus E$. Return $[X_1, \ldots, X_v, X_{i_{v+1}}, \ldots, X_{i_m}]$.

Theorem 6.1. *The complexity of ORDER is $O(nm^2)$ of pairwise unions and intersections with sets of size at most n.*

Proof. The procedure takes at most n recursive steps. At each step one extracts one element from the sets where it occurs, or constructs a $[1, L_m, 0]$-subfamily or a closed component and extracts those elements from the sets where they occur. The construction costs $O(m^2)$ set operations. That implies the theorem. □

7. Reductions

Theorem 7.1. *Let $0 \leq s < n \leq m$. Then* $\mathbf{f}(n, m, s)$

$$\leq \max\{L_m, \mathbf{f}(n - T, m - T, s - T)\} \tag{7}$$

for some $0 < T \leq s$, or

$$\leq \max\{L_m, 1 + \mathbf{f}(n - 1, m, s + t)\}, \tag{8}$$

or

$$\leq \max\{L_m, \mathbf{f}(n - 2, m - t + 1, s)\}, \tag{9}$$

where $t = \lceil \frac{3(m-s)}{n-s} \rceil$. *If* $n - s \geq 3$, *then* $n - 2 \leq m - t + 1$.

Proof. Let $\mathcal{X} = \{X_1, \ldots, X_m\}$ be a family with parameters n, m, s. We apply ORDER to $\mathcal{X}$ and see $\Delta(\mathcal{X})$ is at most (7), (8) or (9). That will imply the theorem.
1. Let $s = 0$, that is stage (1). Then x appears in $u \geq t = \lceil \frac{3m}{n} \rceil$ of $X_1, \ldots, X_m$. So $\Delta(\mathcal{X}) \leq 1 + \Delta(\bar{X}_1, \ldots, \bar{X}_m) \leq 1 + \mathbf{f}(n-1, m, u) \leq 1 + \mathbf{f}(n-1, m, t)$, by Lemma 3. So $\Delta(\mathcal{X})$ is at most (8).

2. Let $s \geq 1$ and stage (a). We put $|\bigcup_{i=1}^{u} X_i| = b$ and let the subfamily contain T (≤ 2)-sets, where $0 < T \leq s$. By Lemma 1, $\Delta(\mathcal{X}) \leq \max\{L_m, b - u + \Delta(\bar{X}_{u+1}, \ldots, \bar{X}_m)\}$. So by Lemmas 5 and 6,

$$\begin{aligned} \Delta(\mathcal{X}) &\leq \max\{L_m, b - u + \mathbf{f}(n-b, m-u, s-T)\} \\ &\leq \max\{L_m, \mathbf{f}(n-u, m-u, s-T)\} \leq \max\{L_m, \mathbf{f}(n-T, m-T, s-T)\}, \end{aligned}$$

that is at most (7). Let stage (b) and $E, e = |E|$ be a closed component whose covering $X_1, \ldots, X_v$, where $v = e - 1$, contains T 2-sets. As E is closed, we have $|X_j \setminus E| = 1$ or 0 for $j = v+1, \ldots, m$.
3. Let stage (i). As E is a small closed component, then $T = v$ and $[X_1, \ldots, X_v]$ is its $[1,1,1]$-covering. If $\{X_1, \ldots, X_v\}$ is not a connected component in $\mathcal{X}$, then

$$\begin{aligned} \Delta(\mathcal{X}) &= \Delta(X_{i_{v+1}}, \ldots, X_{i_k}, X_{j_1}, \ldots, X_{j_v}, X_{i_{k+1}}, \ldots, X_{i_m}) \\ &\leq \Delta(\bar{X}_{i_{v+1}}, \ldots, \bar{X}_{i_m}) \leq \mathbf{f}(n-T, m-T, s-T). \end{aligned}$$

If $\{X_1, \ldots, X_v\}$ is a connected component in $\mathcal{X}$, then

$$\begin{aligned} \Delta(\mathcal{X}) &= \Delta(X_{i_{v+1}}, \ldots, X_{i_m}, X_1, \ldots, X_v) \\ &= \max\{\Delta(X_{i_{v+1}}, \ldots, X_{i_m}), (n-T-1) - (m-T) + \Delta(X_1, \ldots, X_v)\} \\ &\leq \mathbf{f}(n-T-1, m-T, s-T) \leq \mathbf{f}(n-T, m-T, s-T). \end{aligned}$$

In any case, $\Delta(\mathcal{X})$ is at most (7). Let E be a large closed component.

4. Let stage (A). If $n - s \leq 2$, then $\Delta(\mathcal{X}) \leq \mathbf{f}(n, m, s) = 1$ by Lemma 2 and (9) is true. Let $n - s \geq 3$, then $\frac{3(m-s)}{n-s} \leq m - n + 3$ so $n - m + r - 1 \leq n - m + t - 3 \leq 0$. We have $\Delta(\mathcal{X}) = \Delta(X_{i_{v+r+1}}, \ldots, X_{i_m}, X_1, \ldots, X_{v+r})$

$$\begin{aligned} &\leq \max\{\Delta(X_{i_{v+r+1}}, \ldots, X_{i_m}), |V| - (m - e - r + 1) + \Delta(X_1, \ldots, X_v, \bar{X}_{v+1}, \ldots, \bar{X}_{v+r})\} \\ &\leq \max\{\Delta(X_{i_{v+r+1}}, \ldots, X_{i_m}), (n-e) - (m - e - r + 1) + \Delta(X_1, \ldots, X_v)\} \end{aligned}$$

by Lemma 1, where $\bigcup_{i=v+r+1}^{m} X_i \subseteq V = \bigcup_{i=1}^{m} X_i \setminus E$ and $\bar{X}_j = X_j \setminus V$, and as $|\bar{X}_j| \leq 1, v+1 \leq j \leq v+r$. So

$$\begin{aligned} \Delta(\mathcal{X}) &\leq \max\{\mathbf{f}(n-e, m-e-r+1, s-T), n-m+r-1+L_m\} \\ &\leq \max\{\mathbf{f}(n-T-2, m-T-1-r, s-T), L_m\} \\ &\leq \max\{\mathbf{f}(n-2, m-1-r, s), L_m\} \leq \max\{\mathbf{f}(n-2, m-t+1, s), L_m\} \end{aligned}$$

by Lemmas 5, 4 as $e \geq T+2$, and $s \leq m-t+1 \leq m-r-1$, and $n-2 \leq m-t+1$. So $\Delta(\mathcal{X})$ is at most (9).

5. Let stage (B). By Lemmas 1 and 5,

$$\begin{aligned}\Delta(\mathcal{X}) &\leq \max\{\Delta(X_1,\ldots,X_v\}), e-(e-1)+\Delta(\bar{X}_{v+1},\ldots,\bar{X}_m)\\ &\leq \max\{L_m, 1+\mathbf{f}(n-e,m-e+1,s-T+r)\}\\ &\leq \max\{L_m, 1+\mathbf{f}(n-2,m-1,s+r)\} \leq \max\{L_m, 1+\mathbf{f}(n-1,m,s+r+1)\}\\ &\leq \max\{L_m, 1+\mathbf{f}(n-1,m,s+t)\},\end{aligned}$$

because $e-2 \geq T$. So $\Delta(\mathcal{X})$ is at most (8). The theorem is proved. □

8. Auxiliary Functions

Let

$$\alpha_i(s) = \frac{s+i-3}{(s+i-2)(s+i)},$$

for $i+s \geq 3$, and

$$\gamma_k(s) = \alpha_k(s)\prod_{i=0}^{k-1} 1+\alpha_i(s)$$

for $s \geq 3, k \geq 0$.

Lemma 9. *Let $s \geq 3$. Then*

1. $1-(1+s)\gamma_k(s) > 0$ *for all* $s \geq 3$ *and* $k \geq 0$.
2. $\gamma_k(s)$ *is increasing in* k *and tends to a real value* $\gamma(s)$,
3. $\gamma(s+1) = \frac{\gamma(s)}{1+\alpha_0(s)}$ *and so* $\gamma(s)$ *is decreasing in* s,
4. $\gamma_k(s) < \gamma(s) \leq \frac{\gamma_k(s)}{\alpha_k(s)(1+s+k)}$,
5. $\gamma(3) = \gamma(4) = \frac{1}{5.78838..}$,
6. *Let* $A(s)/\alpha_0(s) \to 1$ *as* $s \to \infty$ *and* $A(s+1)(1+\alpha_0(s)) \geq A(s)$, *then* $\gamma(s) \geq A(s)$,
7. $\gamma(s) \geq A_0(s) \geq \frac{s-2}{s(s-1)}$, *where* $A_0(s) = \frac{2(s-2)(s^2-s-3)}{2s^4-4s^3-5s^2+9s+1}$.

Proof. To prove (1), it is easy to check $1-(1+s)\gamma_0(s) > 0$ for any $s \geq 3$ and

$$1-(1+s)\gamma_{k+1}(s) = 1-(1+s)\gamma_k(s+1)(1+\alpha_0(s)) > 1-(2+s)\gamma_k(s+1).$$

Therefore the inequality for any $k \geq 0$ follows by induction. To prove (2), as $s \geq 3$, it is easy to see $\frac{\gamma_{k+1}(s)}{\gamma_k(s)} = \frac{\alpha_{k+1}(s)(1+\alpha_k(s))}{\alpha_k(s)} > 1$ and $\gamma_k(s) < \frac{1}{s+1}$, so $\gamma_k(s)$ has a limit $\gamma(s)$. Statement (3) comes from $\gamma_k(s+1) = \frac{\gamma_{k+1}(s)}{1+\alpha_0(s)}$ by taking $\lim_{k\to\infty}$. To prove (4),

$$\gamma_k(s) < \gamma_{k+j}(s) = \frac{\gamma_k(s)\gamma_j(k+s)}{\alpha_k(s)} < \frac{\gamma_k(s)}{\alpha_k(s)(1+s+k)}$$

where the rightmost inequality is true by the first statement. By taking $\lim_{j\to\infty}$, the statement (4) is true.

One explicitly computes $\gamma_k(3) = \gamma_k(4) = \frac{1}{5.78838..}$ for a large k. Statement (4) guarantees the accuracy $\gamma(s) = \gamma_k(s) + O(\frac{1}{(k+s)^2})$. So (5) is correct.

To prove (6), one writes

$$\begin{aligned}\gamma(3) &= \lim_{k\to\infty} \alpha_k(3) \prod_{i=0}^{k-1} 1+\alpha_i(3) = \lim_{s\to\infty} \alpha_0(s) \prod_{i=3}^{s-1} 1+\alpha_0(i) \\ &= \lim_{s\to\infty} A(s) \prod_{i=3}^{s-1} 1+\alpha_0(i).\end{aligned}$$

By the lemma condition, the sequence of $A(s)\prod_{i=3}^{s-1} 1+\alpha_0(i)$ is increasing in s, so

$$\gamma(3) \geq A(s) \prod_{i=3}^{s-1} 1+\alpha_0(i)$$

for any $s \geq 3$. Then

$$\gamma(s) = \frac{\gamma(3)}{\prod_{i=3}^{s-1} 1+\alpha_0(i)} \geq A(s).$$

$A(s) = A_0(s)$ satisfies the conditions in (6), that implies (7). □

For $k \geq 1$, denote

$$\beta_k(s) = 1 + \sum_{j=0}^{k-1} \quad \frac{\gamma(s)}{\prod_{i=0}^{j} 1+\alpha_i(s)}.$$

Lemma 10. (1) $\gamma(s+1) + \beta_k(s+1) = \beta_{k+1}(s)$.
(2) *Let $k \geq 3$, then $\beta_k(3) < \log_2 k$.*

Proof. The first statement follows from $\alpha_i(s+1) = \alpha_{i+1}(s)$ and $\gamma(s+1) = \frac{\gamma(s)}{1+\alpha_0(s)}$. As $\gamma_0(3) = 0$,

$$\prod_{i=0}^{j} 1+\alpha_i(3) = \gamma_j(3) + \ldots + \gamma_1(3) + \gamma_0(3) + 1 \geq \gamma_1(3)j + 1.$$

Therefore, by using $\sum_{j=2}^{n} \frac{1}{aj+b} \leq \int_2^n \frac{dx}{ax+b}$ for positive a, b, which comes from summation formulae in [15], we get

$$\beta_k(3) \quad \leq \quad 1 + \gamma(3) \sum_{j=0}^{k-1} \frac{1}{\gamma_1(3)j+1} < 1 + \gamma(3) \frac{\ln(\gamma_1(3)k+1)}{\gamma_1(3)} < \log_2 k.$$

As $\gamma(3) = \frac{1}{5.78838..}$, $\gamma_1(3) = \frac{1}{8}$, the rightmost inequality holds for $k = 3$. The derivative on the right hand side is larger, so the inequality is true for any $k \geq 3$.

□

Let $\lambda(n, m, s) = \gamma(s)\, n - [1 - (s+1)\gamma(s)]\,(m - n)$ for any $n \leq m$ and $s \geq 3$.

Lemma 11. *Let $n \leq m$, and $t = \lceil \frac{3m}{n} \rceil$ and $t_1 \geq t$, then*

(1) $\lambda(n,m,t) > 0$,
(2) $\lambda(n,m,s) - \lambda(n,m,s+1) = \gamma(s+1)\,\frac{sn-3m}{s(s-2)}$,
(3) $\lambda(n,m,t) \geq \lambda(n,m,s)$ *for* $s \geq 3$.
(4) $1 + \lambda(n-t-1, m-t, t_1) \leq \lambda(n,m,t) + \gamma(t+1) \ldots + \gamma(t_1)$,
(5) $\lambda(n-2, m-t+1, t) \leq \lambda(n,m,t)$,
(6) *if in addition* $t_1 > t$, *then*

$$1 + \lambda(n-t-3, m-t-t_1+1, t_1-1) \leq \lambda(n,m,t) + \gamma(t+1) \ldots + \gamma(t_1-1),$$

(7) *if in addition* $(t-1)(n+1) < 3m$, *then*

$$\gamma(t) + \lambda(n-4, m-2t+2, t-1) \leq \lambda(n,m,t).$$

Proof. As $\gamma(t) > \gamma_0(t) = \frac{t-3}{(t-2)t}$ and $3m \leq tn$,

$$\frac{1-t\gamma(t)}{1-(1+t)\gamma(t)} > \frac{t}{3} \geq \frac{m}{n}.$$

That implies the first statement. To prove (2), we have $\lambda(n,m,s) - \lambda(n,m,s+1) =$

$$\begin{aligned} & (\gamma(s) - \gamma(s+1))n + [(s+1)\gamma(s) - (s+2)\gamma(s+1)]\,(m-n) \\ = \;& \gamma(s+1)\,\frac{sn-3m}{s(s-2)} \end{aligned}$$

as $\gamma(s) = \gamma(s+1)(1 + \frac{s-3}{(s-2)s})$ and $(s+1)\gamma(s) - (s+2)\gamma(s+1) = \frac{-3\,\gamma(s+1)}{(s-2)s}$. To prove (3), we have $\lambda(n,m,s+1) \leq \lambda(n,m,s+1)$ for $s \geq t$ as $sn - 3m \geq 0$. On the other hand, $\lambda(n,m,s-1) \leq \lambda(n,m,s)$ for $s \leq t$ as $(s-1)n - 3m < 0$. To prove (4), we have

$$\begin{aligned} & 1 + \lambda(n-t-1, m-t, t_1) \\ = \;& 1 + \gamma(t_1)(n-t-1) - (1-(t_1+1)\gamma(t_1))(m-n+1) \\ = \;& \lambda(n,m,t_1) + 1 - (t+1)\gamma(t_1) - (1-(t_1+1)\gamma(t_1)) \\ = \;& \lambda(n,m,t_1) + (t_1-t)\gamma(t_1) \leq \lambda(n,m,t) + \gamma(t+1) + \ldots + \gamma(t_1) \end{aligned}$$

as $\lambda(n,m,t_1) \leq \lambda(n,m,t)$ by (2), and $\gamma(s)$ is decreasing by Lemma 9. To prove (5), we have $\lambda(n-2, m-t+1, t) =$

$$\begin{aligned} & \gamma(t)(n-2) - (1-(t+1)\gamma(t))(m-n-t+3) \\ = \;& \lambda(n,m,t) + (t-3) - (2+(t+1)(t-3))\gamma(t)) \leq \lambda(n,m,t) \end{aligned}$$

as $\gamma(t) \geq \frac{t-2}{t(t-1)} \geq \frac{t-3}{2+(t+1)(t-3)}$ for $t \geq 3$. To prove (6), we have

$$\begin{aligned} & 1 + \lambda(n-t-3, m-t-t_1+1, t_1-1) \\ = \;& 1 + \gamma(t_1-1)(n-t-3) - (1-t_1\gamma(t_1-1))(m-n-t_1+4) \\ = \;& \lambda(n,m,t_1-1) + 1 + (t_1-4) - (t+3+t_1(t_1-4))\gamma(t_1-1) \\ = \;& \lambda(n,m,t_1-1) + (t_1-3) - (t_1-2)(t_1-1)\gamma(t_1-1) + (t_1-t-1)\gamma(t_1-1) \\ \leq \;& \lambda(n,m,t) + \gamma(t+1) + \ldots + \gamma(t_1-1) \end{aligned}$$

as $\lambda(n,m,t_1-1) \le \lambda(n,m,t)$ by (2), and $\gamma(t_1-1) \ge \frac{t_1-3}{(t_1-2)(t_1-1)}$ by Lemma 9. To prove (7), we have $\gamma(t)+\lambda(n-4,m-2t+2,t-1) =$

$$\begin{aligned} & \gamma(t)+\gamma(t-1)(n-4)-(1-t\gamma(t-1))(m-n-2(t-3)) \\ = \;& \lambda(n,m,t-1)+\gamma(t)-4\gamma(t-1)+(1-t\gamma(t-1))2(t-3) \\ \le \;& \lambda(n,m,t)+\gamma(t)-\frac{t\gamma(t)}{(t-1)(t-3)}-4\gamma(t-1)+(1-t\gamma(t-1))2(t-3) \end{aligned}$$

as $\lambda(n,m,t-1)-\lambda(n,m,t) = \gamma(t)\frac{(t-1)n-3m}{(t-1)(t-3)} \le -\frac{t\gamma(t)}{(t-1)(t-3)}$ by (2). We have $\gamma(t) = \frac{\gamma(t-1)}{1+\alpha_0(t-1)}$. So

$$\begin{aligned} 0 \;\ge\; & \gamma(t)-\frac{t\gamma(t)}{(t-1)(t-3)}-4\gamma(t-1)+(1-t\gamma(t-1))2(t-3) \\ = \;& \gamma(t-1)\left[\frac{(t-3)(t-1)}{(t-3)(t-1)+t-4}-\frac{t}{(t-3)(t-1)+t-4}-4-2t(t-3)\right]+2(t-3) \\ = \;& \gamma(t-1)\left[-\frac{2t^4-12t^3+19t^2-x-7}{t^2-3t-1}\right]+2(t-3) \end{aligned}$$

if and only if $\gamma(t-1) \ge A_0(t-1)$. The latter is true by Lemma 9.

□

9. Main Results

Lemma 12. *Let $0 \le s < n \le m$ and $t = \lceil \frac{3(m-s)}{n-s} \rceil$.*

1. *If $n-s \le t$, then $\mathbf{f}(n,m,s) \le L_m$.*
2. *If $n-s \le 2t-2$, then $\mathbf{f}(n,m,s) \le 1+L_m$.*

Proof. Denote $n_1 = n-s, m_1 = m-s$. By Theorem 7.1, one of (7),(8),(9) is true. Let's prove the first statement. If (7), it is true by induction in n. Let (8), then by Lemma 2, $\mathbf{f}(n-1,m,s+t)=1$, and the statement is true. Let (9), we put $t_1 = \lceil \frac{3(m_1-t+1)}{n_1-2} \rceil$ and by the definition of t, t_1 as $n_1 \le t$,

$$(n_1-2)(n_1-3) < (t-1)(n_1-3) < 3m_1-3(t-1) \le t_1(n_1-2).$$

So $n_1-2 \le t_1$ and the first statement is true by induction. Let's prove the second statement. If (7), it is true by induction. Let (8),

$$\mathbf{f}(n,m,s) \le \max\{L_m, 1+\mathbf{f}(n-1,m,s+t)\} \le 1+L_m$$

as $t_1 = \lceil \frac{3(m_1-t)}{n_1-t-1} \rceil \ge t > n-s-t-1$ and so $\mathbf{f}(n-1,m,s+t) \le L_m$ by the first statement. Let (9), then $t_2 = \lceil \frac{3(m_1-t+1)}{n_1-2} \rceil \le t$. If $t_2 = t$ or $t-1$, then $n_1-2 \le 2t_2-2$ and the statement is true by induction. Let $t_2 \le t-2$, so $\frac{3(m_1-t+1)}{n_1-2} \le t-2$. As $(t-1)n_1 < 3m_1$, we have $n_1 \le t$ and by the first statement $\mathbf{f}(n,m,s) \le L_m$. The lemma is proved.

□

Theorem 9.1. *Let $n_1 = n - s \geq 3, m_1 = m - s \geq n_1$ and $t = \lceil \frac{3m_1}{n_1} \rceil$. Then $\mathbf{f}(n, m, s)$ is*

$$\leq \max\{\lambda(n_1, m_1, t) + \beta_{n_1}(t), \lambda(n_1 - 2, m_1 - t + 1, t - 1) + \beta_{n_1-2}(t-1)\} + L_m \tag{10}$$

in case $(t-1)n_1 < 3m_1 \leq (t-1)(n_1+1)$, or

$$\leq \lambda(n_1, m_1, t) + \beta_{n_1}(t) + L_m. \tag{11}$$

in case $(t-1)(n_1+1) < 3m_1 \leq tn_1$.

Proof. If $n_1 \leq t$, the theorem is true by Lemma 12. Assume $n_1 > t$ and prove the theorem by induction in n. By Theorem 7.1, $\mathbf{f}(n, m, s)$ is at most (7),(8), or (9).

If (7), then (10) or (11) is true by induction. Let (8), then $\mathbf{f}(n, m, s) \leq \max\{L_m, 1 + \mathbf{f}(n-1, m, s+t)\}$. We put $t_1 = \lceil \frac{3(m_1-t)}{n_1-t-1} \rceil$ and see $t_1 \geq t$. There are two cases to consider by induction. Firstly, $1 + \mathbf{f}(n-1, m, s+t) \leq 1 + \lambda(n_1 - t - 1, m_1 - t, t_1) + \beta_{n_1-t-1}(t_1)$

$$\begin{aligned} &\leq \lambda(n_1, m_1, t_1) + \gamma(t+1) \ldots + \gamma(t_1) + \beta_{n_1-t-1}(t_1) \\ &= \lambda(n_1, m_1, t) + \beta_{n_1+t_1-2t-1}(t). \end{aligned}$$

Therefore if $t_1 \leq 2t+1$, then (10) or (11) is true. Let $t_1 > 2t+1$. As $t_1 - 1 < \frac{3(m_1-t)}{n_1-t-1}$, and $\frac{3m_1}{n_1} \leq t$, we have

$$2t(n_1 - t - 1) < (t_1 - 1)(n_1 - t - 1) < 3m_1 - 3t \leq tn_1 - 3t$$

so $n_1 \leq 2t - 2$ and $\mathbf{f}(n, m, s) \leq 1 + L_m$ by Lemma 12. Therefore, (10) or (11) is true as $\beta_{n_1}(t) \geq 1$. Secondly, we can assume

$$(t_1 - 1)(n_1 - t - 1) < 3(m_1 - t) \leq (t_1 - 1)(n_1 - t),$$
$$\mathbf{f}(n-1, m, s+t) \leq \lambda(n_1 - t - 3, m_1 - t - t_1 + 1, t_1 - 1) + \beta_{n_1-t-3}(t_1 - 1)$$

otherwise the statement is true by the previous case. If $t_1 = t$, as $(t-1)n_1 < 3m_1$, we have $t = n_1 = m_1 = 3$ which contradicts with $n_1 > t$. Let $t_1 > t$. Then $1 + \mathbf{f}(n-1, m, s+t)$

$$\begin{aligned} &\leq 1 + \lambda(n_1 - t - 3, m_1 - t - t_1 + 1, t_1 - 1) + \beta_{n_1-t-3}(t_1 - 1) \\ &\leq \lambda(n_1, m_1, t) + \gamma(t+1) \ldots + \gamma(t_1 - 1) + \beta_{n_1-t-3}(t_1 - 1) \\ &\leq \lambda(n_1, m_1, t) + \beta_{n_1+t_1-2t-4}(t), \end{aligned}$$

where the second inequality is true by Lemma 11. If $t_1 \leq 2t + 4$ then (10) is true. If $t_1 > 2t + 4$, then (10) is true by the argument above.

Let (9), that is $\mathbf{f}(n, m, s) \leq \max\{L_m, \mathbf{f}(n-2, m-t+1, s)$. We put $t_2 = \lceil \frac{3(m_1-t+1)}{n_1-2} \rceil$. As $n_1 > t$, we have $t_2 = t - 1$ or t. Let $t_2 = t$. There are two cases to consider. Firstly,

$$\mathbf{f}(n-2, m-t+1, s) \leq \lambda(n_1 - 2, m_1 - t + 1, t) + \beta_{n_1-2}(t) \leq \lambda(n_1, m_1, t) + \beta_{n_1}(t),$$

by Lemma 11. So (10) or (11) is true. Secondly, as $t_2 = t$, we have $(t-1)n + t \leq 3m$ and so $\mathbf{f}(n-2, m-t+1, s) \leq \lambda(n_1 - 4, m_1 - 2t + 2, t - 1) + \beta_{n_1-4}(t-1)$

$$\begin{aligned} &= \lambda(n_1 - 4, m_1 - 2t + 2, t - 1) + \gamma(t) + \beta_{n_1-5}(t) \\ &\leq \lambda(n_1, m_1, t) + \beta_{n_1}(t), \end{aligned}$$

by Lemma 11 and (10) or (11) is true. Let $t_2 = t-1$, so $3m_1 \le (t-1)n_1 + (t-1)$. If $3(m_1 - t + 1) \le (t-2)(n_1-2) + (t-2)$, then as $(t-1)n_1 < 3m_1$ we have $n_1 \le 2t-2$ and the statement is true by Lemma 12. Let $3(m_1 - t + 1) \ge (t-2)(n_1-2) + (t-1)$. By induction, $\mathbf{f}(n-2, m-t+1, s) \le \lambda(n_1-2, m_1-t+1, t-1) + \beta_{n_1-2}(t-1)$. So (10) is true. The theorem is proved.

□

Corollary 1. $\mathbf{f}(n,n) \le \gamma(3)n + 1 + 2\log_2 n$.

Theorem 9.2. *There is a procedure ORDER* with the following property. Let $\{X_1, \dots, X_m\}$ be a family with parameters n, m, s and $[X_{i_1}, \dots, X_{i_m}] = \mathrm{ORDER}^*(X_1, \dots, X_m)$. Then $\delta(X_{i_1}, \dots, X_{i_m})$ is at most* (10) *or* (11). *The complexity of ORDER* is $O(nm^2)$ of pairwise unions and intersections with sets of size at most n.*

Proof. The procedure ORDER* is a slow down variation of ORDER. We add some artificial steps to control better the parameters when the procedure is called recursively. The proof then follows those of Theorems 7.1 and 9.1.

In stage (1), x appears in $u \ge t = \lceil \frac{3m}{n} \rceil$ sets. One extracts x from the sets to produce a family $\{\bar{X}_1, \dots, \bar{X}_m\}$ with parameters $n-1, m, u$. Some $u-t$ 2-sets are transformed into 3-sets by adding appropriate elements. That produces a family $\{\tilde{X}_1, \dots, \tilde{X}_m\}$, with parameters $n-1, m, t$. Then $[\tilde{X}_{i_1}, \dots, \tilde{X}_{i_m}] = \mathrm{ORDER}^*(\tilde{X}_1, \dots, \tilde{X}_m)$, and $[X_{i_1}, \dots, X_{i_m}]$ is to return. Then

$$\delta(X_{i_1}, \dots, X_{i_m}) \le 1 + \delta(\tilde{X}_{i_1}, \dots, \tilde{X}_{i_m}). \tag{12}$$

In stage (a), $[X_1, \dots, X_u]$ is a $[1, L_m, 0]$-covering. Let it contain T (≤ 2)-sets and $\bigcup_{i=1}^u X_i \subseteq U \subseteq \bigcup_{i=1}^m X_i$, where $|U| = u$. We extract U from the sets $\tilde{X}_i = X_i \backslash U$ add some $u-T$ 1-sets to the family. That produces $\mathcal{Y} = \{\tilde{X}_{u+1}, \dots, \tilde{X}_m, \tilde{X}_{m+1}, \dots \tilde{X}_{m+u-T}\}$ with parameters $n-T, m-T, s-T$. Let $[\tilde{X}_{i_{u+1}}, \dots, \tilde{X}_{i_m}]$ be a permutation of $[\tilde{X}_{u+1}, \dots, \tilde{X}_m]$ in $\mathcal{Y}_1 = \mathrm{ORDER}^*(\mathcal{Y})$. Then $[X_1, \dots, X_u, X_{i_{u+1}}, \dots, X_{i_m}]$ is to return and

$$\delta(X_1, \dots, X_u, X_{i_{u+1}}, \dots, X_{i_m}) \le \max\{L_m, \delta(\mathcal{Y}_1)\}. \tag{13}$$

If there are no such U, then $m = u$ and $[X_1, \dots, X_m]$ is to return.

In stage (i), ORDER* works as ORDER. The family $\{\bar{X}_{v+1}, \dots, \bar{X}_m\}$ has parameters $n-T, m-T, s-T$. Then $[\bar{X}_{i_{v+1}}, \dots, \bar{X}_{i_m}] = \mathrm{ORDER}^*(\bar{X}_{v+1}, \dots, \bar{X}_m)$ and

$$[X_{i_{v+1}}, \dots, X_{i_k}, X_{j_1}, \dots, X_{j_v}, X_{i_{k+1}}, \dots, X_{i_m}]$$

is to return, where

$$\delta(X_{i_{v+1}}, \dots, X_{i_k}, X_{j_1}, \dots, X_{j_v}, X_{i_{k+1}}, \dots, X_{i_m}) = \delta(\bar{X}_{i_{v+1}}, \dots, \bar{X}_{i_m}). \tag{14}$$

In stage (ii), let $[X_1, \dots, X_v]$ be a cover for E, it contains T (≤ 2)-sets and $e = |E| \ge T+2$.

Let stage (A). Then $\{X_{v+r+1}, \dots, X_m\}$ has parameters $n-e, m-e+1-r, s-T$. One adds some $e-2$ 1-sets to the latter family, and transforms $e-T-2$ of 1-sets into 3-sets, and gets a family with parameters $n-2, m-1-r, s$. As $r+1 \le t-1$, one can remove any $t-r-2$ 3-sets and get a family $\mathcal{Y}$ with parameters $n-2, m-t+1, s$. As $n \ge 3$, we have $m-1-r \ge m-t+1 \ge s$ and that is possible. One permutes $\mathcal{Y}_1 = \mathrm{ORDER}^*(\mathcal{Y})$. The removed 3-sets

are appended to $\mathcal{Y}_1$, and $e-T-2$ 3-sets are transformed back into 1-sets, and $e-2$ 1-sets are removed. That produces $[X_{i_{v+1+r}}, \ldots, X_{i_m}]$. So $[X_{i_{v+1+r}}, \ldots, X_{i_m}, X_1, \ldots, X_{v+r}]$ is to return and

$$\delta(X_{i_{v+1+r}}, \ldots, X_{i_m}, X_1, \ldots, X_{v+r}) \leq \max\{\delta(\mathcal{Y}_1), L_m\}. \tag{15}$$

Let stage (B). Then $\{\bar{X}_{v+1}, \ldots, \bar{X}_m\}$ has parameters $n-e, m-e+1, s+r-T$. One adds some $e-1$ 1-sets to the latter family, and transforms $e-1+r-T-t \geq 0$ of (≤ 2)-sets into 3-sets. That produces a family $\mathcal{Y}$ with parameters $n-1, m, s+t$. One permutes $\mathcal{Y}_1 = \mathrm{ORDER}^*(\mathcal{Y})$. The above $e-1+r-T-t$ 3-sets are transformed back into (≤ 2)-sets, and $e-1$ 1-sets are removed. That produces $[\bar{X}_{i_{v+1}}, \ldots, \bar{X}_{i_m}]$. So $[X_1, \ldots, X_v, X_{i_{v+1}}, \ldots, X_{i_m}]$ is to return and

$$\delta(X_1, \ldots, X_v, X_{i_{v+1}}, \ldots, X_{i_m}) \leq \max\{L_m, 1+\delta(\mathcal{Y}_1)\}. \tag{16}$$

The proof of the theorem is then by induction and similar to that of Theorem 9.1, where the inequalities (12)-(16) are used instead of (7),(8),(9). The complexity estimate is proved by the same argument as in Theorem 6.1.

□

References

[1] M. Bardet, J.-C.Faugére, and B. Salvy, *Complexity of Gröbner basis computation for semi-regular overdetermined sequences over F_2 with solutions in F_2,* Research report RR–5049, INRIA, 2003.

[2] G. V. Bard, N. T. Courtois, and C. Jefferson, *Efficients methods for conversion and solution of sparse systems of low-degree multivariate polynomials over $GF(2)$ via SAT-solvers,* Cryptology ePrint Archive: Report 2007/024.

[3] B. Buchberger, *Theoretical Basis for the Reduction of Polynomials to Canonical Forms,* SIGSAM Bull. 39(1976), 19-24.

[4] M. Davis, H. Putnam, *A Computing Procedure for Quantification Theory.* Journal of the ACM, vol. 7 (1960), pp. 201-215.

[5] M. Davis, G. Logemann, and D. Loveland, *A Machine Program for Theorem Proving,* Communications of the ACM, vol. 5 (1962), pp. 394-397.

[6] N. T. Courtois and G. V. Bard, *Algebraic Cryptanalysis of the Data Encryption Standard*, Cryptography and Coding 2007, LNCS 4887, pp. 152–169, 2007.

[7] N. Courtois, J. Pieprzyk, *Cryptanalysis of Block Ciphers with Overdefined Systems of Equations* in Asiacrypt 2002, LNCS 2501, pp. 267 – 287, Springer-Verlag, 2002.

[8] N. Courtois, A. Klimov, J. Patarin, and A. Shamir, *Efficient Algorithms for Solving Overdefined Systems of Multivariate Polynomial Equations,* in Eurocrypt 2000, LNCS 1807, pp. 392–407, Springer-Verlag (2000).

[9] N. Eén, N. Sörensson, MiniSat home page, http://minisat.se/

[10] J.-C. Faugère, *A new efficient algorithm for computing Grbner bases (F4),* Journal of Pure and Applied Algebra, vol. 139 (1999), pp. 61-88.

[11] J.-C. Faugère, *A new efficient algorithm for computing Gröbner bases without reduction to zero (F5),* in ISSAC 2002, pp. 75 – 83, ACM Press, 2002.

[12] P. Horak and Z. Tuza, *Speeding up deciphering by hypergraph ordering*, to appear in Des.Codes Cryptogr. (2013).

[13] K. Iwama, K. Seto, T. Takai, and S. Tamaki, *Improved Randomised Algorithms for 3-SAT,*in ISAAC 2010, Part I, LNCS 6506, pp. 73–84, 2010.

[14] D. Lazard, *Gröbner-bases, Gaussian elimination and resolution of systems of algebraic equations*, in EUROCAL 1983, pp. 146–156.
[15] K. Prachar, *Primzahlverteilung*, Springer-Verlag, Berlin, 1957.
[16] H. Raddum, *Solving non-linear sparse equation systems over $GF(2)$ using graphs,* University of Bergen, preprint, 2004.
[17] H. Raddum and I. Semaev, *Solving Multiple Right Hand Sides linear equations*, Designs, Codes and Cryptography, vol. 49, pp. 147–160 (2008), extended abstract in Proceedings of WCC'07, 16-20 April 2007, Versailles, France, INRIA (2007)
[18] I. Semaev, *On solving sparse algebraic equations over finite fields*, Des. Codes Cryptogr., vol. 49 (2008), pp.47–60.
[19] I. Semaev, *Sparse algebraic equations over finite fields,* SIAM J. on Comp., vol. 39(2009), pp. 388–409.
[20] I. Semaev, *Improved Agreeing-Gluing algorithm,* Math. in Comp. Science, vol. 7(2013), pp. 321–339.
[21] B.-Y. Yang, J-M. Chen, and N.Courtois, *On asymptotic security estimates in XL and Gröbner bases-related algebraic cryptanalysis,* LNCS 3269, pp. 401–413, Springer-Verlag, 2004.
[22] A. Zakrevskij, I. Vasilkova,*Reducing large systems of Boolean equations,*4th Int.Workshop on Boolean Problems, Freiberg University, September, 21–22, 2000.

Department of Informatics, University of Bergen, Norway, igor@ii.uib.no