MM Research Preprints
KLMM, Chinese Academy of Sciences
Vol. 27, 16–24, December, 2008

Minimal Achievable Approximation Ratio for MAX-MQ in Finite Fields

Shangwei Zhao and Xiao-Shan Gao
Key Laboratory of Mathematics Mechanization
Institute of Systems Science, AMSS
Chinese Academy of Sciences

Abstract. Given a multivariate quadratic polynomial system in a finite field F_q, the problem MAX-MQ is to find a solution satisfying the maximal number of equations. We prove that the probability of a random assignment satisfying a non-degenerate quadratic equation is at least $\frac{1}{q} - O(q^{-\frac{n}{2}})$, where n is the number of the variables in the equation. Consequently, the random assignment provides a polynomial-time approximation algorithm with approximation ratio $q + O(q^{-\frac{n}{2}})$ for non-degenerate MAX-MQ. For large n, the ratio is close to q. According to a result by Håstad, it is NP-hard to approximate MAX-MQ with an approximation ratio of $q - \epsilon$ for a small positive number ϵ. Therefore, the minimal approximation ratio that can be achieved in polynomial time for MAX-MQ is q.

Keywords. Multivariate quadratic polynomial equations, MAX-MQ, approximation algorithm, approximation ratio.

1. Introduction

Multivariate quadratic polynomial equations play an important role in developing new public key cryptosystems such as the HFE system and the oil-vinegar system and they are listed as one of the main computational problems used in cryptography (Chapter 8 in [2]). The security of the related cryptosystems is based on the difficulty of solving such equation systems in finite fields. Moreover, it is known that any equation system of higher degrees can be transformed into quadratic equations by introducing new variables. So it is important to consider the computational complexity of solving multivariate quadratic polynomial equations.

It is known that to test whether a multivariate quadratic polynomial equation system has a solution in a finite field is a NP complete problem [8]. So, it is natural to consider using approximation algorithms to solve these equations. In order to use approximation algorithms, we need to formulate the original problem as an optimization problem, that is, given a system of m equations with n variables in a finite field, find a solution satisfying the maximal number of equations. For linear equation systems and systems of equations with degree at most two, the optimization problems are denoted as MAX-LIN and MAX-MQ respectively. MAX-LIN and MAX-MQ have been proved to be NP-hard, which means that it is hard to find its exact solutions unless $P = NP$.

For MAX-LIN, since the probability of a random assignment satisfying a linear equation in F_q is $\frac{1}{q}$, a random assignment for the linear system will satisfy $\frac{1}{q}$ fraction of the equations.

Hence, random assignment is a polynomial-time approximation algorithm with approximation ratio (see Section 2 for definition) q for MAX-LIN. Surprisingly, Håstad proved in [4] that it is NP-hard to approximate MAX-LIN with an approximation ratio $q - \epsilon$ where ϵ is an arbitrary small positive number by introducing the 3bit PCP Theorem. Therefore, the random assignment method is the best possible polynomial-time algorithm for the problem MAX-LIN. This deep result gives a complete description of using approximation algorithms to solve MAX-LIN.

It is also NP-hard to approximate MAX-MQ with an approximation factor of $q-\epsilon$ where ϵ is arbitrary small, since MAX-LIN is a subproblem of MAX-MQ. Håstad gave an elementary proof for this result when ϵ is an inverse polynomial [5]. Assuming that each equation has no square terms, Håstad gave a polynomial time approximation algorithm with approximation ratio $\frac{q^2}{q-1}$ [5]. There exists a gap between the approximation ratio $\frac{q^2}{q-1}$ and inapproximability ratio $q-\epsilon$. A natural problem is how to fill the gap. In Håstad's algorithm, the approximation ratio $\frac{q^2}{q-1}$ can be made close to q if q is large.

In this paper, by detailed analyzing the structure of multivariate quadratic equations in finite fields, we prove that the probability of a random assignment satisfying a non-degenerate quadratic equation is at least $\frac{1}{q} - O(q^{-\frac{n}{2}})$,where n is the number of the variables. Consequently, the random assignment provides a polynomial-time approximation algorithm with approximation ratio $q + O(q^{-\frac{n}{2}})$ for non-degenerate MAX-MQ. For large n, the ratio is close to q. Therefore, we may conclude that the minimal approximation ratio (definition in Section 2) that can be achieved in polynomial time for MAX-MQ is q.

The result mentioned above has the following advantages comparing to Håstad's result in [5]. Our approximation ratio $q + O(q^{-\frac{n}{2}})$ depends on n, while Håstad's approximation ratio $\frac{q^2}{q-1}$ only depends on q. In cryptosystems, n usually takes a value greater than or equal to 128 and for such an n, $q + O(q^{-\frac{n}{2}})$ is very close to q. Håstad's result assumes that there exist no square terms in the equation, which is not very natural. We assume that the equation is non-degenerate in the sense that, by performing non-singular substitutions, the equation still has n variables. If the equation is degenerate, then our approximation ratio becomes $q + O(q^{-\frac{k}{2}})$ where k is the smallest number of variables in the new equations obtained by performing non-singular linear substitutions.

The rest of this paper is organized as follows. In Section 2, we introduce the basic notations and main results. In Section 3, we consider the case where q is even. In Section 4, we consider the case where q is odd. Section 5 concludes the paper.

2. Basic notions and main result

First, we will give some definitions about approximation algorithms of optimization problems, detailed description of which can be found in [6, 3].

Definition 2.1 *Let O be a maximization problem and let $r \geq 1$ be a real number. For an instance x of O, let OPT(x) be the optimal value. An r-approximation algorithm is an algorithm that on each input x outputs a number $\widetilde{OPT}(x)$ such that $OPT(x)/r \leq \widetilde{OPT}(x) \leq OPT(x)$.*

We also use the notion "having performance ratio (or factor) r" or "approximation ratio (or factor) r" instead of saying "being an r-approximation algorithm".

Obviously, we only concern the approximation algorithms running in polynomial time.

Definition 2.2 *An optimization problem O is said to be hard to approximate within a factor of r if the existence of an r-approximation algorithm for O implies P=NP.*

Using the constant approximation ratio scheme, we can divide the NP-hard optimization problems into three classes:

1. For any constant $r > 1$, it is hard to approximate O with approximation ratio r.

2. There is a constant $r_0 > 1$ such that for $r > r_0$, O has a polynomial-time r-approximation algorithm while for $1 < r < r_0$, it is hard to approximate O with approximation ratio r.

3. For any constant $r > 1$, O has a polynomial-time r-approximation algorithm.

Each of the three classes is not empty (Chapters 9 and 10 of [6]). For the second class, a basic problem is to determine the minimal achievable approximation ratio r_0, or simply, *minimal ratio.*

In this paper, we will consider the following optimization problem MAX-MQ: given a set of m equations $\{f_i(x_1, \dots, x_n) = b_i\}_{i=1}^m$ such that f_i is of the following form:

$$f(x_1, \dots, x_n) = \sum_{i,j=1}^{n} a_{ij} x_i x_j + \sum_{j=1}^{n} b_j x_j = b \tag{1}$$

where the coefficients and b belong to a finite field F_q, find a solution in F_q^n satisfying the maximal number of equations.

A linear substitution can be expressed by the matrix form $\mathbf{x} = C\mathbf{y}$, where C is an $n \times n$ matrix over F_q and $\mathbf{y}$ is the column vector of new indeterminants $y_1, \dots, y_n$. If C is nonsingular, we call it a nonsingular linear substitution. This transformation keeps the number of solutions unchanged. Two quadratic polynomials f and g over F_q are called *equivalent* if f can be transformed into g by means of a nonsingular linear substitution of indeterminants.

Definition 2.3 *If f is not equivalent to a quadratic polynomial in fewer than n indeterminants, we call f non-degenerate. If f is degenerate, the smallest number of variables in the polynomials equivalent to f is called the* rank *of f. It is clear that a non-degenerate polynomial has rank n.*

The main result of this paper is the following theorem, which is a consequence of Corollaries 3.5 and 4.5 since $\frac{q^2}{q^{n/2}-q} > \frac{q(q-1)}{q^{n/2}-q+1}$.

Theorem 2.4 *The random assignment algorithm is a $(q + \frac{q^2}{q^{n/2}-q})$-approximation algorithm for the non-degenerate MAX-MQ in F_q.*

If some equations in the system are degenerate, the above theorem is still valid if n is taken to be the smallest rank of all the equations.

As a direct consequence of Theorem 2.4, we have

Corollary 2.5 *Let F_q be a finite field and $\epsilon > 0$ a small number. For $n > 2\log_q(\frac{q(q+\epsilon)}{\epsilon})$, the random assignment algorithm is a $(q+\epsilon)$-approximation algorithm for the non-degenerate MAX-MQ in F_q where n is the number of variables.*

Combining the above theorem and Håstad's inapproximability result, we may conclude that MAX-MQ belongs to the second class optimization problem with minimal approximation ratio q. This result gives a clear description of MAX-MQ when using constant ratio approximation scheme.

3. Case 1: q is even

A *quadratic form* in n indeterminants over F_q is a homogeneous polynomial of degree two:

$$f(x_1,\ldots,x_n)=\sum_{i,j=1}^{n} a_{ij}x_ix_j$$

Definition 3.1 *For $a \in E = F_{q^l}$ and $K = F_q$, the trace $Tr_{E/K}(a)$ of a over K is defined by*

$$Tr_{E/K}(a)=a+a^q+\cdots+a^{q^{l-1}}$$

If q is a prime number, then $Tr_{E/K}(a)$ is called the absolute trace of a and simply denoted by $Tr_E(a)$.

Definition 3.2 *For any finite field F_q, the integer-valued function v on F_q is defined by $v(b) = -1$ for $b \in F_q^*$ and $v(0) = q-1$.*

We use $N(f(x_1,\ldots,x_n) = b)$ to represent the number of solutions of the equation $f(x_1,\ldots,x_n) = b$ in F^q.

Lemma 3.3 *[7] (Page 287, Page 288) Let $f \in F_q[x_1,\ldots,x_n]$, q even, be a non-degenerate quadratic form.*

1. If n is odd, then f is equivalent to $x_1x_2+x_3x_4+\ldots+x_{n-2}x_{n-1}+x_n^2$, and $N(x_1x_2+x_3x_4+\ldots+x_{n-2}x_{n-1}+x_n^2=b)=q^{n-1}$ in F_q^n.

2. If n is even, then f is either equivalent to $x_1x_2+x_3x_4+\ldots+x_{n-1}x_n$ or to $x_1x_2+x_3x_4+\ldots+x_{n-1}x_n+x_{n-1}^2+ax_n^2$ where $a \in F_q$ satisfies $Tr_{F_q}(a) = 1$, and the corresponding number of solutions is as follows:

$N(x_1x_2+x_3x_4+\ldots+x_{n-1}x_n=b)=q^{n-1}+v(b)q^{(n-2)/2}$,

$N(x_1x_2+x_3x_4+\ldots+x_{n-1}x_n+x_{n-1}^2+ax_n^2=b)=q^{n-1}-v(b)q^{(n-2)/2}$.

Given a system of multivariate equations $\{f_i(x_1,\ldots,x_n) = b_i\}_{i=1}^m$, the random assignment method assigns random values to each x_i and count the number of equations with these values as solutions. To estimate the approximation ratio, we need to give a lower bound for the number of solutions for a single quadratic equation.

Theorem 3.4 *Let $f(x_1,\ldots,x_n)=\sum_{i,j=1}^{n}a_{ij}x_ix_j+\sum_{j=1}^{n}b_jx_j$ be a non-degenerate quadratic polynomial over F_q, where q is even. Then*

$$N(f(x_1,\ldots,x_n)=b)\geq q^{n-1}-q^{n/2}.$$

Proof: Denote $f(x_1,\dots,x_n)=\sum_{i,j=1}^{n}a_{ij}x_ix_j+\sum_{j=1}^{n}b_jx_j=f_1+f_2$. We will discuss it in three cases :

Case 1: If f is equivalent to a polynomial g without linear terms, then the quadratic part of g contains n variables since f is non-degenerate. By Lemma 3.3 we have

$$N(f(x_1,\dots,x_n)=b)\geq q^{n-1}-(q-1)q^{(n-2)/2}>q^{n-1}-q^{n/2}.$$

Case 2: If f is equivalent to a polynomial g whose quadratic part contains less than n variables, that is, rank(f_1)$<n$, then g must have a nonzero linear term c_kx_k such that x_k does not occur in the quadratic terms. It is clear that assigning random values in F_q to the other $n-1$ variables, x_k will be uniquely determined. Hence,

$$N(f(x_1,\dots,x_n)=b)=q^{n-1}.$$

Case 3: If f is equivalent to a polynomial g which has nonzero linear terms, and $g's$ quadratic part contains n variables, that is, rank(f_1)$=n$, then by introducing a new variable x_{n+1}, f becomes

$$\sum_{i,j=1}^{n}a_{ij}x_ix_j+\sum_{j=1}^{n}b_jx_jx_{n+1}=bx_{n+1}^2.$$

We can rewrite it as

$$f'(x_1,\dots,x_n,x_{n+1})=\sum_{i,j=1}^{n+1}a'_{ij}x_ix_j=0.$$

Since f is degenerate, we have rank(f')$\geq n$.

It is obvious that

$$\begin{aligned}&N(f(x_1,\dots,x_n)=b)\\&=N(f'(x_1,\dots,x_n,x_{n+1})=0|_{x_{n+1}=1})\\&=\frac{1}{q-1}N(f'(x_1,\dots,x_n,x_{n+1})=0|_{x_{n+1}\neq0})\\&=\frac{1}{q-1}(N(f'(x_1,\dots,x_n,x_{n+1})=0)-N(f'(x_1,\dots,x_n,x_{n+1})=0|_{x_{n+1}=0})).\end{aligned}$$

The second equality holds for the following reason:

Suppose $\sum_{i,j=1}^{n}a_{ij}y_iy_j+\sum_{j=1}^{n}b_jy_j=b$ has a solution $(y_1,\dots,y_n)$, let $y_i=\frac{x_i}{x_{n+1}}$ with $x_{n+1}\neq0$. Then for each $x_{n+1}=j, j=1,\dots,q-1$, we will obtain a solution $(x_1,\dots,x_n,j)$ for the equation

$$\sum_{i,j=1}^{n}a_{ij}x_ix_j+\sum_{j=1}^{n}b_jx_jx_{n+1}=bx_{n+1}^2,$$

that is, $f'(x_1,\dots,x_n,x_{n+1})=0$.

If

$$N(\sum_{i,j=1}^{n}a_{ij}y_iy_j+\sum_{j=1}^{n}b_jy_j=b)=t$$

then

$$N(f'(x_1,\dots,x_n,x_{n+1})=0|_{x_{n+1}\neq 0})=(q-1)t$$

and the number of solutions for $f'(x_1,\dots,x_n,x_{n+1})=0$ with $x_{n+1}=1$ is t. Hence,

$$N(f'(x_1,\dots,x_n,x_{n+1})=0|_{x_{n+1}=1})=\frac{1}{q-1}N(f'(x_1,\dots,x_n,x_{n+1})=0|_{x_{n+1}\neq 0}).$$

We will next consider two cases:

1: n is even. Since

$$f'(x_1,\dots,x_n,x_{n+1})|_{x_{n+1}=0}=f_1=\sum_{i,j=1}^{n}a_{ij}x_ix_j,$$

and rank(f_1)= n, by Lemma 3.3 we have

$$N(f'(x_1,\dots,x_n,x_{n+1})|_{x_{n+1}=0}=0)\le q^{n-1}+(q-1)q^{(n-2)/2}.$$

Also by Lemma 3.3, $N(f'(x_1,\dots,x_n,x_{n+1})=0)=q^n$, since $n+1$ is odd. Hence,

$$\begin{aligned}N(f(x_1,\dots,x_n)=b)&\ge\frac{1}{q-1}(q^n-(q^{n-1}+(q-1)q^{(n-2)/2}))\\&=q^{n-1}-q^{(n-2)/2}>q^{n-1}-q^{n/2}.\end{aligned}$$

2: n is odd. Since n is odd and $n+1$ is even, by Lemma 3.3, we have

$$N(f'(x_1,\dots,x_n,x_{n+1})|_{x_{n+1}=0}=0)=N(\sum_{i,j=1}^{n}a_{ij}x_ix_j=0)=q^{n-1}$$

and since rank(f')= $k\ge n$,

$$\begin{aligned}N(f'(x_1,\dots,x_n,x_{n+1})=0)&\ge(q^{k-1}-(q-1)q^{(k-2)/2})q^{n+1-k}\\&=q^n-(q-1)q^{n-\frac{k}{2}}\\&\ge q^n-(q-1)q^{\frac{n}{2}}.\end{aligned}$$

Hence,

$$\begin{aligned}N(f(x_1,\dots,x_n)=b)&\ge\frac{1}{q-1}(q^n-(q-1)q^{n/2}-q^{n-1})\\&>q^{n-1}-q^{n/2}.\end{aligned}$$

Therefore, we conclude

$$N(f(x_1,\dots,x_n)=b)\ge q^{n-1}-q^{n/2}$$

no matter n is even or odd and the theorem holds. □

Corollary 3.5 *The random assignment algorithm is a $(q+\frac{q^2}{q^{n/2}-q})$-approximation algorithm for the non-degenerate MAX-MQ in F_q if q is even.*

Proof: By Theorem 3.4, the number of solutions for each non-degenerate quadratic equation is at least $q^{n-1}-q^{n/2}$. Then the probability of a random assignment to the variables satisfying the equation is at least

$$\frac{q^{n-1}-q^{n/2}}{q^n}=\frac{1}{q}-\frac{1}{q^{n/2}}.$$

Therefore, a random assignment to the variables will satisfy at least $\frac{1}{q}-\frac{1}{q^{n/2}}=\frac{1}{q+\frac{q^2}{q^{n/2}-q}}$ fraction of the equations in the equation system. Hence, it is a $q+\frac{q^2}{q^{n/2}-q}$-approximation algorithm for the problem MAX-MQ. □

4. Case 2: q is odd

Consider a quadratic form in n indeterminants over F_q

$$f(x_1,\ldots,x_n)=\sum_{i,j=1}^{n} b_{ij}x_ix_j.$$

If q is odd, we can write each $b_{ij}x_ix_j$ as $\frac{1}{2}b_{ij}x_ix_j+\frac{1}{2}b_{ij}x_ix_j$ and f can be represented as

$$f(x_1,\ldots,x_n)=\sum_{i,j=1}^{n} a_{ij}x_ix_j$$

where $a_{ij}=a_{ji}$. Let A be the $n\times n$ matrix whose (i,j) entry is a_{ij}. Then f can be written as the matrix form $\mathbf{x}^{\mathbf{T}}A\mathbf{x}$ with $A^T=A$. We can also apply nonsingular linear substitutions to reduce a quadratic form to standard forms.

Proposition 4.1 *[7] (Page 280) If q is odd, every quadratic form over F_q is equivalent to a diagonal quadratic form $a_1x_1^2+\ldots+a_kx_k^2$, where $a_i\in F_q$ and $k\le n$ is the rank of f.*

The quadratic form $f=\mathbf{x}^{\mathbf{T}}A\mathbf{x}$ is non-degenerate if and only if A has rank n. For a non-degenerate f, we may define the determinant $\det(f)$ of f to be the determinant of A.

Lemma 4.2 *[7](Page 282) Let f be a non-degenerate quadratic form over F_q in an even number n of indeterminants. Then for $b\in F_q$, the number of solutions of the equation $f(x_1,\ldots,x_n)=b$ in F_q^n is*

$$q^{n-1}+v(b)q^{(n-2)/2}\eta((-1)^{n/2}\det(f))$$

Where η is the quadratic character function of F_q whose value is 1 or -1.

Lemma 4.3 *[7] (Page 283) Let f be a non-degenerate quadratic form over F_q in an odd number n of indeterminants. Then for $b\in F_q$ the number of solutions of the equation $f(x_1,\ldots,x_n)=b$ in F_q^n is*

$$q^{n-1}+q^{(n-1)/2}\eta((-1)^{(n-1)/2}\det(f)b)$$

The following theorem gives a lower bound for the number of solutions for a single quadratic equation.

Theorem 4.4 *Let $f(x_1, \ldots, x_n) = \sum_{i,j=1}^{n} a_{ij}x_ix_j + \sum_{j=1}^{n} b_jx_j$ over the finite field F_q where q is odd. Assume that f is non-degenerate. Then*

$$N(f(x_1, \ldots, x_n) = b) \geq q^{n-1} - (q-1)q^{(n-2)/2}$$

Proof: Note that $q^{n-1} - q^{(n-1)/2} > q^{n-1} - (q-1)q^{(n-2)/2}$. If f is equivalent to a polynomial g without linear terms, then g must have n variables in its quadratic part and $N(f(x_1, \ldots, x_n) = b) \geq q^{n-1} - (q-1)q^{(n-2)/2}$ by Lemma 4.2 and Lemma 4.3.

If f is equivalent to a polynomial g whose linear terms are not zero, since f is non-degenerate, then g has the following two forms by Proposition 4.1:

$$a_1'x_1^2 + \ldots + a_n'x_n^2 + b_1'x_1 + \ldots + b_k'x_k = b, 1 \leq k \leq n$$

or

$$a_1'x_1^2 + \ldots + a_l'x_l^2 + b_1'x_1 + \ldots + b_n'x_n = b, l < n$$

For the first case, substituting $x_i = y_i - b_i(2a_i)^{-1}, i = 1, \ldots, k$, we have $a_1'y_1^2 + \ldots + a_n'y_n^2 = c$ for some c in F_q.

Since nonsingular linear substitution does not change the number of solutions, by Lemma 4.2 and Lemma 4.3,

$$\begin{aligned} N(f(x_1, \ldots, x_n) = b) &= N(a_1'y_1^2 + \ldots + a_n'y_n^2 = c) \\ &\geq q^{n-1} - (q-1)q^{(n-2)/2}. \end{aligned}$$

For the second case, it is obvious that $b_n' \neq 0$, by assigning random values to $x_1, \ldots, x_{n-1}$, the value of x_n is uniquely determined. Hence, $N(f(x_1, \ldots, x_n) = b) = q^{n-1} \geq q^{n-1} - (q-1)q^{(n-2)/2}$. □

For each equation of $\{f_i(x_1, \ldots, x_n) = b_i\}_{i=1}^{m}$, the probability of a random assignment to the variables satisfying the equation is at least

$$\frac{q^{n-1} - (q-1)q^{(n-2)/2}}{q^n} = \frac{1}{q} - \frac{q-1}{q^{(n+2)/2}} = \frac{1}{q + \frac{q(q-1)}{q^{n/2} - q + 1}}.$$

Therefore, we have the following result.

Corollary 4.5 *The random assignment algorithm is a $(q + \frac{q(q-1)}{q^{n/2}-q+1})$-approximation algorithm for the non-degenerate MAX-MQ in F_q if q is odd.*

5. Conclusion

We show that the problem MAX-MQ can be approximated with an approximation ratio $q + O(q^{-\frac{n}{2}})$, where n is the number of variables if each equation is non-degenerate. Combining this result with Håstad's inapproximability result, we may conclude that for any F_q, q is the minimal achievable approximation ratio for MAX-MQ. It is still an interesting problem to find a polynomial approximation algorithm with approximation ratio $q + \epsilon$ for a fixed number $\epsilon \geq 0$.

References

[1] Arora, S., Lund, C., Motawani, R., Sudan, M., Szegedy, M., Proof verification and the hardness of approximation problems, *Journal of the ACM*, 45(3): 501-555, 1998.

[2] Coron, J.S. and Benne de Weger (eds), *Hardness of the main computational problems used in cryptograpghy*, ECRYPT, IST-202-507932, March, 2007.

[3] Du D.Z. and Ko, K., *Theory of computational Complexity*, John Wiley & Sons, New York, 2000.

[4] Håstad, J., Some optimal inapproximability results. *Journal of the ACM*, 48(4): 798-859, 2001.

[5] Håstad, J., Phillips, S., Safra, S., A well-characterized approximation problem. *Theory and Computing Systems*, Proceedings of the 2nd Israel Symposium, 1993.

[6] Hochbaum, D.S., *Approximation algorithms for NP-hard problems*, International Thomson Publishing Company, 1997.

[7] Lidl, R. and Niederreiter, H., *Finite fields.* Addison-Wesley Publishing Company, 1983.

[8] Wolf, C., *Hidden field equations - variations and attacks.* Diplomarbeit, Universitåt Ulm, December, 2002.