Modular Counting of Rational Points over Finite Fields

Daqing Wan

Department of Mathematics
University of California
Irvine, CA 92697-3875
dwan@math.uci.edu

Abstract

Let $\mathbb{F}_q$ be the finite field of q elements, where $q = p^h$. Let $f(x)$ be a polynomial over $\mathbb{F}_q$ in n variables with m non-zero terms. Let $N(f)$ denote the number of solutions of $f(x) = 0$ with coordinates in $\mathbb{F}_q$. In this paper, we give a deterministic algorithm which computes the reduction of $N(f)$ modulo p^b in $O(n(8m)^{(h+b)p})$ bit operations. This is singly exponential in each of the parameters $\{h, b, p\}$, answering affirmatively an open problem proposed in [5].

1 Introduction

Let $\mathbb{F}_q$ be the finite field of q elements, where $q = p^h$ and p is a prime. Let $f(x_1, \cdots, x_n)$ be a polynomial in n variables with coefficients in $\mathbb{F}_q$ and with sparse representation. That is, f is written as a sum of m non-zero monomials:

$$f(x_1, \cdots, x_n) = \sum_{j=1}^{m} a_j x^{V_j}, a_j \in \mathbb{F}_q^*,$$

where

$$V_j = (V_{1j}, \cdots, V_{nj}) \in \mathbb{Z}_{\geq 0}^n, \ x^{V_j} = x_1^{V_{1j}} \cdots x_n^{V_{nj}}.$$

Let $N(f)$ denote the number of $\mathbb{F}_q$-rational points on the affine hypersurface defined by $f = 0$. It is clear that $0 \leq N(f) \leq q^n$. Replacing

x_i^q by x_i, we may assume that the degree of f in each variable is at most $q-1$. The input size for f is $mn\log(q)$. A fundamental problem of great importance is to compute the number $N(f)$ efficiently and deterministically. The nature of this problem varies quite a bit depending on the range of the various parameters. As a consequence, it has been studied extensively in a number of directions, by both mathematicians and computer scientists. For $f(x) = a_1x_1^{q-1} + \cdots + a_nx_n^{q-1} + b$, deciding if the number $N(f)$ is positive is the well known subset sum problem over $\mathbb{F}_q$, which is NP-complete if $p > 2$. Ehrenfeucht and Karpinski [3] showed that computing $N(f)$ is $\#P$-complete even when we restrict the degree to be three. Von zur Gathen et.al [9] showed that it is also $\#P$-complete when we restrict to curves ($n = 2$). The harder problem of computing the zeta function of an algebraic variety over a finite field is well studied by p-adic methods, see Wan [11][12] for a survey and the references there. In particular, a polynomial time algorithm in dense input size to compute the zeta function is obtained in Lauder-Wan [7] if the characteristic p is small and if the number of variables is fixed. In the case that f defines a smooth projective hypersurface of degree not divisible by $p > 2$, Lauder [6] gave a p-adic polynomial time algorithm in dense input size for computing the zeta function if p is small (the number of variables can be large). Note that in the case of zeta functions, one has to use dense input size, as the output size is already as large as the size of the dense input.

For a positive integer $r > 1$, let $N_r(f)$ denote the least non-negative residue of $N(f)$ modulo r. It is clear that $N_r(f) = N(f)$ if $r > q^n$. Thus, it is no easier to compute $N_r(f)$ in general for large r. The problem of computing $N_r(f)$ can be viewed as a non-archimedian approximation problem. It gives the residue class of the integer $N(f)$ modulo r. By the Chinese remainder theorem, we can assume that r is a prime power. One can hope that it may be easier to compute $N_r(f)$ for small fixed r. Even this is hard. For fixed r which is not a power of p, computing $N_r(f)$ is shown to be NP-hard in Gopalan-Guruswami-Lipton (GGL in short) [5]. Thus, we shall assume that r is a power of p throughout this paper. If $r = p^b$ is small, a polynomial time algorithm to compute $N_{p^r}(f)$ using dense input size is described in Wan [11] using the reduction of the Dwork trace formula.

If $r = p^b$ is a power of p, GGL [5] proved that computing $N_{p^b}(f)$ is also NP-hard if either $p \geq 2n$ or if $h \geq 2n$ or if $b > nh$ (i.e., $p^b > q^n$). This shows that exponential dependence on each of $\{p, b, h\}$ cannot be avoided in computing $N_{p^b}(f)$. Using modulus amplifying polynomials, they constructed an algorithm to compute $N_{p^b}(f)$ in time $O(nm^{2qb})$. In the case $q = p$ (i.e.,

$h = 1$), the running time $O(nm^{2pb})$ is singly exponential both in terms of p and b, and thus essentially optimal except for a possible constant multiple in the exponent. The constant 2 comes from the degree of the optimal modulus amplifying polynomial. For general $q = p^h$, the running time $O(nm^{2qb})$ is doubly exponential in h. By reducing equations over $\mathbb{F}_q$ to equations over $\mathbb{F}_p$, GGL [5] constructed a modified version of their algorithm which computes $N_{p^b}(f)$ in time $O(n(mh^w)^{2hpb})$, where w is the p-weight degree of f. This running time is singly exponential in h but with an extra exponential dependence on the p-weight degree of f. GGL [5] raised the open problem: Is there an algorithm to compute $N_{p^b}(f)$ over $\mathbb{F}_q$ which is singly exponential in p and h?

In this paper, we provide an affirmative answer to this question. We have

Theorem 1.1 *Let $f(x)$ be a polynomial in n variables with m monomials over $\mathbb{F}_q$, where $q = p^h$. There is a deterministic algorithm which computes $N_{p^b}(f)$ in*

$$O(nmp^b(\frac{3mh}{(h+b)(p-1)} + 3)^{(h+b)(p-1)}) = O(n(8m)^{(h+b)p})$$

bit operations.

Our methods are completely different from the modulus amplifying polynomial approach used in [5]. Low degree modulo amplifying polynomials first appeared in the work of Toda [8] and further applications were made in the work of Yao [14] and Beigel-Tarui [1]. In contrast, the idea of our algorithm is to use a simple formula [10][13] for $N(f)$ (and more general for exponential sums) in terms of Gauss sums, and then applying the Gross-Koblitz formula relating Gauss sums to the p-adic Γ-function. Candelas and his collaborators [2] have used similar techniques to calculate the zeta function in some special cases.

Remarks: An alternative approach to the main result of this paper is to use Dwork's p-adic analytic construction of the additive character and the Dwork trace formula as used in our previous work [11][7]. In this paper, we only consider the hypersurface (one equation) case. The more general case of a system of polynomial equations can be easily reduced to the one equation case, see Wan [12] for various reductions.

Acknowledgements. It is a pleasure to thank A.C. Yao and Xiaoming Sun for helpful discussions on the subset sum problem over finite fields. This work was partially supported by NSF.

2 A counting formula via Gauss sums

Let $V_1, \cdots, V_m$ be m distinct lattice points in $\mathbb{Z}_{\geq 0}^n$. For $V_j = (V_{1j}, \cdots, V_{nj})$, write

$$x^{V_j} = x_1^{V_{1j}} \cdots x_n^{V_{nj}}.$$

We may assume that $0 \leq V_{ij} \leq q-1$. Let f be the polynomial in n variables over $\mathbb{F}_q$ written in the form:

$$f(x_1, \cdots, x_n) = \sum_{j=1}^{m} a_j x^{V_j}, a_j \in \mathbb{F}_q^*.$$

Let $N(f)$ denote the number of $\mathbb{F}_q$-rational points on the affine hypersurface $f = 0$. We first review the formula in [10][13] for $N(f)$ in terms of Gauss sums.

Let $\mathbb{Z}_p$ be the ring of integers in the field $\mathbb{Q}_p$ of p-adic rational numbers. Let $\mathbb{Z}_q$ be the ring of integers in the unique unramified extension of $\mathbb{Q}_p$ with residue field $\mathbb{F}_q$. Let ω be the Teichmüller character of the multiplicative group $\mathbb{F}_q^*$. For $a \in \mathbb{F}_q^*$, the value $\omega(a)$ is just the $(q-1)$-th root of unity in $\mathbb{Z}_q$ such that $\omega(a)$ modulo p reduces to a. Define the $(q-2)$ Gauss sums over $\mathbb{F}_q$ by

$$G(k) = \sum_{a \in \mathbb{F}_q^*} \omega(a)^{-k} \zeta_p^{\mathrm{Tr}(a)} \quad (1 \leq k \leq q-2),$$

where ζ_p is a fixed primitive p-th root of unity in an extension of $\mathbb{Q}_p$ and Tr denotes the trace map from $\mathbb{F}_q$ to the prime field $\mathbb{F}_p$.

Lemma 2.1 *For all $a \in \mathbb{F}_q$, the Gauss sums satisfy the following interpolation relation*

$$\zeta_p^{\mathrm{Tr}(a)} = \sum_{k=0}^{q-1} \frac{G(k)}{q-1} \omega(a)^k,$$

where

$$G(0) = q-1, \ G(q-1) = -q.$$

Proof. By the Vandermonde determinant, there are numbers $C(k)$ $(0 \leq k \leq q-1)$ such that for all $a \in \mathbb{F}_q$, one has

$$\zeta_p^{\mathrm{Tr}(a)} = \sum_{k=0}^{q-1} \frac{C(k)}{q-1} \omega(a)^k.$$

It suffices to prove that $C(k) = G(k)$ for all k. Taking $a = 0$, one finds that $C(0)/(q-1) = 1$. This proves that $C(0) = q - 1 = G(0)$. For $1 \leq k \leq q-2$, one computes that

$$G(k) = \sum_{a \in \mathbb{F}_q^*} \omega(a)^{-k} \zeta_p^{\mathrm{Tr}(a)} = \frac{C(k)}{q-1}(q-1) = C(k).$$

Finally,

$$0 = \sum_{a \in \mathbb{F}_q} \zeta_p^{\mathrm{Tr}(a)} = \frac{C(0)}{q-1} q + \frac{C(q-1)}{q-1}(q-1).$$

This gives $C(q-1) = -q = G(q-1)$. The lemma is proved.

Now we turn to deriving a counting formula for $N(f)$ in terms of Gauss sums. Write $W_j = (1, V_j) \in \mathbb{Z}_{\geq 0}^{n+1}$. Then,

$$x_0 f = \sum_{j=1}^{m} a_j x^{W_j} = \sum_{j=1}^{m} a_j x_0 x_1^{V_{1j}} \cdots x_n^{V_{nj}},$$

where x now has $n+1$ variables $\{x_0, \cdots, x_n\}$. Using the formula

$$\sum_{t \in \mathbb{F}_q} \omega(t)^k = \begin{cases} 0, & \text{if } (q-1) \not| k, \\ q-1, & \text{if } (q-1) | k \text{ and } k > 0, \\ q, & \text{if } k = 0, \end{cases}$$

one then calculates that

$$\begin{aligned} N(f) &= \frac{1}{q} \sum_{x_0, \cdots, x_n \in \mathbb{F}_q} \zeta_p^{\mathrm{Tr}(x_0 f(x))} \\ &= \frac{1}{q} \sum_{x_0, \cdots, x_n \in \mathbb{F}_q} \prod_{j=1}^{m} \zeta_p^{\mathrm{Tr}(a_j x^{W_j})} \\ &= \frac{1}{q} \sum_{x_0, \cdots, x_n \in \mathbb{F}_q} \prod_{j=1}^{m} \sum_{k_j=0}^{q-1} \frac{G(k_j)}{q-1} \omega(a_j)^{k_j} \omega(x^{W_j})^{k_j} \\ &= \frac{1}{q} \sum_{k_1=0}^{q-1} \cdots \sum_{k_m=0}^{q-1} \Big(\prod_{j=1}^{m} \frac{G(k_j)}{q-1} \omega(a_j)^{k_j} \Big) \sum_{x_0, \cdots, x_n \in \mathbb{F}_q} \omega\big(x^{k_1 W_1 + \cdots + k_m W_m}\big) \\ &= \sum_{\sum_{j=1}^{m} k_j W_j \equiv 0 (\mathrm{mod}\ q-1)} \frac{(q-1)^{s(k)} q^{n-s(k)}}{(q-1)^m} \prod_{j=1}^{m} \omega(a_j)^{k_j} G(k_j), \qquad (1) \end{aligned}$$

where $s(k)$ denotes the number of non-zero entries in the integer vector $k_1W_1 + \cdots + k_mW_m \in \mathbb{Z}_{\geq 0}^{n+1}$. It is clear that

$$0 \leq s(k) \leq n+1.$$

Formula (1) was used in [13] to prove the mirror congruence formula for a strong mirror pair of Calabi-Yau hypersurfaces. In this paper, we use formula (1) to derive an algorithm to compute $N_{p^b}(f)$.

Note that each term in formula (1) is an algebraic integer. The number of terms in formula (1) can be as large as q^m, which is too big and exponential in terms of m. Fortunately, many of these terms are actually zero modulo p^b. This follows from the Stickelberger theorem described in next section.

3 Stickelberger, Gross-Koblitz formula

Let π be the unique element in $\mathbb{Z}_p[\zeta_p]$ satisfying

$$\pi^{p-1} = -p, \ \pi \equiv \zeta_p - 1 (\mathrm{mod}(\zeta_p - 1)^2).$$

Thus, $\mathrm{ord}_p(\pi) = 1/(p-1)$ and π is a uniformizer in the complete discrete valuation ring $\mathbb{Z}_p[\zeta_p]$.

Theorem 3.1 (Stickelberger) *For integer $0 \leq k \leq q-2$, write*

$$k = k_0 + k_1 p + \cdots + k_{h-1}p^{h-1}$$

in p-adic expansion, where $0 \leq k_i \leq p-1$. Let $\sigma(k) = k_0 + \cdots + k_{h-1}$ be the sum of the p-digits of k. Then,

$$G(k) \equiv \frac{-\pi^{\sigma(k)}}{k_0!k_1!\cdots k_{h-1}!} (\mathrm{mod}\pi^{\sigma(k)+p-1}).$$

The Stickelberger theorem gives the first non-zero digit in the π-adic expansion of the Gauss sum. To get higher digits, we can use the Gross-Koblitz formula which gives all digits of the Gauss sum in terms of the p-adic Γ-function.

Let $\mathbb{Z}_p$ be the ring of p-adic integers. The p-adic Γ-function $\Gamma_p(z)$ is a continuous function from $\mathbb{Z}_p$ to $\mathbb{Z}_p$ which is determined by its values on $\mathbb{Z}_{\geq 0}$ defined as follows:

$$\Gamma_p(0) = 1, \ \Gamma_p(n) = (-1)^n \prod_{j \not\equiv 0 (\mathrm{mod}\ p), 1 \leq j \leq n-1}^{n-1} j, \ (n \geq 1).$$

If $0 \leq n_1 \leq n_2$ are two integers such that $n_1 \equiv n_2 (\mathrm{mod} p^a)$, then we have the p-adic continuity relation:

$$\Gamma_p(n_1) \equiv \Gamma_p(n_2) (\mathrm{mod} p^a).$$

For a p-adic integer

$$z = \sum_{i=0}^{\infty} z_i p^i, \ 0 \leq z_i \leq p-1,$$

we can then define

$$\Gamma_p(z) = \lim_{i \to \infty} \Gamma_p(z_0 + z_1 p + \cdots + z_i p^i).$$

This limit exits and is in $\mathbb{Z}_p$. To compute the reduction of $\Gamma_p(z)$ modulo p^b, it suffices to compute

$$\Gamma_p(z_0 + z_1 p + \cdots + z_{b-1} p^{b-1}) \equiv \Gamma_p(z) (\mathrm{mod} p^b),$$

which takes at most p^b multiplications in $\mathbb{Z}/p^b\mathbb{Z}$.

For an integer k, let $\langle k \rangle_{q-1}$ denote the least non-negative residue of k modulo $q-1$. The rational number $\langle k \rangle_{q-1}/(q-1)$ is clearly a p-adic integer in $\mathbb{Z}_p$.

Theorem 3.2 (Gross-Koblitz [4]) *For integer $1 \leq k \leq q-2$, we have the formula*

$$\frac{G(k)}{\pi^{\sigma(k)}} = -\prod_{i=0}^{h-1} \Gamma_p(\frac{\langle p^i k \rangle_{q-1}}{q-1}) \in \mathbb{Z}_p.$$

Note that the quotient $\frac{G(k)}{\pi^{\sigma(k)}} \in \mathbb{Z}_p$ also holds for $k = 0$ and $k = q-1$, given by

$$\frac{G(0)}{\pi^0} = q-1, \ \frac{G(q-1)}{\pi^{h(p-1)}} = \frac{-q}{(-p)^h} = (-1)^{h-1}.$$

4 Modular counting algorithm

We now show that formula (1) can be used to compute $N_{p^b}(f)$, the reduction of $N(f)$ modulo p^b, in the time as stated in Theorem 1.1. In fact, the

Stickelberger theorem shows that in formula (1), it suffices to restrict to those terms satisfying

$$\sigma(k_1) + \cdots + \sigma(k_m) \leq (h+b)(p-1) - 1,\ 0 \leq k_j \leq q-1.$$

Write

$$k_j = \sum_{i=0}^{h-1} k_{ij} p^i,\ 0 \leq k_{ij} \leq p-1.$$

Let S_b be the set of non-negative integer vector solutions $k = (k_1, \cdots, k_m)$ of the system

$$\sum_{j=1}^{m} k_j W_j \equiv 0 (\text{mod } q-1),\ \sum_{i=0}^{h-1} \sum_{j=1}^{m} k_{ij} \leq (h+b)(p-1) - 1.$$

Reducing formula (1), we obtain the following congruence modulo p^b:

$$N_{p^b}(f) \equiv \sum_{k \in S_b} \frac{(q-1)^{s(k)} q^{n-s(k)}}{(q-1)^m} (-p)^{\frac{\sigma(k_1)+\cdots+\sigma(k_m)}{p-1}} \prod_{j=1}^{m} \omega(a_j)^{k_j} \frac{G(k_j)}{\pi^{\sigma(k_j)}}. \qquad (2)$$

Note that the quotient $\frac{\sigma(k_1)+\cdots+\sigma(k_m)}{p-1}$ is always a non-negative integer. This follows from the first equation

$$k_1 + \cdots + k_m \equiv 0 (\text{mod}(q-1))$$

in the linear system $\sum_j k_j W_j \equiv 0(\text{mod}(q-1))$. By our definition of S_b, it is clear that the cardinality of S_b is bounded by the number A of non-negative integer solutions of the inequality

$$\sum_{i=0}^{h-1} \sum_{j=1}^{m} k_{ij} \leq (h+b)(p-1) - 1. \qquad (3)$$

One calculates that

$$A = \binom{mh + (h+b)(p-1) - 1}{(h+b)(p-1) - 1} \leq \frac{(mh + (h+b)(p-1))^{(h+b)(p-1)-1}}{((h+b)(p-1)-1)!}.$$

By the Stirling formula, we deduce that

$$A \leq \left(\frac{emh}{(h+b)(p-1)} + e \right)^{(h+b)(p-1)},$$

where e is the base in the natural logarithm. To compute the Teichmüller lifting $\omega(a_j)$ modulo p^b, one first lifts a_j to b_j modulo p^b, then computes

$$\omega(a_j) \equiv b_j^{p^{b-1}} (\mathrm{mod} p^b).$$

This takes $O(b\log(p))$ multiplications in $\mathbb{Z}/p^b\mathbb{Z}$. For each term in (1), the product of the Gauss sums can be computed via the p-adic Γ-function in mp^b operations in $\mathbb{Z}/p^b\mathbb{Z}$. Thus, the total time to compute $N_{p^b}(f)$ is bounded by

$$O(nmp^b(\frac{emh}{(h+b)(p-1)} + e)^{(h+b)(p-1)}) = O(n(8m)^{(h+b)p}).$$

The theorem is proved.

We summarize our algorithm below.

Algorithm 4.1

Input: $f(x) = \sum_{j=1}^m a_j X^{V_j}$ over $\mathbb{F}_q$.

1. Compute the Teichmüller liftings $\omega(a_j)$ modulo p^b for all $1 \leq j \leq m$.
2. List the solutions $k = (k_1, ..., k_m)$ of the inequality

$$\sum_{i=0}^{h-1} \sum_{j=1}^{m} k_{ij} \leq (h+b)(p-1) - 1,\ 0 \leq k_{ij} \leq p-1,$$

where $k_j = \sum_{i=0}^{h-1} k_{ij} p^i$. This can be done in time A.

3. For each solution $k = (k_1, \cdots, k_m)$ in Step 2 which satisfies

$$\sum_{j=1}^{m} k_j W_j \equiv 0 (\mathrm{mod}(q-1)), h(n-s(k)) + \frac{1}{p-1} \sum_{j=1}^{m} \sigma(k_j) \leq b-1,$$

use the Gross-Koblitz formula to compute the corresponding term in formula (2) modulo p^b.

4. Add the results and output the sum modulo p^b, which is the reduction of $N(f)$ modulo p^b.

References

[1] R. Beigel and J. Tarui, On ACC, Computational Complexity, 4(1994), 350-366.

[2] P. Candelas, X. de la Ossa and F. R. Villegas, Calabi-Yau manifolds over finite fields, I. http://xxx.lanl.gov/abs/hep-th/0012233.

[3] A. Ehrenfeucht and M. Karpinski, The computational complexity of counting problems, Tech. Rep. 8543-CS, ICSI, Berkeley, 1990.

[4] B. Gross and N. Koblitz, Gauss sums and the p-adic Γ-function, Ann. Math., 109(1979), 569-581.

[5] P. Gopalan, V. Guruswami and R.J. Lipton, Algorithms for modular counting of roots of multivariate polynomials, in LATIN 2006, 544-555.

[6] A. Lauder, Counting solutions in equations in mnay variables over finite fields, Found. Comp. Math., Vol. 4., No. 3, (2004), 221-267.

[7] A. Lauder and D. Wan, Counting points on varieties over finite fields of small characteristic, to appear in Algorithmic Number Theory, J.B. Buhler and P. Stevenhagen (eds), Cambridge University Press, 2007.

[8] S. Toda, PP is as hard as the polynomial-time hierarchy, SIAM Journal on Computing, 20(5) (1991), 865-877.

[9] J. von zur Gathen, M. Karpinski and I. Shparlinski, Counting curves and their projections, Computational Complexity, 6(1996), 64-99.

[10] D. Wan, Newton polygons and congruence decompositions of L-functions over finite fields, Contemporary Mathematics, 133(1992), 221-241.

[11] D. Wan, Computing zeta functions over finite fields, Contemporary Mathematics, 225(1999), 135-141.

[12] D. Wan, Algorithmic theory of zeta functions over finite fields, to appear in Algorithmic Number Theory, J.B. Buhler and P. Stevenhagen (eds), Cambridge University Press, 2006.

[13] D. Wan, Mirror symmetry for zeta functions, in Mirror Symmetry IV, 2006.

[14] A.C. Yao, On ACC and threshold circuits, in 31^{st} IEEE Symposium on Foundations of Computer Sciences (FOCS'90), 1990, 619-627.