Detecting Rational Points on Hypersurfaces over Finite Fields

Swastik Kopparty*
CSAIL, MIT
swastik@mit.edu

Sergey Yekhanin
IAS
yekhanin@ias.edu

Abstract

We study the complexity of deciding whether a given homogeneous multivariate polynomial has a nontrivial root over a finite field. Given a homogeneous algebraic circuit C that computes an n-variate polynomial $p(x)$ of degree d over a finite field $\mathbb{F}_q$, we wish to determine if there exists a nonzero $x \in \mathbb{F}_q^n$ with $C(x) = 0$.

For constant n there are known algorithms for doing this efficiently. However for linear n, the problem becomes NP hard. In this paper, using interesting algebraic techniques, we show that if d is prime and $n > d/2$, the problem can be solved over sufficiently large finite fields in randomized polynomial time. We complement this result by showing that relaxing any of these constraints makes the problem intractable again.

1 Introduction

Given a homogeneous polynomial $p(X)$ over $\mathbb{F}_q$ in n variables and degree d, consider the projective hypersurface of $\mathbb{F}_q$-rational points, $V_p = \{x \in \mathbb{PF}_q^n : p(x) = 0\}$, defined by $p(X)$. We wish to efficiently determine whether V_p is nonempty. This is equivalent to deciding whether there is an $x \in \mathbb{F}_q^n$, with $x \neq \mathbf{0}$, such that $p(x) = 0$.

Let us impose the following conditions on n, d and q:

- d is a prime,
- $d < 2n$
- $q \geq \Omega(n^4)$.

*Supported in part by NSF Award CCR-0514915

Our main result is that under these conditions, given black-box access to evaluations of p, we can decide whether V_p is nonempty in randomized time polynomial in n, d and $\log q$. Furthermore, relaxing any one of these conditions, makes the question NP hard. In particular, "2" cannot be replaced by "$2 + \epsilon$" for any $\epsilon > 0$, we cannot drop the condition that d is prime, nor can we allow q to be constant!

Given access to merely $\mathrm{poly}(n, d, \log q)$ evaluations of $p(X)$, our algorithm infers the existence of a nontrivial root of $p(X)$, without actually finding one. This inference is made by exploiting the rich algebraic structure of polynomials and their factorization patterns. Indeed, we use some powerful algebraic tools to prove these results. We believe that that our techniques are of independent interest and may find further applications in algebraic complexity.

1.1 Detecting Rational Points: History and Motivation

The problem of determining whether there is a rational point on a variety over a finite field is a natural and well studied problem. This problem has been implicitly studied in number theory for centuries, perhaps beginning with the quadratic reciprocity theorem of Gauss.

In [7], Huang and Wong gave a randomized polynomial time algorithm for detecting rational points on varieties for constant n. They further asked if one could find an algorithm for growing n, aware that the general case is NP complete. Our work studies this question for the case of projective hypersurfaces.

Detecting rational points on projective hypersurfaces also arises naturally in the problem (due

to Saks and Wigderson) of detecting nonsingular spaces of matrices. A *nonsingular* space of matrices is a linear subspace of $\mathbb{F}_q^{d\times d}$ such that all its nonzero elements are nonsingular. Given $A_1, \ldots, A_n$, a basis for a linear subspace L of $\mathbb{F}_q^{d\times d}$, consider the homogeneous degree d polynomial $p(X) = \det(\sum_{i=1}^n A_i X_i) \in \mathbb{F}_q[X_1, \ldots, X_n]$. It is easy to check that the projective hypersurface defined by $p(X)$ has no nontrivial $\mathbb{F}_q$ rational point iff L is a nonsingular space of matrices. Furthermore, for any $x \in \mathbb{F}_q^n$, there is a polynomial time algorithm to evaluate $p(x)$. Thus our main result implies that over sufficiently large fields, when d is prime and the dimension n of the space L is greater than $d/2$, there is a randomized polynomial time algorithm for detecting nonsingular spaces of matrices.

A number of problems on singularity, nonsingularity, minimum rank and maximum rank of spaces of matrices have been studied extensively (cf. [3]). The problem of detecting singular spaces of matrices was studied by Lovász [12], and is intimately connected with the problem of polynomial identity testing. We believe that our results shed some light on the structure and algorithmics of such problems in linear algebra.

There are several algorithms [1], [11], [14], [15] for *counting* rational points on projective hypersurfaces. The most powerful, an algorithm of Lauder [10] counts points on *smooth* hypersurfaces in time polynomial in $\log q$ and the number of nonzero coefficients, which may be as large as d^n in general.

Recently, Gopalan, Guruswami and Lipton [6] and Wan [17] considered the problem of counting rational roots of a polynomial *mod* m, for some integer m. They gave algorithms and hardness results for different choices of m.

1.2 Results

We will be considering homogeneous polynomials $p(X)$ in n variables and of degree d over $\mathbb{F}_q$. Our algorithmic result will only require black-box access to evaluations of $p(X)$ at points in $\mathbb{F}_q^n$. For the corresponding NP hardness results, we will assume that $p(X)$ is presented as a homogeneous algebraic circuit[1] $\mathcal{C}$ over $\mathbb{F}_q$. We assume that the field $\mathbb{F}_q$ is represented by an irreducible polynomial over a prime subfield. We are interested in the complexity of detecting nontrivial roots of $p(X)$ after imposing various constraints on n, d and q.

For a set of "constraints" $S \subseteq \mathbb{N}^3$, we define the language $L_{\mathrm{proj}}(S)$ to be the set of all 4-tuples $(\mathcal{C}, 1^n, 1^d, \mathbb{F}_q)$ such that

1. $\mathcal{C}$ is a homogeneous algebraic circuit of degree d over $\mathbb{F}_q$,
2. $(n, d, q) \in S$,
3. There exists $x \in \mathbb{F}_q^n$, $x \neq 0$ with $\mathcal{C}(x) = 0$.

We can now state our main algorithmic result.

Theorem 1.1 *Let $S \subseteq \mathbb{N}^3$ be the set of all (n, d, q) such that:*

1. *d is prime,*
2. *$d < 2n$,*
3. *$q \geq 32n^4$.*

Then $L_{\mathrm{proj}}(S)$ is in RP .

This result will be proved in Section 3. Our main technical result underlying the algorithm is a structure theorem for rational-point-free hypersurfaces over finite fields for the above setting of the parameters n, d and q. The randomized algorithm of this theorem "pretends" that the hypersurface is rational-point-free, attempts to recover the underlying structure guaranteed by the structure theorem, and finally verifies its attempt via a polynomial identity test. Any failure along the way indicates that the hypersurface must have a rational point.

We complement the above algorithmic result by showing that, in a certain sense, relaxing any of the constraints on S above makes $L_{\mathrm{proj}}(S)$ NP hard. We state these results below. They will be proved in Section 4.

Theorem 1.2 (Relaxing primality of the degree) *Let $S_1 \subseteq \mathbb{N}^3$ be the set of all (n, d, q) such that:*

1. *$d < 2n$,*
2. *$q \geq 32n^4$.*

Then $L_{\mathrm{proj}}(S_1)$ is NP *complete.*

Theorem 1.3 (Relaxing the degree bound) *Let $\epsilon > 0$ and let $S_2 \subseteq \mathbb{N}^3$ be the set of all (n, d, q) such that:*

[1] i.e., every intermediate gate computes a homogeneous polynomial

1. d *is a prime,*
2. $d < (2+\epsilon)n$*,*
3. $q \geq 32n^4$*.*

Then $L_{\mathrm{proj}}(S_2)$ *is* NP *complete.*

One may also ask if the condition requiring a growing field size in Theorem 1.1 is really necessary. Here we can only show a slightly weaker kind of hardness. Let $L_{\mathrm{proj}}^{\mathrm{weak}}(S)$ denote the set of $(\mathcal{C}, n, d, \mathbb{F}_q)$, where $\mathcal{C}$ is an algebraic circuit, such that there is a homogenous polynomial $p(X)$ of degree d in n variables over $\mathbb{F}_q$ such that $\mathcal{C}(x) = p(x)$ for each $x \in \mathbb{F}_q^n$ (i.e., we no longer require that $\mathcal{C}$ formally computes $p(X)$, merely that its evaluations agree with evaluations of $p(X)$ at points of $\mathbb{F}_q^n$). Clearly, the NP hardness of $L_{\mathrm{proj}}^{\mathrm{weak}}(S)$ still shows the computational intractibility of detecting nontrivial roots of $p(X)$, given blackbox access to its evaluations in $\mathbb{F}_q^n$ (for $(n, d, q) \in S$).

Theorem 1.4 (Relaxing the growing field size) *Let* q_0 *be a prime power, and let* $S_3 \subseteq \mathbb{N}^3$ *be the set of all* (n, d, q) *such that:*

1. d *is a prime,*
2. $d < 2n$*,*
3. $q = q_0$*.*

Then $L_{\mathrm{proj}}^{\mathrm{weak}}(S_3)$ *is* NP *hard.*

Organization of this paper: In Section 2, we introduce the algebraic tools that we will use. In Section 3 we prove our main technical lemma, and use it to prove Theorem 1.1 . Finally, in Section 4 we prove the complementary hardness results: Theorem 1.2 and Theorem 1.3.

2 Algebraic Preliminaries

In this section, we will introduce some algebraic preliminaries necessary for the proofs of our results.

2.1 Norm polynomials

We give a brief overview of norm polynomials and some of their many interesting properties. They indirectly inspired our algorithm and are crucial to our hardness results.

Let $\mathbb{F}_q$ be a finite field. For any positive integer n, a norm polynomial is a homogeneous polynomial $N_n \in \mathbb{F}_q[x_1, \ldots, x_n]$ with the following properties:

- $\deg(N_n) = n$,
- For any $(c_1, \ldots, c_n) \in \mathbb{F}_q^n$, if $N_n(c_1, \ldots, c_n) = 0$ then $c_1 = \ldots = c_n = 0$.
- An algebraic circuit computing N_n can be generated in time polynomial in n and q.

Following ([13] p.272), we now give a construction of norm polynomials N_n. Let $\{\alpha_1, \ldots, \alpha_n\}$ be a basis of $\mathbb{F}_{q^n}$ over $\mathbb{F}_q$. Put

$$N_n(x_1, \ldots, x_n) = \prod_{j=0}^{n-1} (\alpha_1^{q^j} x_1 + \ldots + \alpha_n^{q^j} x_n). \quad (1)$$

Since the $\{\alpha_i^{q^j}\}_{j \in \{0,1,\ldots,n-1\}}$ are conjugates of α_i over $\mathbb{F}_q$ (i.e., they are the roots of the minimal polynomial of α_i over $\mathbb{F}_q$), the coefficients of N_n are in $\mathbb{F}_q$. It is clear that $\deg(N_n) = n$ and that N_n is a homogeneous polynomial. Now let $(c_1, \ldots, c_n) \in \mathbb{F}_q^n$ and put $\gamma = c_1\alpha_1 + \ldots + c_n\alpha_n \in \mathbb{F}_{q^n}$. Then

$$N_n(c_1, \ldots, c_n) = \prod_{j=0}^{n-1} (\alpha_1^{q^j} c_1 + \ldots + \alpha_n^{q^j} c_n)$$

$$= \prod_{j=0}^{n-1} (\alpha_1 c_1 + \ldots + \alpha_n c_n)^{q^j} = \mathrm{N}_{\mathbb{F}_{q^n}/\mathbb{F}_q}(\gamma),$$

where $\mathrm{N}_{\mathbb{F}_{q^n}/\mathbb{F}_q}$ denotes the field norm from $\mathbb{F}_{q^n}$ to $\mathbb{F}_q$. Thus $N_n(c_1, \ldots, c_n) = 0$ is equivalent to $\mathrm{N}_{\mathbb{F}_{q^n}/\mathbb{F}_q}(\gamma) = 0$, which holds only for $\gamma = 0$, i.e., only when $c_1 = \ldots = c_n = 0$.

It remains to note that one can generate irreducible polynomials of arbitrary degree n in time polynomial in n and q over arbitrary finite fields [2], [16]. Moreover over fields of constant characteristic this can be achieved in running time polynomial in n and $\log q$. Thus given n and $\mathbb{F}_q$ one can efficiently generate an algebraic circuit computing a norm polynomial of degree n in two steps:

- Generate an irreducible polynomial $f(y) \in \mathbb{F}_q[y]$ of degree n. Set $\alpha_1 = 1, \alpha_2 = y, \ldots, \alpha_n = y^{n-1}$.
- Generate an algebraic circuit that on input $x_1, \ldots, x_n \in \mathbb{F}_q^n$ computes the expression in formula (1) modulo $f(y)$.

2.2 Rational points and factorization

The algebraic closure of $\mathbb{F}_q$ is denoted $\overline{\mathbb{F}}_q$. A polynomial $p(X) \in \mathbb{F}_q[X_1, \ldots, X_n]$ is *absolutely irreducible* if it is irreducible in $\overline{\mathbb{F}}_q[X_1, \ldots, X_n]$.

For $r \in \mathbb{N}$, let $Gal(\mathbb{F}_{q^r}/\mathbb{F}_q)$ be the Galois group of $\mathbb{F}_{q^r}$ over $\mathbb{F}_q$. This group is generated by the $\mathbb{F}_q$-automorphism of $\mathbb{F}_{q^r}$ that maps $x \to x^q$. For a polynomial $p(X) = \sum p_i X_1^{a_{i1}} X_2^{a_{i2}} \cdots X_n^{a_{in}} \in \mathbb{F}_{q^r}[X_1, \ldots, X_n]$ and $\sigma \in Gal(\mathbb{F}_{q^r}/\mathbb{F}_q)$, define $\sigma(p(X)) = \sum \sigma(p_i) X_1^{a_{i1}} X_2^{a_{i2}} \cdots X_n^{a_{in}}$. We say $\sigma(p(X))$ is a conjugate of $p(X)$.

We will appeal to several powerful theorems on rational points during the course of our proof. We list them below.

The theorem of Lang and Weil gives an estimate for the number of $\mathbb{F}_q$-rational points on any absolutely irreducible $\mathbb{F}_q$-hypersurface in terms of its degree. The following refinement (due to Cafure-Matera, relying heavily on Weil's theorem and results of Kaltofen) gives a sufficient condition for an absolutely irreducible polynomial to have several roots.

Theorem 2.1 (Cafure-Matera [4]) *Suppose $p(X) \in \mathbb{F}_q[X_1, \ldots, X_n]$ is a degree d absolutely irreducible polynomial. Then for any positive constant c, there is a degree bound $d_0(c)$, such that if $d \geq d_0(c)$ and $q > 2d^4$, then*

$$\left|\{x \in \mathbb{F}_q^n : p(x) = 0\}\right| \geq c.$$

Tracing the dependence of d_0 on c in their proof, one can check that $d_0(2)$ may be taken to be 1.

The Chevalley-Warning theorem gives another sufficient condition for a polynomial to have more than one root. It implies that if a polynomial has low enough degree and has at least one root, then it has more than one root.

Theorem 2.2 (Chevalley-Warning [13]) *Suppose $p(X) \in \mathbb{F}_q[X_1, \ldots, X_n]$ is a degree d polynomial. If $d < n$, then*

$$\left|\{x \in \mathbb{F}_q^n : p(x) = 0\}\right| \equiv 0 \mod char(\mathbb{F}_q)$$

If $d = n$, there are examples of polynomials that have only one root. Indeed, norm polynomials have this property. Our main lemma implies that, for some n, q, the norm polynomials are the *only* such examples.

3 The Algorithm

In this section we prove Theorem 1.1.

3.1 The Main Lemma

Given a homogeneous polynomial $p(X)$ in n variables of degree d over $\mathbb{F}_q$, recall that we wish to decide whether V_p is nonempty or empty; equivalently whether $p(X)$ has a nontrivial $\mathbb{F}_q$ root or not. In this subsection we prove our main technical result, Lemma 3.2, which gives structural information about any polynomial that has no nontrivial $\mathbb{F}_q$ root, under certain conditions on n, q and d.

We begin with a preliminary lemma about factorization of $\mathbb{F}_q$-irreducible polynomials over $\overline{\mathbb{F}}_q$.

Lemma 3.1 *Suppose $p(X) \in \mathbb{F}_q[X_1, \ldots, X_n]$ is of degree d and is irreducible in $\mathbb{F}_q[X_1, \ldots, X_n]$. Then there exists r with $r|d$ and an absolutely irreducible polynomial $h(X) \in \mathbb{F}_{q^r}[X_1, \ldots, X_n]$ of degree d/r such that*

$$p(X) = c \prod_{\sigma \in G} \sigma(h(X))$$

where $G = Gal(\mathbb{F}_{q^r}/\mathbb{F}_q)$ and $c \in \mathbb{F}_q$. Furthermore, if $p(X)$ is homogeneous, then so is $h(X)$.

Proof Let $h(X)$ be an absolutely irreducible factor of $p(X)$, scaled so that one of its nonzero coefficients is in $\mathbb{F}_q$. Let r be the smallest integer such that the coefficients of h lie in $\mathbb{F}_{q^r}$. Furthermore, for any $\sigma \in Gal(\mathbb{F}_{q^r}/\mathbb{F}_q)$, $\sigma(h(X))|\sigma(p(X)) = p(X)$. Thus all conjugates of $h(X)$ are also factors of $p(X)$.

For $\sigma \in Gal(\mathbb{F}_{q^r}/\mathbb{F}_q), \sigma \neq$ identity, we claim that $h(X)$ and $\sigma(h(X))$ are relatively prime. Indeed, suppose $h(X)$ and $\sigma(h(X))$ have a common factor in $\overline{\mathbb{F}}_q[X_1, \ldots, X_n]$. By the absolute irreducibility of $h(X)$, this means that $\sigma(h(X))$ must be a scalar multiple of $h(X)$. This means that for any two nonzero coefficients $\beta, \gamma \in \mathbb{F}_{q^r}$, $\sigma(\beta)/\beta = \sigma(\gamma)/\gamma$, and so $\beta/\gamma \in \mathbb{F}_{q^r}^{\sigma}$ (the subfield of $\mathbb{F}_{q^r}$ fixed by σ). Thus, by the initial scaling of $h(X)$, every coefficient of $h(X)$ lies in $\mathbb{F}_{q^r}^{\sigma}$. By definition of r, this means that $\mathbb{F}_{q^r} = \mathbb{F}_{q^r}^{\sigma}$, which implies that $\sigma =$ identity, a contradiction.

We can now conclude that $\prod_{\sigma \in G} \sigma(h(X)) | p(X)$. However $\prod_{\sigma \in G} \sigma(h(X)) \in \mathbb{F}_q[X_1, \ldots, X_n]$ and p is irreducible. The result follows. ■

We can now state and prove our main lemma.

Lemma 3.2 *Suppose d is prime, $d < 2n$ and $q > 32n^4$. Let $p(X) \in \mathbb{F}_q[X_1, \ldots, X_n]$ be a homogeneous degree d polynomial in n variables with coefficients in $\mathbb{F}_q$ with no nontrivial $\mathbb{F}_q$-rational points. Then there exists a homogeneous degree 1 polynomial $h(X) \in \mathbb{F}_{q^d}[X_1, \ldots, X_n]$ such that:*

$$p(X) = c \prod_{\sigma \in G} \sigma(h(X))$$

where $G = Gal(\mathbb{F}_{q^d}/\mathbb{F}_q)$ and $c \in \mathbb{F}_q$.

Proof Note that there is one trivial rational point (the origin). To prove our theorem, we will infer facts about the factorization of $p(X)$ using the hypothesis (i.e.,the absence of another rational point).

1. $p(X)$ **is irreducible over** $\mathbb{F}_q[X_1, \ldots, X_n]$**:** If $p(X) = g_1(X)g_2(X)$, where $g_i(X) \in \mathbb{F}_q[X_1, \ldots, X_n]$ are of positive degree, then they are both homogeneous polynomials and at least one of them has degree $< n$. This polynomial has a nontrivial rational point by Theorem 2.2, and thus so does p, a contradiction.
2. Therefore, by Lemma 3.1, for some $r|d$, $p(X)$ factors as a product of r conjugates, each of degree d/r. However d is prime, and so $r = 1$ or $r = d$. $r = 1$ corresponds to $p(X)$ being absolutely irreducible.
3. $p(X)$ **is not absolutely irreducible:** If $p(X)$ was absolutely irreducible, we would have a contradiction to Theorem 2.1 with $c = 2$.

Thus $r = d$ and the result follows.

In fact, we can say something further about $h(X)$. Suppose $h(X) = \sum_{i=1}^{n} a_i X_i$. Then, by the absence of nontrivial $\mathbb{F}_q$ rational points, $\{a_1, \ldots, a_n\}$ are linearly independent over $\mathbb{F}_q$. Conversely, if the a_i are independent, then $p(X)$ has no nontrivial roots. ■

The lemma says that if $p(X)$ has no rational points, then $p(X)$ is a product of conjugate degree 1 polynomials. This lays out a natural plan of attack for our algorithm. The algorithm first pretends that $p(X)$ has no rational point, and through some judicious substitutions[2], determines the coefficients of the degree 1 factors that $p(X)$ is supposed to factor into. Then, the algorithm will verify that $p(X)$ is indeed the product of these conjugate degree 1 polynomials via a randomized polynomial identity test.

If $p(X)$ has no rational point, then the substitutions will go through successfully, we will correctly determine the coefficients of the linear factors, and the polynomial identity test will pass. If the polynomial identity test passes with high probability, then $p(X)$ is the product of some known degree 1 polynomials, and hence we can directly check if it has a rational point.

3.2 Substitutions

Suppose $p(X)$ has no nontrivial root. The lemma above tells us that we may write $p(X) = c \prod_{\sigma \in G} \sigma(h(X))$, where $G = Gal(\mathbb{F}_{q^d}/\mathbb{F}_q)$ Let $h(X) = \sum_{i=1}^{n} a_i X_i$. Indeed, there are many representations of $p(X)$ in this form (an $\mathbb{F}_q$-multiple of a product of degree 1 conjugate polynomials over $\mathbb{F}_q^d$), and the purpose of the first phase of our algorithm will be to determine one such representation (c,$h(X)$).

In our representation, we can assume $a_1 = 1$ (possibly changing c). The algorithm will rely on the following observations.

- $p(1, 0, 0, \ldots, 0) = c$
- $p(1, X_2, 0, 0, \ldots, 0) = c \prod_{\sigma \in G} \sigma\left(1 + a_2 X_2\right) = c \prod_{\sigma \in G} \left(1 + \sigma(a_2) X_2\right)$
- $p(1, X_2, 0, \ldots, 0, X_k, 0, \ldots, 0) = c \prod_{\sigma \in G} \sigma\left(1 + a_2 X_2 + a_k X_k\right) = \prod_{\sigma \in G} \left(1 + \sigma(a_2) X_2 + \sigma(a_k) X_k\right)$

As remarked earlier, we will only require black-box access to evaluations of the polynomial. In this model, note that we can efficiently compute $p(1, T, 0, \ldots, 0, S, 0, \ldots, 0)$ (where T and S are indeterminates) by substituting values in $\mathbb{F}_q$ for T and S and then interpolating.

[2]Indeed, at this point one could use Kaltofen factorization, but this will introduce two-sided error

3.3 The Algorithm

The algorithm has two phases. The first phase of the algorithm will attempt to recover a representation $(c, h(X))$ of $p(X)$, as suggested in the previous subsection. We note that univariate and bivariate polynomial factorization and the generation of an explicit description of field $\mathbb{F}_{q^d}$ can all be done in randomized polynomial time (with zero error) [2, 8]. In our algorithm, if any of these procedures fail, the algorithm REJECTs.

We describe the first phase of the algorithm below (here b_i will be our guess for a_i):

Phase 1

1. Compute $c = p(1, 0, \ldots, 0)$. If $c = 0$, ACCEPT
2. Set $b_1 = 1$
3. Compute $g(T) = p(1, T, 0, 0, \ldots, 0)$, where T is an indeterminate. If $g(T) = 0$, ACCEPT
4. Factor $g(T)$ over $\mathbb{F}_{q^d}$ as $c\prod_{j=1}^{d}(1 + \beta_j T)$.
5. Set $b_2 = \beta_1$ (say). If the β_j are not all the distinct conjugates of β_1 over $\mathbb{F}_{q^d}$, ACCEPT
6. For each $k \in \{3, 4, \ldots, n\}$, do the following:
   - Compute $g_k(T, S) = p(1, T, 0, \ldots, 0, S, 0, 0)$, where T and S are indeterminates, and S is substituted into the k^{th} input variable. If $g_k(T, S) = 0$, ACCEPT
   - Factor $g_k(T, S)$ over $\mathbb{F}_{q^d}$ in the form $\prod_{j=1}^{d}(1 + \beta_j T + \gamma_{k,j} S)$ (if possible). If factorization into this form is not possible, ACCEPT
   - Set $b_k = \gamma_{k,1}$
7. Store $(c, \sum_{i=1}^{n} b_i X_i)$ for use in the second phase

In the second phase, the algorithm harvests the information gathered in the first. Using the purported representation of $p(X)$, it verifies that it is correct via an identity test.

Phase 2

1. If the b_i are linearly dependent over $\mathbb{F}_q$, ACCEPT
2. Perform a Randomized Identity Test (with failure probability at most $1/2$) for the identity

$$\text{“}p(X) = c \prod_{\sigma \in G} \sigma(\sum_{i=1}^{n} b_i X_i)\text{”}$$

   - If they are not equal ACCEPT
   - Otherwise REJECT

The following claim completes the proof of Theorem 1.1.

Claim 3.3 *If $p(X)$ has no nontrivial root, then the above algorithm REJECTs with probability 1. If $p(X)$ has a nontrivial root, then the above algorithm ACCEPTs with probability at least $1/2$.*

Proof We consider two cases.

- **Suppose $p(X)$ has no nontrivial root:** Then $p(X)$ is of the form $c\prod_{\sigma \in G} \sigma(h(X))$ where $h(X) = \sum_{i=1}^{n} a_i X_i$, for some $\mathbb{F}_q$-linearly-independent $a_i \in \mathbb{F}_{q^d}$ with $a_1 = 1$. The algorithm will proceed “as planned”; in particular it will not ACCEPT during the first phase. By the observations made earlier, b_2 will be set to $\tau(a_2)$ for some $\tau \in G$, and furthermore for each k, b_k will be set to $\tau(a_k)$. Thus $\sum_{i=1}^{n} b_i X_i = \tau(\sum_{i=1}^{n} a_i X_i)$ and so $p(X) = c\prod_{\sigma \in G} \sigma(\sum_{i=1}^{n} b_i X_i)$. Therefore the b_i are linearly independent over $\mathbb{F}_q$, and the algorithm always REJECTs in the second phase.

- **Suppose $p(X)$ has nontrivial roots:** Now we do not care about what happens in the first phase. If we do not make it to the final identity test, then we must have ACCEPTed during some earlier step. So let us assume we have reached the identity test; in particular, the b_i are linearly independent over $\mathbb{F}_q$. In the final test of $p(X)$ against $c\prod_{\sigma \in G} \sigma(\sum b_i X_i)$, observe that the second polynomial has no nontrivial rational points, while the first does, and hence with probability at least $1/2$ the polynomials will be exposed as unequal by the identity test, and the algorithm will ACCEPT.

This completes the proof of the claim, and hence of Theorem 1.1. ■

3.4 Finding roots

Given a homogeneous degree d polynomial in n variables over $\mathbb{F}_q$, one could also consider the algorithmic question of finding a nontrivial root (if any). The Chevalley-Warning theorem guarantees that every d-dimensional projective hyperplane contains a rational point. Our algorithm allows us to identify a projective hyperplane of dimension at most $d/2$ with a rational point in time polynomial in n, d, q, when d is prime and $q > 32n^4$. Indeed, using the detection algorithm one can successively find smaller dimensional projective hyperplanes containing a rational point until the dimension becomes at most $d/2$.

4 Hardness results

In this section we establish Theorem 1.2 and Theorem 1.3, that show that relaxing the conditions in Theorem 1.1 makes L_{proj} NP complete.

4.1 Proofs of hardness results

Proof of Theorem 1.2:

Consider the language L of 3DNF formulae F for which there exists an assignment x, such that $F(x) = 0$ and $x \neq 0$. It is easy to see that language L is NP complete. We will now give a reduction from L to $L_{\mathrm{proj}}(S_1)$.

Let F be a 3-DNF formula of length l in variables $x_1, \ldots, x_m$. Assume F has t clauses. Clearly, $m \leq l$ and $t \leq l$. Fix some integer $s \geq 2$. Fix a finite field $\mathbb{F}_q$ of size at least $100s^4l^8$. This is the right field size to meet the requirements of the theorem. However one can carry out our reduction over arbitrary finite fields independent of their size.

In what follows we demonstrate an efficient procedure that given F generates an algebraic circuit over $\mathbb{F}_q$ computing a homogeneous polynomial p_F in $n = m + m(s-1)(3t+1)$ variables such that the following holds,

$$\exists\, x \in \{0,1\}^m \text{ such that } F(x) = 0 \text{ and } x \neq 0 \tag{2}$$

$$\Leftrightarrow \exists\, x \in \mathbb{F}_q^n, \text{ such that } p_F(x) = 0 \text{ and } x \neq 0. \tag{3}$$

First for every clause $c_i = x_{i_1}^{\sigma_1} \wedge x_{i_2}^{\sigma_2} \wedge x_{i_3}^{\sigma_3}$ construct a polynomial $w_{c_i}(x_{i_1}, x_{i_2}, x_{i_3}) = l_{\sigma_1}(x_{i_1}) l_{\sigma_2}(x_{i_2}) l_{\sigma_3}(x_{i_3})$, where $l_\sigma(x)$ is a linear form defined by $l_0(x) = x$ and $l_1(x) = 1 - x$. Our first attempt for the polynomial $p_F(x)$ would be to define

$$\hat{p}_F(x) = N_t(w_{c_1}, \ldots, w_{c_t}), \tag{4}$$

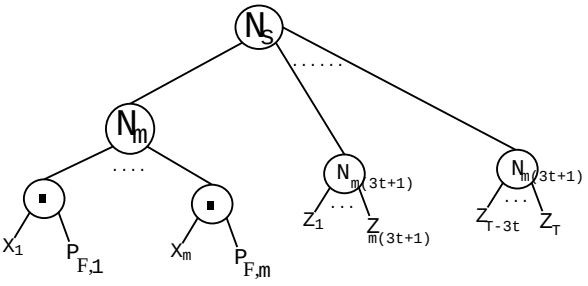
where N_t is a norm polynomial. One can verify that the polynomial $\hat{p}_F(x)$ defined above has a root in $\mathbb{F}_q$ if and only if there exists an $x \in \{0,1\}^n$ such that $F(x) = 0$. The obvious problem with this reduction is that $\hat{p}_F(x)$ need not be a homogeneous polynomial.

We go around this problem by defining homogenizing the polynomial $\hat{p}_F$, once with respect to each input variable x_j. Introduce new polynomials $w_{c_i,j}(x_{i_1}, x_{i_2}, x_{i_3}) = l_{\sigma_1,j}(x_{i_1}) l_{\sigma_2,j}(x_{i_2}) l_{\sigma_3,j}(x_{i_3})$, where $l_{\sigma,j}(x)$ is a homogeneous linear form defined by $l_{0,j}(x) = x$ and $l_{1,j}(x) = x_j - x$. We also define polynomials

$$p_{F,j}(x) = N_t(w_{c_1,j}, \ldots, w_{c_t,j}), \tag{5}$$

for all $j \in \{1, 2, \ldots, m\}$. Essentially, the polynomial $p_{F,j}(x)$ is obtained by homogenization of the polynomial $\hat{p}_F(x)$ defined by (4), where variable x_j is used to carry out the homogenization procedure. Polynomials $p_{F,j}$ are homogeneous. However they may have roots that do not correspond to boolean assignments to F that set it to zero.

The picture below represents the structure of our final polynomial p_F. p_F is a polynomial in $n = m + T$ variables, where $T = m(s-1)(3t+1)$. Of these, m variables are just $x_1, \ldots, x_m$ and the other T variables are labelled $z_1, \ldots, z_T$. We need z_i's to achieve the right ratio between the number of variables and the degree of the polynomial.



The general idea behind our construction is the following. We combine the products

$x_j p_{F,j}(x_1, \ldots, x_m)$ using the norm polynomial N_m. This polynomial already has the property (2). The only problem that we still have at this point is that the degree of the norm of products is $d = m(3t+1)$ while the number of variables is only m. So we are quite far from the the desired relation $d < 2m$. To resolve this problem we add yet another norm polynomial N_s on top of N_n. Other arguments of N_s are norms in new variables $z_1, \ldots, z_T$.

Note that for every norm gate in the p_F the arguments are polynomials of the same degrees. Thus p_F is a homogeneous polynomial. p_F is a polynomial in $n = m + m(s-1)(3t+1)$ variables of degree $d = sn(3t+1)$. Therefore we have the relation

$$d < ((s)/(s-1))n. \tag{6}$$

Recall that p_F is a polynomial over $\mathbb{F}_q$ where $q \geq 100 s^4 l^8$. Thus we also have the relation $q \geq 32n^4$. It remains to verify that the property (2) does hold.

Assume we are given a nonzero boolean assignment $a \in \{0,1\}^m$ such that $F(a) = 0$. We set all variables $z_1, \ldots, z_T$ to zero. We shall now verify that $p_F(a_1, \ldots, a_m, 0, \ldots, 0) = 0$. First note that for every $j \in \{1, 2, \ldots, n\}$ the product $a_j p_{F,j}(a_1, \ldots, a_m) = 0$. This happens either because $a_j = 0$ or because $a_j = 1$ and $p_{F,j}(a_1, \ldots, a_m)$ computes the value of $F(a)$. Further recall that the norm polynomials on zero inputs evaluate to zero. Therefore $(a_1, \ldots, a_n, 0, \ldots, 0)$ is a nontrivial zero of p_F.

The other direction is slightly more complicated. Let $(a_1, \ldots, a_m, v_1, \ldots, v_T) \in \mathbb{F}_q^m$ be a nontrivial zero of p_F. Note that v_i needs to be zero for all $i \in \{1, \ldots, T\}$, since otherwise the value of N_s will be nonzero. Therefore there should be a non-zero value among $(a_1, \ldots, a_m)$. Without loss of generality assume that $a_1 \neq 0$. Observe the following chain of implications:

$$N_s = 0 \;\Rightarrow\; N_m = 0 \;\Rightarrow\; p_{F,1}(a_1, \ldots, a_m) = 0. \tag{7}$$

For all $i = 1, n$ set $b_i = 0$ if $a_i = 0$ and set $b_i = 1$ otherwise. Observe that $b_1 = 1$. The structure of polynomials $w_{c_i,1}$ implies that $p_{F,1}(b_1, \ldots, b_m) = 0$. It remains to notice that $b_1 = 1$ and $p_{F,1}(b_1, \ldots, b_m) = 0$ imply that $F(b_1, \ldots, b_m) = 0$. ■

Remark The above proof shows that we may even add the additional constraint $d < (1+\epsilon)n$ for any $\epsilon > 0$ to the definition of S_1. Indeed, we may choose s to be a suitably large constant so that (6) gives us the desired bound on d.

Proof of Theorem 1.3: We exhibit a reduction from the same NP complete language as in the proof of theorem 1.2. Given a 3DNF formula F of length l we follow the procedure of theorem 1.2 to construct an algebraic circuit computing a polynomial $\hat{p}(x_1, \ldots, x_n)$ of degree d over a field of size at least $32n^4$ such that the condition (2) is satisfied. We also require $d \leq (1 + \epsilon/2)n$. The possibility of such a construction is implied by the above remark.

Next we choose a prime r between $(2 + \epsilon/2)n$ and $(2+\epsilon)n$. Note that the existence of such a prime for sufficiently large n follows from the prime number theorem [5]. We can find r by brute force search since we are allowed to run in time polynomial in n. Put $g = r - d$. Note that $g \geq n$. We now define our final polynomial to be

$$p_F(x_1, \ldots, x_n) = \hat{p}_F(x_1, \ldots, x_n) \cdot N_g(x_1, \ldots, x_n, x_1, \ldots, x_1).$$

The first n arguments of N_g are $x_1, \ldots, x_n$ and all other arguments are x_1. One can easily see that a circuit for p_F can be constructed in polynomial time. p_F has prime degree r and the required relations between n and r are satisfied. Proving the condition (2) is also easy. It suffices to show that every nontrivial root of p_F is also a nontrivial root of $\hat{p}_F$. This follows from the fact that N_g depends on all the variables $x_1, \ldots, x_n$ and always evaluates to nonzero at nonzero inputs. ■

Proof of Theorem 1.4: Again, our reduction is from the same NP complete language as in the previous proofs. We fix some value of $\epsilon > 0$. Given a formula F we use the construction from the remark after the proof of Theorem 1.2 to get an algebraic circuit that evaluates a homogeneous polynomial $\hat{p}_F(x_1, \ldots, x_n)$ such that $\hat{p}_F$ satisfies the property (2). Moreover $\deg \hat{p}_F \leq (1+\epsilon)n$. Also, we choose the field size q to be q_0.

The polynomial $\hat{p}_F$ has all the properties that we want except that the degree of $\hat{p}_F$ is not prime. Our idea to go around this problem is the following. We will prove that there exists a polynomial $p_F(x_1, \ldots, x_n)$ of prime degree such that p_F and the $\hat{p}_F$ are identical over $\mathbb{F}_q$. Then we can can use

the circuit that computes $\hat{p}_F$ to compute p_F and our reduction will be complete. Note that the further argument is aimed only to prove the existence of p_F of prime degree and we do not need to construct another circuit (recall that we are showing hardness of $L^{\mathrm{weak}}_{\mathrm{proj}}(S_3)$, not $L_{\mathrm{proj}}(S_3)$).

For a later part of the argument, we need d to be relatively prime to q. Recall that $d = ms(3t+1)$, where m, s and t are defined in the proof of theorem 1.2. Closer look at that proof shows that we can increase each of these quantities by one without affecting the reduction. In particular,

- In order to increase m by one just add a new variable y to F and consider a new formula $F \vee (y \wedge y \wedge y)$.
- Increasing s is trivial since s is just a parameter of the reduction and it can set to be arbitrarily large.
- It order to increase t one can just duplicate some clause of F.

Applying the tricks from above one can force $d = ms(3t+1)$ to be relatively prime to q. The simple way to increase the degree of a polynomial over a finite field of size q is to increase the degree of one variable in each monomial from b to $b+q$. Clearly, this does not affect the values of polynomial over $\mathbb{F}_q$ but increases the degree by q.

Therefore to complete the reduction we only need to show that assuming $d \leq (1+\epsilon)n$ an arithmetic progression of the form $\{d + qk\}_{k \geq 1}$, contains a prime that is less than $2n$. Assuming that n is sufficiently large this follows from the known facts about the distribution of primes in arithmetic progressions [5] [3]. ■

5 Conclusions

We studied the complexity of deciding whether a given multivariate polynomial has a root over a finite field. In this paper we found a randomized polynomial time algorithm for solving this problem given black-box access in a setting amenable to interesting algebraic techniques. We hope the techniques find wider applicability.

[3] We may actually choose q to be slightly superconstant

Acknowledgement

We would like to thank Avi Wigderson for introducing us to the problem of detecting nonsingular spaces of matrices, and many helpful discussions during this work.

References

[1] L. Adleman and M. Huang, Counting points on curves and Abelian varieties over finite fields. *Journal of Symbolic Computation*, v. 32 n. 3, pp. 171-189, 2001.

[2] E.R. Berlekamp, Factoring polynomails over large finite fields, *Mathematics of Computation*, pp. 713-735, 1970.

[3] J.F.Buss, G.S. Frandsen, J.O.Shallit, The Computational Complexity of Some Problems of Linear Algebra, *Journal of Computer and Systems Sciences*, v. 58, pp. 572-596, 1999.

[4] A. Cafure, G. Matera, Improved explicit estimates on the number of solutions of equations over a finite field, *http://arxiv.org*

[5] H. Davenport, Multiplicative number theory. Springer, 2000.

[6] P. Gopalan, V. Guruswami, R. Lipton, Algorithms for Modular Counting of Roots of Multivariate Polynomials, *LATIN*, 2006.

[7] M. Huang, Y. Wong, Solving Systems of Polynomial Congruences Modulo a Large Prime, *FOCS*, 1996.

[8] Kaltofen, E. A polynomial-time reduction from bivariate to univariate integral polynomial factorization, *FOCS*, 1982.

[9] K. Kedlaya, Counting points on hyperelleiptic curves using Monsky-Washintzer cohomology, *Journal of the Ramanujan Mathematical Society*, v. 16, pp. 323-338, 2001.

[10] A. Lauder, Counting solutions to equations in many variables over finite fields. *Foundations of Computational Mathematics*, vol. 4 n. 3, pp. 221-267, 2004.

[11] A. Lauder and D. Wan, Counting points on varieties over finite fields of small characteristic, "Algorithmic Number Theory: Lattices, Number Fields, Curves and Cryptography" (Mathematical Sciences Research Institute Publications), J. P. Buhler and P. Stevenhagen (eds.), Cambridge University Press.

[12] L. Lovasz, Singular spaces of matrices and their applications in combinatorics, *Bol. Soc. Bras. Mat.* v. 20, pp. 87-99, 1989.

[13] R. Lidl and H. Niederreiter, Finite Fields. Cambridge: Cambridge University Press, 1985.

[14] J. Pila, Frobenius maps of abelian varieties and finding roots of unity in finite fields, *Mathematics of Computation* v. 55, pp. 745-763, 1990.

[15] R. Schoof, Elliptic curves over finite fields and the computation of square roots mod p. *Mathematics of Computation*, v. 44, n. 170, pp. 483-494, 1985.

[16] V. Shoup, Removing randomness from the computational number theory, Ph.D. Theses, University of Wisconsin-Madison, 1989.

[17] D. Wan, Modular counting of rational points over finite fields, *Foundations of Computational Mathematics*, to appear.