Improved low-degree testing and its applications

Sanjeev Arora[*] Madhu Sudan[†]
Princeton University MIT

August, 1, 2001

Abstract

NP = PCP$(\log n, 1)$ and related results crucially depend upon the close connection between the probability with which a function passes a *low degree test* and the distance of this function to the nearest degree d polynomial. In this paper we study a test proposed by Rubinfeld and Sudan [30]. The strongest previously known connection for this test states that a function passes the test with probability δ for some $\delta > 7/8$ iff the function has agreement $\approx \delta$ with a polynomial of degree d. We present a new, and surprisingly strong, analysis which shows that the preceding statement is true for arbitrarily small δ, provided the field size is polynomially larger than d/δ. The analysis uses a version of *Hilbert irreducibility*, a tool of algebraic geometry.

As a consequence we obtain an alternate construction for the following proof system: A constant prover 1-round proof system for NP languages in which the verifier uses $O(\log n)$ random bits, receives answers of size $O(\log n)$ bits, and has an error probability of at most $2^{-\log^{1-\epsilon} n}$. Such a proof system, which implies the NP-hardness of approximating Set Cover to within $\Omega(\log n)$ factors, has already been obtained by Raz and Safra [29]. Raz and Safra obtain their result by giving a strong analysis, in the sense described above, of a new low-degree test that they present.

A second consequence of our analysis is a self tester/corrector for any buggy program that (supposedly) computes a polynomial over a finite field. If the program is correct only on δ fraction of inputs where $\delta = 1/|\mathrm{F}|^{\epsilon} \ll 0.5$, then the tester/corrector determines δ and generates $O(\frac{1}{\delta})$ values for every input, such that one of them is the correct output. In fact, our results yield something stronger: Given the buggy program, we can construct $O(\frac{1}{\delta})$ randomized programs such that one of them is correct on every input, with high probability. Such a strong self-corrector is a useful tool in complexity theory - with some applications known.

[*] `arora@cs.princeton.edu`. Supported by an NSF CAREER award, an Alfred P. Sloan Fellowship, and a Packard Fellowship.

[†] `madhu@lcs.mit.edu.` Laboratory for Computer Science, MIT, Cambridge, MA 02139. Part of this work was done when this author was at the IBM Thomas J. Watson Research Center.

1 Introduction

The use of algebraic techniques has recently led to new (probabilistic) characterizations of traditional complexity classes. These characterizations involve an interaction between an untrustworthy prover (or many provers) and a probabilistic polynomial-time verifier. In MIP= NEXPTIME [7], and NP = PCP($\log n, 1$) [6, 5] the verifier has to probabilistically verify the satisfiability of a boolean formula by reading very few bits in a "proof string" presented by a prover. In IP=PSPACE [25, 32] the verifier has to probabilistically verify tautologyhood of a quantified boolean formulae by interacting with a prover. All these results fundamentally rely on the same idea: the verifier first *arithmetizes* (or *algebraizes*) the boolean formula, which involves viewing a boolean assignment not as a sequence of bits but as values of a polynomial [25]. From then on, verifying satisfiability or tautologyhood involves verifying — using some efficient algebraic procedures — specific properties of a polynomial that has been provided by the prover.

In this paper we present an improved analysis of the *low degree test*, an algebraic procedure used in the result NP=PCP($\log n, 1$). We expect this result to have many applications; some are already known. For example, the new analysis is known to lead to new characterizations of NP in terms of PCP, which in turn lead to improved results about the hardness of approximation. Recall that NP=PCP($\log n, 1$) implies the hardness of computing approximate solutions to many optimization problems such as CLIQUE [13, 6], CHROMATIC NUMBER and SET COVER [26], and MAX-3SAT [5]. For most of these problems it implies NP-hardness, but for some —most notably the problem of approximating SET COVER within a $\Omega(\log n)$-factor and an entire set of problems in [4] — it is only known to imply *quasi-NP-hardness* (a *quasi-NP-hard* problem has a polynomial-time algorithm only if NP $\subseteq$ Time($n^{polylog(n)}$)).

Plugging our improved analysis of the low degree test into known constructions leads to very efficient *constant-prover* 1*-round proof systems* for NP. Such systems imply the NP-hardness of approximating Set Cover to within a factor of $O(\log n)$ (see the reduction of [26], adapted for more than 2 provers in [10]). Prior to this work, Raz and Safra [29] had constructed such systems; our construction can be viewed as an alternative proof of their result.

In our proof system, a probabilistic polynomial-time verifier checks that a given string is in the language by using $O(\log n)$ random bits, and one round of interaction with a constant number of provers during which it receives $O(\log n)$ bit long answers from the provers. If the input is in the language, the provers can answer in a way that makes the verifier accept with probability 1. If the input is not in the language, then regardless of the prover's answers the verifier accepts with probability at most $2^{-\log^{(1-\epsilon)} n}$, for any $\epsilon > 0$. The number of provers in our construction grows as $O(1/\epsilon)$. If we are willing to increase the error probability to $2^{-\log^{1/3} n}$ then the number of provers can be reduced to 7. The number of provers can possibly be reduced further while maintaining a $o(1)$ error probability, using some techniques of Tardos [37]. However, no known technique seems to be able to reduce the number of provers to two. Thus, the question of whether two-prover proof systems can achieve $o(1)$ error with logarithmic randomness and answer size, remains an open one.

Now we briefly describe low degree tests; see Section 2 for more details. Given an m-variate function $f : \mathrm{F}^m \to \mathrm{F}$ over a finite field F, the test wishes to determine whether or not there exists a degree d polynomial that agrees with f in δ fraction of points in F^m. (The function is presented *by value*, and the test has random access into this table of values. Both d and δ are inputs to the test.) The low degree test is allowed to be probabilistic and it has to read as few values of f as possible.

We are interested in a test proposed by [30] that works roughly as follows: Pick a random "line" in $\mathrm{F}^m$and verify that the restriction of f to this line agrees significantly with some univariate degree d polynomial. If this is the case, accept. This test is similar in flavor to all other known low degree tests, such as the original tests in [11, 7] and later ones in [8, 13, 18]. Our interest in the test of [30] stems from the fact that it has a potential to yield something meaningful about the function being tested, even when the function passes the test with very low probability. (In contrast, several of the other tests above, e.g. those in [7, 8, 13], explicitly test the degree of the polynomial in each variable and consequently are unable to say anything useful unless the function passes the test with probability $1 - O(1/m)$.) However the best prior analysis of this test was not strong enough to match the potential performance of the test.

Let δ denote the probability with which f passes the low-degree test. Prior analyses of all low degree tests (with the exception of that of [29] could not say anything meaningful about f if $\delta < 1/2$; in fact the analyses of [13, 18, 30, 6] required $\delta > 1-O(1/d)$ or $\delta > 1-O(1/m)$. A crucial ingredient of NP=PCP($\log n, 1$) was an analysis (actually just a combination of the analyses of [30, 6]) of the above test that worked for any $\delta > 1-\epsilon$, where $\epsilon > 0$ is fixed. This analysis showed that if a function f passes the test with probability $\delta > 1-\epsilon$, then there exists a degree d polynomial that has agreement $\approx \delta$ with f. (The value of ϵ for which this is true was later improved to 1/8 [17].)

In this paper we present an analysis (see Theorem 17) that continues to say something meaningful about f even when δ is fairly close to 0. We show that if $\delta > (md)^c / |\mathrm{F}|^\epsilon$ for some fixed $c, \epsilon > 0$, then there exists a degree d polynomial that agrees with f in $\approx \delta$ fraction of the inputs. We remark that a similar statement had earlier been proved for really large fields, satisfying $|\mathrm{F}| > 2^{\Omega(m+d+1/\delta)}$ [2, 35]. However, most applications require the field size to be poly$(m, d, 1/\delta)$.

We also prove a related result, Theorem 16, which is more useful for constructing efficient PCP-style verifiers. It says that every function f that passes the low degree test with probability δ has a set of polynomials $P_1, P_2, \ldots$ such that the test fails with high probability if it encounters a point where f does not agree with one of the P_i's. In fact, the set $P_1, P_2, \ldots$ contains every polynomial that has any significant agreement with f. This result is useful because all known verifiers work by checking the properties of some function f provided by the prover. If f is a polynomial, the verifier is extremely unlikely to produce an erroneous answer . Errors creep in when f is not a polynomial but may have some agreement with one or more polynomials, say $g_1, g_2, \ldots$. In this case, if the verifier has the bad luck to examine f at a point where f doesn't agree with any of $g_1, g_2, \ldots$, then it could accept erroneously. Our corollary provides the means to combat such errors, since any such $g_1, g_2, \ldots$ turn out to be exactly the set of polynomials $P_1, P_2, \ldots$, mentioned in Theorem 16. Thus the verifier can just subject f to a low degree test: at any point where f doesn't agree agree with any of $P_1, P_2, \ldots$, the test is guaranteed to fail with high probability, thus averting an erroneous accept. This intuition is formalized to some extent in Section B of the appendix.

Application to program testing/correcting. Suppose we are given a potentially buggy program that purportedly computes an (unknown) m-variate polynomial over a finite field F. Program testing/correcting [11] concerns the following problems: (i) *testing:* determine δ, the fraction of points at which this program is correct and (ii) *correction:* for each input, correct the output of the program in case it is incorrect. It was open how to do testing if $\delta < 1/2$; our low-degree test (specifically, the version that doesn't use an auxiliary table) closes this open problem when $|\mathrm{F}|^\epsilon > poly(md)$. As for correction, note that its meaning is not clear when $\delta < 1/2$, since as many as $O(1/\delta)$ polynomials could have agreement δ with the program. Two notions of correction are possible, as noted in [1]. The weaker one is that for each input, the corrector outputs $O(1/\delta)$ values, one of which is correct. Such a corrector can be derived easily based on [1, 33]. The stronger notion is that the corrector creates $O(\frac{1}{\delta})$ programs (polynomials) such that w.h.p. one of them is correct. Finding such a corrector was an open problem. We provide such a corrector in Section 4.2.

Subsequent to this paper, Sudan, Trevisan and Vadhan [34], have shown how the corrector of Section 4.2 can be used to provide an alternate route to "amplify" the hardness of Boolean functions. Such amplification results in turn play a central role in weakening the assumptions under which it can be shown that probabilistic polynomial time can be simulated deterministically (see [20]). Sudan et al. [34] also give a simpler analysis with improved parameters for the self-correction problem.

Past work. The first construction of a nontrivial constant prover 1-round proof system for NP appeared in [24]; others appeared in [16, 10, 36, 14, 28]. These systems could not reduce the error probability to below a constant while using $O(\log n)$ random bits (the best construction needs $O(k \log n)$ random bits to make the error probability 2^{-k}; see [28]). It was also known [15] that some obvious ideas (such as "recycling randomness") cannot let us get around this. Slightly prior to our work, Raz and Safra [29] found a construction of a proof system achieving subconstant error. Their analysis also relies on a low degree test, albeit a new one and with a very different correctness proof than ours. Raz and Safra also extend their ideas to design constant prover systems in which the error probability is $2^{-O(\#\text{ of answer bits})}$, while the number of random bits stays $O(\log n)$.

Paper organization. We state and explain our main theorem (Theorem 1) in Section 2. We also prove some corollaries of this theorem (Theorems 16 and 17) in Section 2.2 after introducing some basic algebra and geometric facts in Section 2.1. We prove Theorem 1 in Section 3. This proof resembles proofs of earlier results [30, 5, 3, 17], in that it has two parts. First in Sections 3.1 and 3.2 we prove the theorem when m is constant (specifically, $m = 2, 3$); this uses algebraic arguments inspired by Sudan's [33] work on reconstructing polynomials from very noisy data and Kaltofen's work on "Effective Hilbert Irreducibility" [21, 23]. Then in Section 3.3 we "bootstrap" to allow larger m. This part uses probabilistic arguments and relies upon the cases $m = 2, 3$ (including Theorems 16 and 17 for the cases $m = 2, 3$). It is inspired by the "symmetry-based" approach of Arora [3]. Finally, the appendix contains the proof of the version of Hilbert irreducibility theorem that we use (Section A) and the construction of constant prover 1-round proof systems (Section B).

2 The Low-degree Test

Let F be a finite field and m, d be integers. An *m-variate polynomial over F* is a sum of terms of the form $ax_1^{j_1}x_2^{j_2}\cdots x_m^{j_m}$, where $a \in \mathrm{F}$. The set of such polynomials forms an integral domain, denoted $\mathrm{F}[x_1, \ldots, x_m]$. We will often view such a polynomial as a function from F^m to F. The *degree* of the polynomial is its total degree (thus $j_1 + \cdots + j_m$ is the degree of the above monomial). We will reserve the symbol q for $|\mathrm{F}|$, the cardinality of F.

The *distance* between two functions $f, g : \mathrm{F}^m \to \mathrm{F}$, denoted $\Delta(f, g)$, is the fraction of points in F^m they differ on. The *agreement* between the functions is $1 - \Delta(f, g)$.

The low-degree test is given a function $f : \mathrm{F}^m \to \mathrm{F}$. Using randomness, it checks that f looks "locally" like a degree-d polynomial. Magically, it can infer from this that f has significant agreement with a degree-d polynomial.

To be more formal we need to define a *line* in F^m. This is a set of q points with a parametric representation of the form $\{(u_1 + tv_1, u_2 + tv_2, \ldots, u_m + tv_m) : t \in \mathrm{F}\}$ for some $u = (u_1, \ldots, u_m) \in \mathrm{F}^m$ and $v = (v_1, \ldots, v_m) \in \mathrm{F}^m \setminus \{0\}$. We refer to the point $(u_1 + av_1, u_2 + av_2, \ldots, u_m + av_m)$ as the *point* $t = a$ of the line.

In the above definition, replacing u by any other point on the line and v by any non-zero multiple of itself, would leave the line unchanged; our convention is to fix one of these $(|\mathrm{F}|)(|\mathrm{F}| - 1)$ representations as canonical.

Note that the number of lines over a finite field is finite (and equals $q^{m-1}(q^m - 1)/(q - 1)$). Thus it makes sense to define the uniform distribution over lines and we will often refer to a random line in this paper to describe a line chosen at random from the uniform distribution.

Definition 1 Let l be the line $\{(u_1 + tv_1, u_2 + tv_2, \ldots, u_m + tv_m) : t \in \mathrm{F}\}$ and $f : \mathrm{F}^m \to \mathrm{F}$ be a function. Let $g(t)$ be a univariate polynomial. Then g *describes* f *at the point* $t = a$ *of* l if

$$g(a) = f(u_1 + av_1, u_2 + av_2, \ldots, u_m + av_m).$$

□

Note that if $f : \mathrm{F}^m \to \mathrm{F}$ is a degree d polynomial, then on every line the restriction of f to that line is described by a univariate degree-d polynomial in the line parameter t. The converse can also be shown to be true: if on every line in F^m, the values of f are described by a univariate degree-d polynomial and F is sufficiently large ($q \geq (d+1)(\frac{p}{p-1})$, where p is the characteristic of the field [17]), then f must be a degree-d polynomial.

The low degree test is presented with $f : \mathrm{F}^m \to \mathrm{F}$, and an integer d. It is also presented a table that is meant to be a "proof" that f is a degree d polynomial. This table contains, for each line in F^m, a univariate degree d polynomial that supposedly describes the restriction of f to that line. We will use the term *d-oracle* for any table that contains, for each line in F^m, a univariate degree d polynomial. (The number of variables m can be inferred from the context.)

The Low Degree Test:

Pick a random line l in F^m and read the univariate polynomial, say $p_l(t)$, which the given d-oracle contains for this line. Randomly pick a point x on line l and check whether p_l correctly describes f at x. If so, ACCEPT, else REJECT.

We denote by $\delta_d(f, B)$ the probability that the low degree test accepts a function f and a d-oracle B. We will prove the following result about the low degree test.

Theorem 1 (Main) *There are positive integers c_0, c_1, c_2, and c_3 such that the following are true. Let $f : F^m \to F$ be any function and $d \geq 0$ be any integer.*

1. *For any $\delta > 0$, if f has agreement δ with some degree d polynomial, then there is a d-oracle B such that $\delta_d(f, B) \geq \delta$.*
2. *If $\delta > 0$ satisfies $q > c_0((d+1)/\delta)^{c_1}$ and there is a d-oracle B such that $\delta_d(f, B) \geq \delta$, then f has agreement at least $\frac{1}{c_2}(\delta/(d+1))^{c_3}$ with some degree d polynomial.*

We remark that the first half of this theorem is trivial, since any degree d polynomial that has agreement δ with f can be used to create a d-oracle by using the polynomial's restriction to all the lines. A moment's thought shows that f passes the low degree test with probability δ using this d-oracle. We will prove only the more nontrivial second half of the theorem. As mentioned earlier, previous results show that the statement in the second half has been shown true for some $\delta < 1$. This paper shows that the statement is true for $\delta \ll 0.5$, and in fact for δ as small as $(d+1)(c_0/q)^{1/c_1}$, which is tiny if q is, say, $(c_0(d+1))^{2c_1}$.

Before going on to prove Theorem 1 we first introduce some notation that will be used throughout the paper. From now on we will reserve the symbol f for a function from F^m to F which is the subject of the low degree test.

Definition 2 The *line polynomial for f on line l for degree d*, denoted $P_d^f(l)$, is the univariate degree d polynomial (in the line parameter t) that describes f on more points of l than any other degree d polynomial. (We arbitrarily break ties among different polynomials that describe f equally well on the line.) The *d-success-rate of f on line l*, denoted $\mu_d^f(l)$, is defined as

$$\mu_d^f(l) = \text{fraction of points on } l \text{ where } P_d^f(l) \text{ describes } f.$$

The *d-success-rate of f at point $x \in F^m$* is the fraction of lines through x whose line polynomial describes f at x.

The *d-success rate of f* is the average of its d-success rates on all lines. (Note: By symmetry this is also equal to its average d-success rate at all points.)

Note that the probability that a function $f : \mathrm{F}^m \to \mathrm{F}$ passes the low degree test is maximised when the accompanying d-oracle contains, for each line l, the polynomial $P_d^f(l)$. Hence it suffices to prove the following.

Theorem 2 (Restatement of Theorem 1 part 2) *There are integers c_0, c_1, c_2, c_3 such that the following is true. If $f : F^m \to F$ is any function whose d-success rate is at least δ and $q > c_0\left(\frac{d+1}{\delta}\right)^{c_1}$, then there exists a degree d polynomial that has agreement at least $\frac{1}{c_2}(\delta/(d+1))^{c_3}$ with f.*

In the rest of this section we present some background material and show how two "strengthenings" of Theorem 1 that can be obtained, once the theorem is proven.

2.1 Preliminaries

In this section we present some elementary algebraic, probabilistic and geometric results which will be used later in our proofs. We start with some basic properties of polynomials.

Lemma 3 (Schwartz [31]) *A non-zero m-variate degree D polynomial over a field F can be zero at no more than $D/|H|$ fraction of points in H^m, for any set $H \subseteq F$.*

Using Lemma 3 it is easy to prove the following well-known fact that there are not "too many" polynomials that have significant agreement with a given function.

Proposition 4 *Let $f : F^m \to F$ be any function. Suppose integer $d \geq 0$ and fraction ρ satisfy $\rho > 2\sqrt{\frac{d}{q}}$. Then there are at most $2/\rho$ degree d polynomials that have agreement at least ρ with f.*

Proof: Suppose $k = \lfloor 2/\rho \rfloor + 1$ distinct polynomials $P_1, P_2, \ldots, P_k$ each have agreement ρ with f. Then P_1 describes f in at least ρ fraction of the points. By applying Lemma 3 to $P_1 - P_2$ we find that P_1 and P_2 can agree in at most d/q fraction of the points. Thus P_2 describes f in at least $\rho - d/q$ fraction of the points where $P_1 \neq f$. Similarly, P_3 describes f in at least $\rho - 2d/q$ fraction of the points where $P_1 \neq f$ and $P_2 \neq f$, etc.

Thus the polynomials together describe f in at least

$$\rho + (\rho - \frac{d}{q}) + (\rho - \frac{2d}{q}) + \cdots + (\rho - (k-1)\frac{d}{q})$$

fraction of the points. This fraction is at least

$$k\rho - \binom{k}{2}\frac{d}{q}.$$

Since $\rho > 2\sqrt{d/q}$, we have that the above fraction is more than 1 for every choice of $k \in [\frac{2}{\rho}, \frac{2}{\rho} + 1]$. In particular this holds for $k = \lfloor 2/\rho \rfloor + 1$, which is impossible. □

Proposition 5 *There exists a function $h : F^m \to F$ that has agreement at most $(d+1)/q$ with every degree d polynomial.*

Proof: Consider the function $h(x_1, \ldots, x_m) = x_1^{d+1}$. Thus by Lemma 3 it has agreement at most $\max\{\deg(p), \deg(h)\}/q$ with any polynomial $p \neq h$. In particular, letting p be any polynomial of degree at most d, we find $h \neq p$ and thus the agreement is at most $(d+1)/q$. □

Before going on to the next lemma, we will introduce a slightly modified notion of a line, that we will call a quasi-line. Recall that a line was specified by a point $u \in \mathrm{F}^m$ and a point $v \in \mathrm{F}^m - \{0\}$. If we allow the vector v to take on the value 0 as well, then we get a collection of objects that we will call "quasi-lines". The uniform distribution on quasi-lines is obtained by picking $u, v \in \mathrm{F}^m$ independently, uniformly at random and let $\{(u_1 + tv_1, \ldots, u_m + tv_m) | t \in \mathrm{F}\}$ be the chosen quasi-line. Note that a random quasi-line is obtained by picking a random singleton subset of F^m with probability $1/q^m$ and picking a random line with the remaining probability (and thus is a close approximation to the uniform distribution on lines). We will often rely on some properties of quasi-lines in the proofs below.

Lemma 6 *Let $p \in F[x_1, x_2, \ldots, x_m]$ be a polynomial of degree exactly D. Then on at least $1 - D/|F|$ fraction of lines in F^m, the restriction of p has degree no less than D*

Proof: We will actually show that on at most $D/|\mathrm{F}|$ fraction of quasi-lines, the restriction of p to the quasi-line may be a polynomial of degree less than D. This suffices, since if a quasi-line is not a line, the p restricted to it is a constant and hence a degree 0 polynomial. Thus the fraction of lines for which this holds is even smaller.

Let $Q(y_1, \ldots, y_m, z_1, \ldots, z_m, t) = p(y_1 + z_1 t, \ldots, y_m + z_m t)$. View it as a univariate polynomial, of degree at most D, in t whose coefficients are in the ring $\mathrm{F}[y_1, \ldots, y_m, z_1, \ldots, z_m]$. Let $Q_D(y_1, \ldots, z_m)$ denote the coefficient of t^D in Q. Notice that p restricted to the set $\{(a_1 + b_1 t, \ldots, a_m + b_m t) : t \in \mathrm{F}\}$ is a polynomial of degree exactly D iff $Q_D(a_1, \ldots, a_m, b_1, \ldots, b_m)$ is non-zero. Note that Q_D is of degree at most D. We claim that Q_D is non-zero and thus is non-zero with probability at least $1 - D/|\mathrm{F}|$ for a random setting of $y_1, \ldots, y_m, z_1, \ldots, z_m$. This suffices to prove the lemma.

To prove the claim, let $p = p_1 + p_2$, where p_1 is homogenous of degree D and p_2 is of degree less than D. Note that $p_1 \neq 0$, since p is exaactly of degree D. The claim follows from the observation that $Q_D(0, \ldots, 0, b_1, \ldots, b_m) = p_1(b_1, \ldots, b_m) \neq 0$. □

Lemma 7 *Let F be a field and $\overline{F}$ be any field extending F. Let H be a finite subset of F. Let $p \in \overline{F}[x_1, x_2, \ldots, x_m]$ be a polynomial of total degree D that takes values in F on more than $D/|H|$ fraction of the points in H^m. Then p is a polynomial in $F[x_1, \ldots, x_m]$.*

Proof: Let $S = \{x \in H^m | p(x) \in \mathrm{F}\}$. We claim there exists a polynomial $q \in \mathrm{F}[x_1, x_2, \ldots, x_m]$ of degree at most D such that $q(x) = p(x)$ for every $x \in S$. Note that q, if it exists, is the solution to a linear system over F. Furthermore, the linear system has a solution over $\overline{\mathrm{F}}$, namely p. Hence a solution exists over F and this yields the polynomial q as desired. Now viewing p and q as polynomials in $\overline{\mathrm{F}}[x_1, \ldots, x_m]$, note that they agree with probability more than $D/|H|$ over inputs from H^m. Applying Lemma 3 to the polynomial $p - q$, we find $p - q = 0$ and thus $p \in \mathrm{F}[x_1, \ldots, x_m]$. □

Lemma 8 *Let F be a finite field and $\overline{F}$ be any field extension of F. Let $p \in \overline{F}[x_1, x_2, \ldots, x_m]$ be a polynomial of degree at most D. If on more than $D/|F|$ fraction of the lines in F^m, the restriction of p to the line is in $F[t]$ (where t is the line parameter), then p must be a polynomial in $F[x_1, \ldots, x_m]$.*

Proof: Let $K = \mathrm{F}(t)$ be the field of rational functions in t over F and let $\overline{K} = \overline{\mathrm{F}}(t)$. Let $H = \{a + tb | a, b \in \mathrm{F}\} \subseteq K$. View p as a polynomial in $K[x_1, \ldots, x_m]$ and note that the restriction of p to the set $\{(a_1 + b_1 t, \ldots, a_m + b_m t) : t \in \mathrm{F}\}$ is given by evaluating p at the point $(a_1 + b_1 t, \ldots, a_m + b_m t) \in H^m$. Applying Lemma 7 to the polynomial $p \in \overline{K}[x_1, \ldots, x_m]$, we find that if p takes values in $\mathrm{F}(t)$ for more than $D \cdot |H|^{m-1}$ points in H^m, then $p \in K[x_1, \ldots, x_m]$ (which implies $p \in \mathrm{F}[x_1, \ldots, x_m]$). Since every line in F^m corresponds to $|\mathrm{F}|(|\mathrm{F}| - 1)$ points in H^m, we also find that if p restricted to more than $D \cdot |\mathrm{F}|^{2m-3} / (|\mathrm{F}| - 1)$ lines takes on a value in $\mathrm{F}[t]$, then $p \in \mathrm{F}[x_1, \ldots, x_m]$. The lemma follows easily since

$$\frac{D \cdot |\mathrm{F}|^{2m-3}}{|\mathrm{F}| - 1} \cdot \frac{|\mathrm{F}|(|\mathrm{F}| - 1)}{|\mathrm{F}|^m (|\mathrm{F}|^m - 1)} = \frac{D\,|\mathrm{F}|^{m-2}}{|\mathrm{F}|^m - 1} \leq \frac{D}{|\mathrm{F}|}.$$

□

Our next lemma is a basic probabilistic fact that we will use repeatedly throughout Section 3.

Lemma 9 (Markov's inequalities) *Let $r_1, r_2, \ldots, \in [0, 1]$ be some real numbers whose average is α. Then (i) For any $\rho < 1$, at least $\frac{\alpha - \rho}{1 - \rho}$ fraction of them are greater than ρ (ii) at most $1/k$ fraction of them are more than $k \cdot \alpha$.*

Proof: (i) If the desired fraction is s, then s satisfies $s + (1 - s)\rho \geq \alpha$.(ii) If the desired fraction is t then t satisfies $t \cdot k\alpha \leq \alpha$. □

To give a typical example of the way we will use Lemma 9, suppose the indices correspond to the different lines in F^m, and r_i is the the d-success rate of a function f on the ith line. If the d-success rate of f is ϵ, then the Lemma implies that for $\epsilon/2$ fraction of the lines l, the d-success rate of f on l is at least $\epsilon/2$.

The rest of this section presents some "geometric" results about F^m. The results will be especially useful in Section 3.3 where we analyze the general case of the low-degree test. The notions below as well as the results are based on Arora's work [3].

Definition 3 Let $m, k \in \mathcal{Z}^+$ and $k < m$. A *k-dimensional affine subspace* of F^m is a set of points with a parametrization of the form

$$\{\overline{u_0} + t_1 \cdot \overline{u_1} + t_2 \cdot \overline{u_2} + \cdots + t_k \cdot \overline{u_1} : t_1, t_2, \ldots, t_k \in \mathrm{F}\},$$

for some $\overline{u_0} \in \mathrm{F}^m$ and linearly independent vectors $\overline{u_1}, \overline{u_2}, \ldots, \overline{u_k} \in \mathrm{F}^m$. □

Thus a *line* is a 1-dimensional subspace, for example. We will refer to a 2-dimensional subspace as a *plane* and a 3-dimensional subspace as a *cube.* A function defined on a k-dimensional subspace of F^m is called a degree d polynomial if the function can be expressed as a degree d polynomial in the parameters $t_1, \ldots, t_k$.

Note that each set of $k+1$ affine independent points in F^m determines a unique k-dimensional subspace. For example, a line and a point outside it determine a unique plane, two lines that are not in the same plane determine a unique cube, and so on. We will often use the fact that picking a random cube in F^m is essentially the same as picking four random points (or two random lines) in F^m and taking the cube that contains these points (resp. lines). The main difference arises since the points (or lines) may turn out to be "coplanar" - i.e., they lie on the same 2-dimensional affine subspace. The following lemma bounds the probability of this event from above.

Lemma 10 *For $m \geq 3$, the probability that four points chosen uniformly and independently from F^m are coplanar is at most $\frac{1}{q} + \frac{1}{q^2} + \frac{1}{q^3} \leq \frac{2}{q}$. For $m \geq 3$, the probability that two lines chosen uniformly and independently from F^m are coplanar is at most $\frac{1}{q} + \frac{1}{q^2} \leq \frac{2}{q}$.*

Proof: Note $q \geq 2$. Notice that points x_1, x_2, x_3, x_4 are coplanar iff $x_2 - x_1, x_3 - x_1, x_4 - x_1$ are linearly dependent. In turn this happens iff $x_2 - x_1 = 0$, or $x_3 - x_1$ is linearly dependent on $x_2 - x_1$ or if $x_4 - x_1$ is linearly dependent on the first two. The probabilities of these events is upper bounded by $\frac{1}{q^m}$, $\frac{1}{q^{m-1}}$, and $\frac{1}{q^{m-2}}$ respectively. Thus the union of these events has probability at most $\frac{1+q+q^2}{q^m}$. Using $m \geq 3, q \geq 2$ the first part of the lemma follows immediately.

The second part is reasoned similarly picking x_1, x_2 from the first line and x_3, x_4 from the second line. The only difference (to our advantage) is that we can pick $x_1 \neq x_2$ and this gives a bound of $\frac{1}{q} + \frac{1}{q^2}$ on the probability of coplanarity. □

Our argument will rely on symmetry, such as the following facts: (i) all points in F^m are part of exactly the same number of k-dimensional subspaces (ii) All lines in F^m are part of exactly the same number of k-dimensional subspaces, etc. We give an illustrative example of a symmetry-based calculation.

Example 1 Suppose $f : \mathrm{F}^m \to \mathrm{F}$ is any function whose d-success-rate is exactly β. For any plane s let t_s be the average d-success-rate of f among lines in s. Then symmetry implies that $E_s[t_s]$, the average of t_s among all planes, is exactly β. The reason is that $\sum_s t_s$ counts every line in F^m an equal number of times.

The following lemmas provide other examples of symmetry-based arguments that will be used later in this section.

Lemma 11 (Well-distribution lemma for lines) *Let $S \subseteq F^m$ be a set whose size is $\mu \cdot q^m$. For every $K > 0$, at least $1 - \frac{1}{K^2}$ fraction of lines in F^m have between $\mu q(1 - \frac{K}{\sqrt{\mu q}})$ and $\mu q(1 + \frac{K}{\sqrt{\mu q}})$ points from S.*

Proof: Imagine picking a line $l = \{u + tv : t \in \mathrm{F}\}$ randomly from F^m. For $a \in \mathrm{F}$ let the random variable X_a be 1 if $u + av \in S$ and 0 otherwise. Notice that the assertion is equivalent to showing that

$$\Pr_{u,v}[|\sum_{a \in \mathrm{F}} X_a - \mu q| \geq K\sqrt{\mu q}] \leq \frac{1}{K^2}.$$

Then $E_{u,v}[X_a] = \mu$. We would like to argue next that X_a and X_b are independent if $a \neq b$, but this is not strictly true (it would be true for quasi-lines, but not lines). Instead we bound the variance of $\sum_a X_a$. Let $Y_a = X_a - \mu$. A simple calculation shows that $E[X_a X_b] = \mu(\mu - \frac{1-\mu}{|\mathrm{F}|^m - 1})$ and thus $E[Y_a Y_b] = -\frac{\mu(1-\mu)}{|\mathrm{F}|^m - 1} \leq 0$, and thus $E[(\sum_a Y_a)^2] \leq \mu(1-\mu)q \leq \mu$. The lemma follows from Chebyshev's inequality

$$\Pr_{u,v}[|\sum_{a \in \mathrm{F}} Y_a| \geq K\sqrt{\mu q}] \leq \frac{1}{K^2},$$

which holds for random variables with expectation 0. □

We choose to state the next lemma in terms of the d-success-rate, though it is true in general if instead of d-success-rate we associate any set of positive fractions with the lines of F^m and look at their average value in a random cube.

Lemma 12 (Well-distribution lemma for cubes) *For any $\alpha > 0$ and $m > 3$, if any function $f : F^m \to F$ has d-success-rate δ, then in a random cube C,*

$$\Pr_{\text{cube } C}[\textit{Average d-Success-rate of } f \textit{ on lines in } C \leq (1-\alpha)\delta] \leq \frac{3}{\alpha^2 \delta^2 |F|}.$$

Proof: Let the random variable X_C denote the average d-success-rate of f on lines in cube C. When C is chosen uniformly at random from the space of all cubes, by symmetry (see the note on symmetry in Example 1) we have $E_C[X_C] = \delta$. (Here and later $V[\cdot]$ denotes variance.) Let Y_C be the random variable $X_C - \delta$, so $E_C[Y_C] = 0$. By Chebyshev's inequality, we have

$$\Pr_C[|Y_C| \geq \alpha\delta] \leq \frac{V[Y_C]}{\alpha^2\delta^2}.$$

We show next (using an argument similar to one in [9, Section 7.1]) that $V[X_C] \leq 3/|\mathrm{F}|$, thus proving the lemma.

Let $q = |\mathrm{F}|$ and let K denote the number of lines in F^3 (thus $K = \binom{q^3}{2}/\binom{q}{2}$). Let us number these lines in some arbitrary way, so that we can refer to the "ith line of F^3." We similarly talk about the "ith line of cube C." Now for $1 \leq i \leq K$, let $Y_{i,C}$ be the r.v. $Y_{i,C} =$ (success rate of f on ith line of C) $- \delta$. Note that $Y_C = E_{i \leq K}[Y_{i,C}]$. Hence

$$V[Y_C] \quad = \quad E_C\,[\,E_{i,j \leq K}[Y_{i,C}Y_{j,C}]\,] \tag{1}$$

To upperbound the expression in (1), we assert the following:

1. By Lemma 10 we have that at most $1/q + 1/q^2$ fraction of all pairs (i,j) correspond to a pair of coplanar lines in F^3.
2. $-1 \leq E[Y_{i,C}Y_{j,C}] \leq 1$ if $Y_{i,C}$ and $Y_{j,C}$ are coplanar.
3. Finally we show that $E[Y_{i,C}Y_{j,C}] \leq \frac{(1/q+1/q^2)}{(1-1/q-1/q^2)}$ if i,j are indices to two random non-coplanar lines in F^3. Note that picking a random cube C in F^m and then picking any two (fixed) non-coplanar lines in C is equivalent to picking two random non-coplanar lines in F^m. Let $Y_l =$ (d-success rate of l) $-\delta$ and let τ denote the quantity $E[Y_{l_1}Y_{l_2}]$ when l_1 and l_2 are two random non-coplanar lines in F^m. We now give a bound on τ. Notice that if we pick l_1 and l_2 at random, independently then $E[Y_{l_1}Y_{l_2}] = 0$. Further, once we have fixed l_1, the fraction of lines in F^m that are coplanar with it is at most $\rho = 1/q + 1/q^2$. Further, we have $0 = E[Y_{l_1}Y_{l_2}] \geq \rho \cdot (-1) + (1-\rho) \cdot \tau$. Thus $\tau \leq \rho/(1-\rho) \leq (1/q + 1/q^2)/(1 - 1/q - 1/q^2)$.

Putting the above together we get:

$$V[Y_C] \leq (\frac{1}{q} + \frac{1}{q^2}) \cdot 1 + (1 - (\frac{1}{q} + \frac{1}{q^2})) \cdot \frac{(1/q + 1/q^2)}{(1 - 1/q - 1/q^2)} = 2/q + 2/q^2 \leq \frac{3}{q},$$

where the last inequality uses $q \geq 2$. □

2.2 Stronger Forms of Theorem 1

We now show how analyses of the low-degree test can be strengthened to obtain larger agreement, or an explanation of the test's behaviour at most points. Specifically, we show two stronger forms (Theorems 16 and 17) one of which will be useful in constructing proof systems. The first strong form says, heuristically speaking, that if $q > \text{poly}(\frac{d}{\rho})$, then every function that passes the low degree test with probability ρ has agreement at least $\rho - \epsilon$ with some degree d polynomial. (Note: Theorem 1 only guaranteed an agreement ρ^{c_3}/c_2). The second strong form states, roughly, that "almost all" of the success probability of the low degree test happens at points where f agrees with (one of) a small set of polynomials, where the size of the set is bounded by Proposition 4. While the theorems show how to use Theorem 1 to get these stronger results, the proofs themselves are more general and work, for example, for specific choices of m as well. We abstract the more general results in the next three lemmas before using them to prove the stronger forms of the main theorem. The lemmas will be used to strength our analysis of the low-degree test for the cases $m = 2, 3$ in Sections 3.1 and 3.2. We will then be able to apply the strong forms when we analyze the general case of the low-degree test in Section 3.3.

Our first lemma shows how to explain the behaviour of the low-degree test at most points of the space, even when the success rate is very small.

Lemma 13 *Let $m, d, q, \epsilon, \delta$ be such that the following holds: "For every function $g : F^m \to F$ with d-success rate at least δ, g has agreement at least ϵ with some m variate polynomial of degree at most d." Let $f : F^m \to F$ and let $P_1, \ldots, P_k$ be a list of all polynomials that have agreement $(\epsilon - (d+1)/q)$ with f. Then, for every d-oracle, with probability at least $1 - \delta$, at least one of the following events happen on a run of the low-degree test:*

- *The low-degree test outputs REJECT.*
- *The low-degree test picks a point $x \in F^m$ such that there exists an $i \in \{1, \ldots, k\}$ s.t. $f(x) = P_i(x)$.*

Proof: Suppose the probability mentioned in the theorem statement is less than $1 - \delta$. We derive a contradiction.

Let $h : \mathrm{F}^m \to \mathrm{F}$ be a function with agreement at most $(d + 1)/q$ with any degree d polynomial (as known to exist from Prop 5). Let $S \subseteq \mathrm{F}^m$ be the set of points at which f does *not* agree with any of $P_1, \ldots, P_k$. Let $g : \mathrm{F}^m \to \mathrm{F}$ be defined to as follows: $g(x) = f(x)$ if $x \in S$ and $g(x) = h(x)$ otherwise.

We first note that g has d-success rate at least δ due to points in S alone. Note that by the contradiction assumption, the probability that the low-degree test picks a point in S and the test passes is at least δ. Since $g(x) = f(x)$ in such cases, there exists a d-oracle such that g also satisfies the same condition (i.e., the low-degree test picks a point in S and g passes the low-degree test).

By hypothesis, we now have that there exists a degree d polynomial P with agreement at least ϵ with g. Next we note that this agreement must largely be on points in S. This is so because g restricted to points outside S equals h which does not have much agreement with any degree d polynomial. Specifically g and P have agreement at most $(d + 1)/q$ on points not in S. Thus g and P must have agreement at least $\epsilon - (d + 1)/q$ on points in S. But this implies f and P have agreement at least $\epsilon - (d + 1)/q$, which contradicts the hypothesis that $P_1, \ldots, P_k$ is an exhaustive list of all polynomials with agreement at least $\epsilon - (d + 1)/q$ with f. (Note that P can not be one of the P_i's since none of the P_i's agree with f at any point in S.) This yields the desired contradiction. □

The following lemma gives an analogous result, with a slightly different conclusion: this time the conclusion says that the success of the low-degree test is due to some nice lines, rather than some nice points.

Lemma 14 *Let $m, d, q, \epsilon, \delta$ be such that the following holds: "For every function $g : F^m \to F$ with d-success rate at least δ, g has agreement at least ϵ with some m variate polynomial of degree at most*

d." Let $f : F^m \to F$ and let $\epsilon > 2\sqrt{d/q} + (d+1)/q$. Let $P_1, \ldots, P_k$ be a list of all polynomials that have agreement $\epsilon - (d+1)/q$ with f. Then, for every d-oracle, with probability at least $1 - (\delta + \frac{2d}{\epsilon q - (d+1)})$, at least one of the following events happen on a run of the low-degree test:

- *The low-degree test outputs REJECT.*
- *The low-degree test picks a line such that the response of the d-oracle agrees with P_i restricted to the line, for some $i \in \{1, \ldots, k\}$.*

Proof: The proof is similar to that of Lemma 13. We define S, h and g as in that proof. We claim that if the lemma is not true, then g still has high success rate. This part needs to be argued differently and this is done so next. Note first that, by Proposition 4, we have $k \leq \frac{2}{\epsilon - (d+1)/q}$, since $\epsilon - (d+1)/q > 2\sqrt{d/q}$. Thus the contradiction assumption says that with probability at least $\delta + kd/q$, the low-degree test picks a line such that the response of the d-oracle does not agree with any of the P_i's but the test still accepts. Note that on such a line l, at most a (kd/q)-fraction of the points satisfy the condition that $f(x) = P_i(x)$ (for some i) and the pair (x, l) are accepted by the low-degree test. Thus the probability that $g(x) \neq f(x)$ on such a line is at most kd/q. Thus we get that the probability that g passes the low-degree test is at least δ.

We now conclude as before: g has agreement ϵ with some degree d polynomial P. This polynomial has agreement at least $\epsilon - (d+1)/q$ with f also (as argued in Lemma 13 and thus we find a polynomial P distinct from $P_1, \ldots, P_k$ that has agreement $\epsilon - (d+1)/q$ with f, yielding the desired contradiction. □

Our next lemma uses Lemma 14 to show larger agreement between f and the low-degree polynomial that is close to it.

Lemma 15 *Let $m, d, q, \epsilon, \delta$ be such that the following holds: "For every function $g : F^m \to F$ with d-success rate at least δ, g has agreement at least ϵ with some m variate polynomial of degree at most d." Further, let $\rho, \alpha > 0$ be such that $\frac{\alpha}{4}(\rho - \alpha/2) > \delta$ and $q > \max\{\frac{16(d+1)}{\epsilon^2}, \frac{4d}{\epsilon(\frac{\alpha}{4}(\rho - \frac{\alpha}{2}) - \delta)}, \frac{64\rho}{\epsilon\alpha^3}\}$. Then, if $f : F^m \to F$ has d-success rate at least ρ, then it has agreement $\rho - \alpha$ with some degree d polynomial.*

Proof: Let $P_1, \ldots, P_k$ be all polynomials with agreement at least $\epsilon - (d+1)/q$ with f. From the condition $q > 16(d+1)/\epsilon^2$, it follows that $\epsilon - (d+1)/q > \epsilon/2 > 2\sqrt{d/q}$. Thus (from Proposition 4) we have $k \leq 4/\epsilon$. For $i \in \{1, \ldots, k\}$, let ρ_i denote the agreement between f and P_i. We wish to show that $\rho_i \geq \rho - \alpha$ for some i. Assume for contradiction, that $\rho_i < \rho - \alpha$ for every i. We will argue that this contradicts the fact that f has high d-success rate, on many lines.

Pick a line l randomly from F^m. By Lemma 9, with probability at least $\alpha/2$ it is the case that the d-success rate of f on l is $\geq \rho - \alpha/2$. In other words,

$$\Pr_l[\text{some univ. deg. } d \text{ polynomial } g_l \text{ describes } f \text{ on } \rho - \alpha/2 \text{ fraction of points of } l] \geq \frac{\alpha}{2} \qquad (2)$$

Where could these univariate degree d polynomials g_l of Assertion (2) come from? We divide the analysis into two cases and we show that each happens with probability less than $\alpha/4$ (over the choice of l), which leads to the desired contradiction.

Case (i): g_l is not the restriction of any of the P_i's to the line l. Let τ be the fraction of lines for which this case holds. Fix any d-oracle whose response on such a line is g_l. Then notice that with probability at least $\tau \cdot (\rho - \alpha/2)$ it holds that the low-degree test picks a line for which the line polynomial is not the restriction of one of the P_i's and the low-degree test accepts. From Lemma 14 it follows that $\tau \cdot (\rho - \alpha/2) < \delta + kd/q$. Using $k \leq 4/\epsilon$ and $q > \frac{4d}{\epsilon(\frac{\alpha}{4}(\rho - \alpha/2) - \delta)}$, we find that $\tau < \alpha/4$ as desired.

Case (ii): There exists an $i \in \{1, \ldots, k\}$ such that g_l is the restriction of P_i to the line l. For any fixed i, we argue that this happens with probability at most $\alpha/4k$. Let $\rho_{i,l}$ denote the fraction of points on line l for which P_i agree with f. Note that $\mathrm{E}_l[\rho_{i,l}] = \rho_i < \rho - \alpha$. By Lemma 11 it follows that

$$\Pr_l[\rho_{i,l} - \rho_i > \frac{\alpha}{2}] \leq \frac{4\rho_i}{\alpha^2 q} \leq \frac{4(\rho - \alpha)}{\alpha^2 q}. \qquad (3)$$

Thus we get that the probability that g_l has agreement at least $\rho - \alpha/2$ with f on the line l is at most $\frac{4(\rho-\alpha)}{\alpha^2 q}$. Using the fact that $q > \frac{64\rho}{\epsilon\alpha^3}$, we get that the last probability is at most $\alpha/(4k)$ as desired. □

We now move onto the promised "strong forms of Theorem 1". Applying Lemma 15 to Theorem 1, we get the first strong form of the low-degree test. Applying Lemma 13 to Theorem 16, we get the second strong form of the low-degree test.

Theorem 16 *There exist constants c_0, c_1 such that for every m, d, ρ and field F of cardinality $q > c_0((d+1)/\rho)^{c_1}$, if $f : F^m \to F$ has d-success rate ρ, then f has agreement at least $2\rho/3$ with some degree d polynomial.*

Proof: Let c_0', c_1', c_2', c_3' be the constants given by Theorem 1. We apply Lemma 15 with $\alpha = \rho/3$, $\delta = \rho^2/18$, $\epsilon = \frac{1}{c_2'}(\delta/(d+1))^{c_3'}$. The lemma, combined with Theorem 1, guarantees that some degree d polynomial has $2\rho/3$-agreement with f provided

$$q > \max\left\{ c_0'(\frac{(d+1)}{\delta})^{c_1'}, 16\frac{d+1}{\epsilon^2}, \frac{72d}{\epsilon\rho^2}, 64\rho/(\epsilon\alpha^3) \right\}.$$

It is easily verified that the lower bound on q is of the form $c_0((d+1)/\rho)^{c_1}$ for appropriately chosen c_0, c_1. □

Theorem 17 *There exist constants c_0, c_1 such that if γ, d and F satisfy $|F| \geq c_0((d+1)/\gamma)^{c_1}$, then for every function $f : F^m \to F$ there exists a set of at most $\frac{4}{\gamma}$ polynomials $Q_1, \ldots, Q_k$ that have at least $\frac{\gamma}{2}$ agreement with f, such that the probability that the low-degree test picks a point where f does not equal any of the Q_i's, and still accepts, is at most γ.*

Proof: This theorem follows by applying Lemma 13 to Theorem 16. Let c_0', c_1' be constants guaranteed to exist by Theorem 16. We will show that the corollary holds for $c_0 = \max\{c_0', 16\}$ and $c_1 = \max\{c_1', 2\}$. Under this condition, we have $q > \max\{c_0'((d+1)/\gamma)^{c_1'}, 6(d+1)/\gamma, 16d/\gamma^2\}$. Let $Q_1, \ldots, Q_k$ be all polynomials with agreement at least $\gamma/2$ with f. By Proposition 4 the number of such polynomials is at most $4/\gamma$. Furthermore by the constraints on q, we have $2\gamma/3-(d+1)/q > \gamma/2$. Since Theorem 16 holds with $\delta = \gamma$, we may apply Lemma 13, with $m' = 2$, $d' = d$, $\epsilon = 2\gamma/3$ and $\delta = \gamma$, we get that the average d-success rate of f on any of the points where $f(x)$ does not equal one of $Q_i(x)$ is at most γ. □

Remark: Using Lemma 14 instead of Lemma 13 in the proof above, one can also obtain the following alternative statement, in which the conclusion is replaced by "... such that the average d-success rate of f on lines where the d-oracle does not agree with the restriction of any of the Q_i's is at most 2γ."

3 Analysis of the low-degree test

We start by proving the correctness of the low-degree test for the bivariate case in Section 3.1. The trivariate case is similar (though a little bit more complicated) and is described in Section 3.2. The proofs of Sections 3.1 and 3.2 can be extended to the m-variate case, but require a field size that is exponential in m. In Section 3.3 we give a better proof that works with fields of polynomial size. It relies on the correctness of the test for $m = 2, 3$.

3.1 The Bivariate Case

In this section we prove Theorem 2 for $m = 2$. Let $f : \mathrm{F}^2 \to \mathrm{F}$ be a function with d-success-rate at least δ. Our proof goes in two steps.

(Step 1). Show that there is a polynomial $Q \in \mathrm{F}[z, x, y]$ of "not too large degree" such that for a "reasonably large" set of points $S \subseteq \mathrm{F}^2$, the following are true:

$$Q(f(a, b), a, b) = 0 \qquad \forall (a, b) \in S \tag{4}$$

$$\text{the } d\text{-success rate of } f \text{ at every point in } S \text{ is "nontrivial."} \tag{5}$$

(Step 2). Show that any Q that satisfies the conditions in Step 1 has a factor $z - g(x, y)$, such that $g \in \mathrm{F}[x, y]$ is a degree d polynomial and for "many" $(a, b) \in S$

$$f(a, b) = g(a, b) \tag{6}$$

By the end of Step 2, we have concluded that f has significant agreement with the degree d bivariate polynomial g. Step 2 depends on a fairly difficult technical fact, Theorem 34, which will be proved separately in Section A in the appendix. Step 1 is motivated by Sudan's [33] technique for reconstructing univariate polynomials from very noisy data, which we describe next.

Sudan makes the following observation. (We note that exactly the same proposition was used in an earlier paper on low-degree testing, namely [27], though in a different context.)

Proposition 18 *Let $(a_1, y_1), \ldots, (a_n, y_n)$ be any set of n pairs from F^2, and d_z, d_x be any positive integers satisfying $d_x d_z > n$. Then there exists a nonzero bivariate polynomial $\Gamma \in F[z, x]$ with $deg_z(\Gamma) \leq d_z$ and $deg_x(\Gamma) \leq d_x$, satisfying*

$$\Gamma(y_i, a_i) = 0 \qquad \textit{for } i = 1, \ldots, n \tag{7}$$

Remark: We can view Γ as an implicit description of the sequence $(a_1, y_1), \ldots, (a_n, y_n)$, in the following sense: for each a_i, one of the roots of $\Gamma(z, a_i)$ is y_i.

Proof: If we let γ_{ij} be the coefficient of $z^i x^j$ in Γ, then the contraints in (7) define the following homogeneous linear system with $(1 + d_x)(1 + d_y)$ unknowns and n constraints. (Note that $a_1, \ldots, a_n, y_1, \ldots, y_n$ are "constants.")

$$\begin{aligned}
\sum_{i=0}^{d_z} \sum_{j=0}^{d_x} \gamma_{ij} y_1^j a_1^i &= 0 \\
\sum_{i=0}^{d_z} \sum_{j=0}^{d_x} \gamma_{ij} y_2^j a_2^i &= 0 \\
\ldots &= 0 \\
\sum_{i=0}^{d_z} \sum_{j=0}^{d_x} \gamma_{ij} y_n^j a_n^i &= 0
\end{aligned}$$

Since $(1 + d_x)(1 + d_y)$, the number of variables, exceeds n, the number of constraints, a nontrivial solution exists. □

To use Γ in polynomial reconstruction, Sudan relies on a lemma from Ar et al. [1].

Lemma 19 *Let $(a_1, y_1), \ldots, (a_n, y_n) \in F^2$ be distinct points such that for some $\rho > 0$, there is a degree d polynomial $h \in F[x]$ satisfying $h(a_i) = y_i$ for ρn values of i. Let $\Gamma \in F[z, x]$ be any polynomial satisfying (7). If $deg_x(\Gamma) + d \cdot deg_z(\Gamma) < \rho n$, then $(z - h(x)) \mid \Gamma$.*

Proof: The polynomial $\Gamma(h(x), x)$ has degree at most $deg_x(\Gamma) + d \cdot deg_z(\Gamma)$ and has at least ρn roots. So if $deg_x(\Gamma) + d \cdot deg_z(\Gamma) < \rho n$, this polynomial must be identically 0. Since $\Gamma(h(x), x)$ is the remainder obtained by dividing Γ with $z - h(x)$, we conclude that $z - h(x) \mid \Gamma(z, x)$. □

Remark: Sudan's observations lead to efficient algorithms because both Lemma 19 and Proposition 18 have "constructive" versions: efficient algorithms exist for polynomial factorization (needed for Lemma 19) and solving linear equations (needed for Proposition 18) over finite fields. The current paper is not about algorithm design, but nevertheless the key algebraic facts used in polynomial factorization and solving linear equations also drive our result. See for example the use of "effective Hilbert irreducibility" in Section A and Cramer's Rule in Lemma 20.

We will need the following generalization of Sudan's ideas to F$[y]$, the ring of univariate polynomials in the formal variable y.

Lemma 20 *Let $S_1, S_2 \subseteq F$ be any subsets of F and $l = |S_1|$. For each $a, b \in F$, let $C_a \in F[y], R_b \in F[x]$ be polynomials of degree at most d. Suppose there is a fraction $\rho > (d+1)/\sqrt{l}$ such that for all $b \in S_2$,*

$$C_a(b) = R_b(a) \qquad \text{for at least } \rho l \text{ values of } a \in S_1. \tag{8}$$

Then there is a non-zero polynomial $Q \in F[z, x, y]$ satisfying $\deg_z(Q) \leq \sqrt{l}$, $\deg_x(Q) \leq \sqrt{l}$, $\deg_y(Q) \leq dl^{3/2}$ such that

$$\forall a \in S_1, \;\; Q(C_a(y), a, y) = 0 \;\; \text{and} \tag{9}$$

$$\forall b \in S_2, \;\; (z - R_b(x)) \,|\, Q(z, x, b) \tag{10}$$

(Note that (10) can be restated as $Q(R_b(x), x, b) = 0$.)

Proof: Let F$[y][z, x]$ be the ring of polynomials in the formal variables z and x whose coefficients are from F$[y]$.

We use the same idea as in Proposition 18, but work over the integral domain F$[y]$ instead of over F. Consider the following sequence of $|S_1|$ pairs from F $\times$ F$[y]$: $((a, C_a(y)) : a \in S_1)$. Note that there exists a nonzero polynomial $Q \in$ F$[y][z, x]$ with $deg_z(Q), deg_x(Q) \leq \sqrt{l}$ such that

$$Q(C_a(y), a) = 0 \qquad \forall a \in S_1 \tag{11}$$

The reason is that if we let $Q_{ij} \in$ F$[y]$ be the coefficient of $z^j x^i$ in Q, then the constraints in (11) define a homogeneous system of linear equations over F$[y]$ with $(1 + deg_x(Q))(1 + deg_z(Q)) > l$ unknowns and l constraints.

$$\sum_{i=0}^{\sqrt{l}} \sum_{j=0}^{\sqrt{l}} Q_{ij}(C_a(y))^j a^i \;\; = \;\; 0 \qquad \forall a \in S_1$$

Since the number of unknowns exceeds the number of constraints, a nontrivial solution exists.

Now we claim that we can find a nontrivial solution Q that in addition is in F$[y][z, x]$ and satisfies $deg_y(Q) \leq dl^{3/2}$. To do this, apply Cramer's Rule on the system of l constraints. Doing this carefully yields a solution in which each coordinate (i.e., coefficient Q_{ij}) is a signed determinant of some square submatrix of this system. (This simple fact is worked out as Fact 25 in [6].) In general, the determinant of a $k \times k$ matrix is a polynomial of degree k in the matrix entries. Here the matrix entries are themselves degree $d\sqrt{l}$ polynomials in F$[y]$, so the determinant of a submatrix is a polynomial of degree at most $dl^{3/2}$ in y. Hence $deg_y(Q) \leq dl^{3/2}$.

Finally we show that Q satisfies condition (10) using Lemma 19. For every $b \in S_2$, $\Gamma_b(x, z) = Q(z, x, b)$ is a bivariate polynomial of degree $\sqrt{l}$ in x and z. Further $\Gamma_b(x, z)$ is zero on all the l points $\{(a, z_a = C_a(b)) | a \in S_1\}$. Further, there exists ρl points (where $C_a(b) = R_b(a)$) where $z_a = R_b(a)$ for some degree d polynomial R_b. Since $\rho > (d+1)/\sqrt{l}$, Lemma 19 shows that $z - R_b(x)$ divides $\Gamma_b(x, z)$. □

Now we return to the proof of Theorem 2 for $m = 2$. The following lemma finishes Step 1 of our proof.

Lemma 21 *There exist constants e_1, e_2 such that the following is true: Let $f : F^2 \to F$ have d-success rate at least δ, let $t = e_1(d+1)^2/\delta^6$. Then there exists a non-zero polynomial $Q \in F[z, x, y]$ of total degree at most $(d+1)t^{3/2}$ and a set of points $S \subseteq F^2$ containing at least $(\delta^9/e_2)\,|F|^2$ points such that*

1. *$Q(f(a, b), a, b) = 0 \qquad \forall (a, b) \in S$.*

2. *The d-success rate of f at each point in S is at least $\delta/2$*

Proof: Note that the lemma is vacuously true if $q < t$. So we assume $q \geq t$ (which implies in particular that $q \geq 18/\delta$, as will be used later), and prove the lemma for $e_1 = 36^2$ and $e_2 = 2 \cdot 18^3$. The main idea is to use averaging to show that the coordinate system can be rotated so that with respect to the new x and y axes, the conditions of Lemma 20 are satisfied for $\rho = \Omega(\delta^6)$. Lemma 20 yields a polynomial Q' of low-degree in the new variables, satisfying some nice properties. Rotating our axes back to the old x and y yields the polynomial Q of the lemma. Note that we are interested only in the total degree of Q and this coordinate rotation does not affect the total degree. Below we argue the details.

For $h \in \mathrm{F}^2$, $h \neq 0$, a "line in direction h," is one of the form $\{(u + \theta \cdot h) : \theta \in \mathrm{F}\}$. We consider two directions h_1 and h_2 to be the same if they satisfy $h_1 = \theta \cdot h_2$ for some $\theta \in \mathrm{F} - \{0\}$. Note that there are exactly $q + 1$ distinct directions. Further, note that for a point $x \in \mathrm{F}^2$ and direction h, there is exactly one line in direction h passing through x.

A point $x \in \mathrm{F}^2$ is *good* for a pair of distinct directions (h_1, h_2) if the line polynomials $P_d^f(l_1)$ and $P_d^f(l_2)$ correctly describe f at x, where l_1, l_2 are the lines that pass through x and lie in directions h_1 and h_2 respectively.

Let $G \subseteq \mathrm{F}^2$ consist of points where the d-success rate of f is at least $\delta/2$. Since the overall success rate is at least δ, averaging (Lemma 9) shows G contains at least $\delta/2$ fraction of F^2. (Looking ahead, the set S that we output finally will be some subset of G, and thus will satisfy the second condition of the lemma immediately.)

Claim 1: *Let $\delta_1 = \delta^3/9$. There exist two directions h_1, h_2 and a set of points $H \subseteq G$ with size $|H| \geq \delta_1 |F|^2$ such that every point in H is good for (h_1, h_2).*

Proof: We use the probabilistic method and show that a random point $x \in \mathrm{F}^2$ is good for a random pair of distinct directions (h_1, h_2) with high probability. (We will use below that q is sufficiently large. In particular $\delta q \geq \delta t \geq (36(d+1))^2 \geq 18$, where the last constant is just what suffices for the inequalities below.)

$$\begin{aligned}
\Pr_{x,h_1,h_2}[x \in G \bigwedge x \text{ is good for } (h_1,h_2)] &= \Pr_x[x \in G] \times \Pr_{x,h_1,h_2}[x \text{ is good for } (h_1,h_2) \mid x \in G] \\
&\geq \frac{\delta}{2} \times \Pr_{x,h_1,h_2}[x \text{ is good for } (h_1,h_2) \mid x \in G] \\
&\geq \frac{\delta}{2} \times \frac{\frac{\delta(q+1)}{2}(\frac{\delta(q+1)}{2} - 1)}{q(q+1)} \\
&\geq \frac{\delta^3}{8}(1 - \frac{2}{\delta q}) \\
&\geq \frac{\delta^3}{9} \quad \text{(Using } \delta q \geq 18\text{)}.
\end{aligned}$$

In other words, when a pair of directions (h_1, h_2) is picked randomly, then the expected size of the set $\{x \in \mathrm{F}^2 : x \in G \bigwedge x \text{ is good for } (h_1, h_2)\}$ is at least $\delta^3 |\mathrm{F}|^2 /9$. Hence there exists a pair of directions for which this set has size at least $\delta^3 |\mathrm{F}|^2 /9$. Thus Claim 1 is proved. □

Let δ_1, h_1, h_2, H be as in Claim 1. Rotate the coordinates so that h_1 becomes the x-axis and h_2 becomes the y-axis. From now on, coordinates are stated in this new system. We use *columns* and *rows* to refer to lines parallel to the y and x axes respectively.

For $a, b \in \mathrm{F}$ let R_b and C_a denote the line polynomials in the row $\{(x, b) : x \in \mathrm{F}\}$ and the column $\{(a, y) : y \in \mathrm{F}\}$ respectively. By the defining property of H, if $(a, b) \in H$, then $C_a(b) = R_b(a) = f(a, b)$. Next we show how to select sets $S_1, S_2 \subseteq \mathrm{F}$ that satisfy the hypothesis of Lemma 20 *and* some additional properties. In particular they satisfy the properties for $\gamma = \delta_1/2$:

1. $|S_1| = t$.

2. $|S_2| \geq (\gamma^2/2)q$.

3. For every $b \in S_2$, at least $(\gamma/2)t$ values of $a \in S_1$ satisfy $(a, b) \in H$.

4. For every $b \in S_2$, at least γq values of $a \in \mathrm{F}$ satisfy $(a, b) \in H$.

By hypothesis, $\Pr_{a,b\in\mathrm{F}}[(a, b) \in H] \geq 2\gamma$. Let W be the set of rows

$$W = \left\{ b \in \mathrm{F} : \Pr_{a\in\mathrm{F}}[(a, b) \in H] \geq \gamma \right\}.$$

S_2 will be chosen to be a subset of W and this will yield property (4) from the above list.

Averaging (Lemma 9 again) shows that W contains at least γ fraction of all rows. Let S_1 denote the indices of the t columns that contain the most elements in H from among the rows of W. (Thus S_1 satisfies property (1).) Clearly, $\Pr_{a\in S_1,b\in W}[(a, b) \in H] \geq \Pr_{a\in\mathrm{F},b\in W}[(a, b) \in H] \geq \gamma$.

Now let

$$S_2 = \left\{ b \in W : \Pr_{a\in S_1}[(a, b) \in H] \geq \gamma/2 \right\}.$$

By definition S_2 satisfies property (3). Averaging shows that S_2 contains at least $\gamma/2$ fraction of the rows in W. In other words, S_2 contains at least a $\gamma^2/2$ fraction of all rows, and thus property (2) is satisfied.

By properties (1) and (3) above, we have proven the existence of $S_1, S_2 \subseteq \mathrm{F}$ such that they satisfy the hypothesis of Lemma 20 with $\rho = \gamma/2$ and $l = t$. (Notice that by the condition $t \geq (36(d+1)/\delta^3)^2$, we have that $\rho \geq (d + 1)/\sqrt{t}$ as required by Lemma 20.) Let $Q \in \mathrm{F}[z, x, y]$ be the polynomial given by Lemma 20. Then $deg_x(Q), deg_z(Q) \leq \sqrt{t}$ and $deg_y(Q) \leq dt^{3/2}$, and thus the total degree of Q is at most $2\sqrt{t} + dt^{3/2} \leq (d + 1)t^{3/2}$ (using $t \geq 2$, which is certainly true).

The polynomial Q is the one that the lemma statement guarantees to exist. To finish we need to define the set S mentioned in the lemma. Let

$$S = \left\{(a, b) \in \mathrm{F}^2 : b \in S_2 \text{ and } (a, b) \in H\right\}.$$

Since $S \subseteq H \subseteq G$, part (2) of the conclusion is true, i.e., the d-success rate for every $(a, b) \in S$ is at least $\delta/2$. It remains to show that $|S| \geq \frac{\delta^9}{e_2} |\mathrm{F}|^2$ and that $Q(f(a, b), a, b) = 0$ for every $(a, b) \in S$.

Since each row $b_2 \in S_2$ has at least γ fraction of its points in H and $|S_2| > \gamma^2 |\mathrm{F}| /2$, we have

$$|S| \geq \frac{\gamma^3}{2} |\mathrm{F}|^2 = \frac{\delta^9}{2 \cdot 18^3} |\mathrm{F}|^2 = \frac{\delta^9}{e_2} |\mathrm{F}|^2 .$$

Finally, let $(a, b) \in S$. Since $b \in S_2$, the property of Q implies $(z - R_b(x)) \mid Q(z, x, b)$ and so $Q(R_b(x), x, b) = 0$. Since $(a, b) \in H$, we have $f(a, b) = R_b(a)$. Hence $Q(f(a, b), a, b) = 0$. Thus the lemma has been proved. □

Now we move to Step 2 of our argument. The main technical tool used in the lemma is Theorem 34 on "effective Hilbert irreducibility" whose statement and proof are deferred to the appendix.

The next lemma shows how to exploit the consequence of Lemma 21 to complete the analysis of the low-degree test. While we need the Lemma only for the case of bivariate polynomials ($m = 2$), we prove it for general m, since the more general lemma will be useful later.

Lemma 22 *Let $f : F^m \to F$ be a function, d be a positive integer, $Q \in F[z, x_1, \ldots, x_m]$ be a polynomial of total degree $D \geq d$ and $S \subseteq F^m$ be a set of points of size at least $\gamma \cdot |F|^m$ such that: (i) $\forall \hat{a} \in S$, $Q(f(\hat{a}), \hat{a}) = 0$. (ii) The d-success-rate of f at every point in S is at least γ.*

If $|F| > 12D^4/\gamma^2$, then there is a degree d m-variate polynomial $g \in F[x_1, \ldots, x_m]$ that has agreement at least $\gamma^4/(8D)$ with f and such that $z - g(x, y)$ is a factor of Q.

Proof: The main idea is to use Lemma 19 to show that the restriction of Q on "many" lines has a linear factor that describes f on "many" points of that line. Then we will use Theorem 34 on "effective Hilbert irreducibility" to conclude that Q itself must have a linear factor that describes f in "many" points.

We say a point $\hat{a} \in \mathrm{F}^m$ is *nice* for a line l in F^m if (i) $\hat{a} \in S$ and (ii) $P_d^f(l)$, the line polynomial of l, describes f at $\hat{a}$. We say that a line l is *nice* if at least $\gamma^2/2$ fraction of the points on it are nice for l.

Claim 1: *At least $\gamma^2/2$-fraction of lines are nice.*
Proof: We will show that the expected fraction of nice points on a random line is γ^2. The claim follows by averaging (Lemma 9).

Imagine reversing the order in which the point and the line are picked: first pick a point $\hat{a}$ randomly and then a random line l that passes through it. With probability at least γ, the point is in S and with a further probability at least γ, line polynomial through l describes f at the point. Thus the probability that the point is nice for l is at least $\gamma \cdot \gamma = \gamma^2$. By linearity of expectations, it follows that the expected fraction of nice points on a random line is at least γ^2. □

The next claim relates nice lines to factorization of restrictions of Q. For a line l let us denote the restriction of Q to l by $Q|_l \in \mathrm{F}[z, t]$, where t is the line parameter. In proving the claim, we use the fact that $q > 4Dd/\gamma^2$.

Claim 2: *For a nice line l, $Q|_l(z, t)$ has a factor of the form $z - h(t)$ where $h \in F[t]$ is a polynomial that agrees with f on at least $(\gamma^2/2)q$ points on the line.*
Proof: Let h be the line polynomial for l. Since $\gamma^2/2$ fraction of the points on l are nice for l, it follows that $h(t)$ has agreement $\gamma^2/2$ with f on l. Since $(\gamma^2/2)q > 2Dd$, we can apply Lemma 19 and conclude that $z - h(t)$ divides $Q|_l$. The claim follows. □

Since many lines are nice, the Hilbert Irreducibility Theorem (Theorem 34) now allows us to conclude that Q has a linear factor (i.e., of the form $z - g(\hat{x})$), as argued in the next claim. The theorem only states the existence of such a factor with no guarantees on its degree (could be as large as D), or its coefficients (which could lie in the algebraic closure of F, rather than F itself), or about the agreement of f with g. The claim below also establishes these additional details.

Let $\overline{\mathrm{F}}$ denote the algebraic closure of F and let $Q_1, \ldots, Q_k \in \overline{\mathrm{F}}[z, \hat{x}]$ be all the distinct factors (over the algebraic closure of field F) of Q that involve z. Note that $k \leq D$. For a line l let us denote the restriction of Q_i to l by $Q_i|_l \in \mathrm{F}[z, t]$, where t is the line parameter. In proving the claim we use the fact that $q > 12D^4/\gamma^2$.
Claim 3: *One of the Q_i's is of the form $c(z - g(\hat{x}))$ where $c \in F$, and $g \in F[\hat{x}]$ has degree at most d and has agreement at least $\gamma^4/(8D)$ with f.*
Proof: We first prove a weaker claim, where g is allowed to take coefficients from $\overline{\mathrm{F}}$ and allowed to have arbitrary degree.

We associate with every nice line l a factor Q_i of Q as follows: Let the line polynomial for l be $h_l(t)$. We find an i such that $Q_i|_l$ has $z - h_l(t)$ as a factor (such a factor does exist by Claim 2) and associate Q_i with l. (Ties are broken arbitrarily.)

As a first step, we show that Q has a linear factor (i.e., of the form $c(z - g(\hat{x}))$) that has many nice lines associated with it. We first invoke Theorem 34 to conclude that any non-linear factor of Q has at most a $(3D^3/q)$-fraction of lines associated with it. Using $3D^3/q < \gamma^2/(4D)$ and the fact that there are at most D non-linear factors of Q, we conclude that the non-linear factors have at most $\gamma^2/4$ fraction of lines associated with them. Thus Q must have some linear factors and in particular one with $\gamma^2/(4D)$-fraction of lines associated with it (since it can have at most D linear factors). Let this factor be $Q_1 = c(z - g(\hat{x}))$. We now show that g is the polynomial as asserted to exist by the claim.

First note that g has large agreement with f. In particular if we pick a random line and then a random point on the line, then we get that this line is a nice line associated with Q_1 with probability at least $\gamma^2/(4D)$. Conditioned on this line being nice, the point is nice for the line l with probability at least $\gamma^2/2$ and in such a case g agrees with f at this point. Combining the bounds we get that f and g agree with probability at least $\gamma^4/(8D)$, as desired.

Now we claim that g has coefficients in F and has degree at most d. (Thus far we only know that $g \in \overline{\mathrm{F}}[\hat{x}]$ and has degree at most D.) To this end, note that on a nice line l associated with Q_1, g restricted to the line l is a degree d polynomial with coefficients in F. Since $\gamma^2/(4D)$ fraction of lines are nice and associated with Q_1, we get: *(a) g has all its coefficients in F.* The reason is that that

the restriction of g on at least $\frac{1}{D} \cdot \frac{\gamma^2}{4}$ fraction of lines is in $\mathrm{F}[t]$ and $\frac{\gamma^2}{4D} > D/|\mathrm{F}|$. Thus Lemma 8 implies that g has no coefficients in $\overline{\mathrm{F}} \setminus \mathrm{F}$.
(b) g is a degree d polynomial. The reason is its restriction to at least $\frac{\gamma^2}{4D}$ fraction of the lines is a degree d polynomial and $\frac{\gamma^2}{4D} > D/|\mathrm{F}|$. Thus Lemma 6 implies its degree is at most d. □

Thus concludes the proof of the lemma. □

The bivariate case of Theorem 1 now follows easily.

Theorem 23 *There are constants $c_0 < \infty$ and $c_1 > 0$ such that the following is true. Let $F = GF(q)$ and $f : F^2 \to F$ be a function that has d-success rate at least δ. If $q > c_0(d+1)^{16}/\delta^{54}$, then there is a bivariate degree d polynomial g that has agreement at least $c_1\delta^{45}/(d+1)^4$ with f.*

Proof: By Lemma 21 we get that there exists a trivariate polynomial Q of degree at most $e_1^{3/2}(d+1)^4/\delta^9$ and a set $S \subseteq \mathrm{F} \times \mathrm{F}$ of density δ^9/e_2 such that (a) $Q(f(a,b),a,b) = 0$ for all $(a,b) \in S$, and (b) The d-success rate of each point in S is at least $\delta/2$. Thus we may apply Lemma 22 to f, S and Q with $\gamma = \delta^9/e_2$ and $D = e_1^{3/2}(d+1)^4/\delta^9$ to find that there exists a bivariate polynomial g that has $\gamma^4/(8D)$ agreement with f, provided $|\mathrm{F}| > 12D^4/\gamma^2$. Substituting the values of D and γ we find that we require $|\mathrm{F}| > 12e_1^6 e_2^2(d+1)^{16}/\delta^{54}$, and if this is satisfied, we get an agreement of at least $\frac{1}{8e_1^{3/2}e_2^4}\frac{\delta^{45}}{(d+1)^4}$. The theorem follows with for $c_0 = 12e_1^6 e_2^2$ and $c_1 = \frac{1}{8e_1^{3/2}e_2^4}$. □

Since we have proved the bivariate case of the low degree test, the bivariate cases of Theorems 16 and 17 now follow.

Corollary 24 *There exist constants c_0, c_1 such that if δ, d and F satisfy $|F| \geq c_0((d+1)/\delta)^{c_1}$, then if $f : F^2 \to F$ has d-success rate δ, then it has agreement at least $2\delta/3$ with some degree d polynomial.*

Corollary 25 *There exist constants c_0, c_1 such that if γ, d and F satisfy $|F| \geq c_0((d+1)/\gamma)^{c_1}$, then for every function $f : F^2 \to F$ and every d-oracle, there exists a set of at most $\frac{4}{\gamma}$ polynomials $Q_1, \ldots, Q_k$ such that the probability that the low-degree picks a point where $f(x)$ does not equal $Q_i(x)$ for any $i \in \{1, \ldots, k\}$ and the low-degree test accepts is at most γ.*

3.2 The Trivariate Case

We restate Theorem 2 for the case $m = 3$ and prove it. The proof is a minor modification of the proof for $m = 2$.

Lemma 26 *There exist constants c_0, c_1, c_2, c_3 such that for all δ, d and F such that $|F| \geq c_0((d+1)/\delta)^{c_1}$ if $f : F^3 \to F$ has d-success-rate at least δ, then f has agreement at least $\frac{1}{c_2}(\delta/(d+1))^{c_3}$ with some degree d polynomial.*

Proof: The steps of the proof are similar to the corresponding steps in the bivariate case. Again, we first perform a random transformation of the coordinates. We identify three directions h_1, h_2 and h_3 in F^3 and for all $z \in \mathrm{F}^3$ call all lines of the form $\{z + th_1 | t \in \mathrm{F}\}$ as *vertical lines*, all lines of the form $\{z + t(h_2 + vh_3) | t \in \mathrm{F}\}$ for any $v \in \mathrm{F}$ as *horizontal* lines.

By averaging we know that the d-success-rate of f is at least $\delta/2$ for at least $\delta/2$ fraction of the points. Let G denote this set of points.

Let $\delta_1 = \delta^3/64$. We say a point $z \in \mathrm{F}^3$ is *good* for directions h_1, h_2, h_3 if it satisfies

- The vertical line through the point passes the low-degree test at z (i.e., the line polynomial for that line describes f at z).
- δ_1 fraction of the horizontal lines through the point pass the low-degree test at z.

Claim 1 *If $q \geq 64/\delta^2$, then there exist directions h_1, h_2 and h_3 s.t. δ_1 fraction of the points in F^3 are in G and are good for h_1, h_2, h_3.*

Proof: We use the probabilistic method. We randomly pick three directions h_1, h_2, h_3 and a point $z \in \mathrm{F}^3$ and show that

$$\Pr_{z,h_1,h_2,h_3}[z \in G \bigwedge z \text{ is good for } h_1, h_2, h_3 \bigwedge h_1, h_2, h_3 \text{ are noncoplanar}] \geq \delta_1,$$

whence the existence of the desired three directions follows.

Note that $\Pr_z[z \in G] \geq \delta/2$, so it suffices to show that

$$\Pr_{z,h_1,h_2,h_3}[h_1, h_2, h_3 \text{ are noncoplanar} \bigwedge z \text{ is good for } h_1, h_2, h_3 \mid z \in G] \geq 2\delta_1/\delta = \delta^2/32.$$

We will actually show that the above holds for every $z \in G$. Now, if we pick a random line through z, then it passes the low degree test at z with probability at least $\delta/2$. So if we pick random directions h_1, h_2, h_3 and then a random $v \in \mathrm{F}$, then

$$\Pr_{h_1,h_2,h_3,v}[h_1, h_2, h_3 \text{ are noncoplanar} \bigwedge \text{line } \{z + t(h_2 + vh_3) : t \in \mathrm{F}\} \text{ passes the test at } z] \geq \delta/2 - \frac{2}{q},$$

where the "$2/q$" term upperbounds the probability that that h_1, h_2, h_3 are coplanar.

Hence we conclude by averaging that for at least $\delta/4 - 1/q$ choices of h_2, h_3, the fraction of $v \in \mathrm{F}$ for which this event happens is at least $\delta/4 - 1/q$.

Thus for all $z \in G$,

$$\Pr_{h_1,h_2,h_3}[z \text{ is good for } h_1, h_2, h_3] \geq (\frac{\delta}{4} - \frac{1}{q})^2 \geq \delta^2/16 - \frac{2}{q} \geq \frac{\delta^2}{32},$$

where the last inequality uses $q \geq 64/\delta^2$. □

From now on we assume that h_1, h_2 and h_3 as guaranteed above have been found and that the coordinates have been transformed so that $h_1 = (1, 0, 0)$, $h_2 = (0, 1, 0)$ and $h_3 = (0, 0, 1)$. The set of points of the form $\{(a, x, y) | x, y \in \mathrm{F}\}$ will be called the *horizontal plane at height a*. The set of points $\{(w, b, c) : w \in \mathrm{F}\}$ will be called the *vertical line through* (b, c). For each vertical line through (b, c) let $V_{b,c}(w)$ denote the line polynomial for that line.

Let c'_0, c'_1 be the two constants given by Corollary 24. Below we use the fact that $q > c'_0(2^{13}(d + 1)/\delta^4)^{c'_1} \geq c'_0((d + 1)/(\delta_1^2/2))^{c'_1}$, and thus Corollary 24 can be applied with $\delta' = \delta_1^2/2$ (to some appropriately chosen functions).

Claim 2 *Let $\delta_2 = \delta_1^4/8$. Suppose $q > \max\{c'_0((d+1)/(\delta_1^2/2))^{c'_1}, 12(d+1)/\delta_1^2\}$. Then, for each $a \in F$ we can associate a bivariate polynomial $H_a[x, y]$ with the horizontal plane at height a such that the following is true. For at least δ_2 fraction of the points $(a, b, c) \in F^3$:*

- *$f(a, b, c) = V_{b,c}(a) = H_a(b, c)$.*
- *δ_2 fraction of all lines through (a, b, c) pass the low-degree test at (a, b, c).*

Proof: Let S denote the set of points referred to in Claim 1 (i.e., are in G and are good for $h_1, h_2.h_3$; thus $S/|\mathrm{F}|^3 \geq \delta_1$. Let a function $h : \mathrm{F}^2 \to \mathrm{F}$ have agreement less than $(d+1)/q$ with every bivariate degree d polynomial (its existence is proved in Proposition 5). On each horizontal plane, replace the values of f on $\mathrm{F}^3 \setminus S$ by using values of h and let $g : \mathrm{F}^3 \to \mathrm{F}$ be the new function thus obtained. (I.e., $g(a, b, c) = h(b, c)$, if $(a, b, c) \notin S$.) Note that at every point of S, the low degree test still accepts the function g, with the d-oracle used for f, on at least δ_1 fraction of horizontal lines. We conclude that the average d-success rate of g among horizontal planes is at least δ_1^2. Say that a horizontal plane is *nice* if the average d-success rate of lines in it is at least $\delta_1^2/2$. Averaging shows that at least $\delta_1^2/2$ of the horizontal planes are nice. The correctness of the low-degree test for the bivariate case (Corollary 24) implies that the restriction of g on a nice plane has agreement at least $\delta_1^2/3$ with a degree d bivariate polynomial. For $a \in \mathrm{F}$, let $H_a[x, y]$ denote the bivariate degree d polynomial that has maximum agreement (break ties arbitrarily) with g on the horizontal plane at height a. In a nice plane, all but $2/q$ of this agreement must happen on points in S, since values of g outside

S have agreement at most $(d+1)/q$ with every degree d polynomial. Using $(d+1)/q < \delta_1^2/12$, we conclude that on nice planes, H_a agrees with f on at least $\delta_1^2 q/4$ points of S.

Notice that we have shown that if (w,x,y) is a random point then with probability at least $(\delta_1^2/2)(\delta_1^2/4) = \delta_2$ we have: (a) the horizontal plane at height w is nice (b) $(w,x,y) \in S$ and $f(w,x,y) = H_w(x,y)$. This proves the claim. $\Box$

Call a point in F^3 *great* if it satisfies the properties listed in the Claim 2. Call a vertical line *fine* if at least $\delta_2/2$ fraction of the points on it are great. Since every point is on a unique vertical line, averaging shows that at least $\delta_2/2$ fraction of the vertical lines are fine.

Claim 3 *Let $l < q\delta_2^2/8$, and $\delta_3 = \delta_2^2/16$. There exists a set $U \subseteq F$ of size l such that for at least δ_3 fraction of $(b,c) \in F^2$, the following is true: the vertical line through (b,c) is fine and contains at least $\delta_3 l$ great points of the type (a,b,c) where $a \in U$.*

Proof: As already noted,

$$\Pr_{a,b,c}[\text{the line through } (b,c) \text{ is fine} \bigwedge (a,b,c) \text{ is great}] \geq \frac{\delta_2}{2} \cdot \frac{\delta_2}{2} = \delta_2^2/4.$$

Averaging shows that for at least $\delta_2^2/8$ fraction of $a \in \mathrm{F}$,

$$\Pr_{b,c}[\text{the line through } (b,c) \text{ is fine} \bigwedge (a,b,c) \text{ is great}] \geq \delta_2^2/8. \tag{12}$$

Let $U \subseteq \mathrm{F}$ be the points a satisfying the above. Then we have $|U| \geq l$ as required. Further averaging shows that at least $\delta_2^2/16$ fraction of (b,c) are such that the vertical line through (b,c) is fine and contains at least $\delta_2^2 l/16$ great points of the type (a,b,c) for $a \in U$. $\Box$

Claim 4 *Let $l \geq 4d^2/\delta_3^2$ and $l \leq q\delta_2^2/4$. Let $\delta_4 = \delta_3\delta_2/2$. There exists a polynomial $Q(w,x,y,z)$ with $\deg_w(Q) = \deg_z(Q) = \sqrt{l}$ and $\deg_x(Q) = \deg_y(Q) = dl^{3/2}$ such that at least δ_4 fraction of the points (w,x,y) satisfy the following properties:*

- *$Q(w,x,y,f(w,x,y)) = 0$.*
- *δ_4 fraction of all lines through (w,x,y) pass the low-degree test at (w,x,y).*

Proof: Let U denote the same set as in Claim 3. First we show the existence of a polynomial $Q(w,x,y,z)$ of degree $\sqrt{l}$ in w and z and degree $dl^{3/2}$ in x and y such that for all $a \in U$, $Q(a,x,y,H_a(x,y)) = 0$. Consider the following sequence of l pairs from $\mathrm{F}[x,y] \times \mathrm{F}$: $\{(H_a(x,y),a) : a \in U\}$. By an argument identical to the one in Lemma 20, there exists a polynomial $\Gamma(z,w)$ in $\mathrm{F}[x,y][z,w]$ (i.e., bivariate polynomial in variables z,w whose coefficients are in $\mathrm{F}[x,y]$) such that $\Gamma(H_a(x,y),a) = 0 \ \forall a \in U$, and the degrees of Γ are as stated in the Claim. Rewrite Γ as a polynomial in $\mathrm{F}[w,x,y,z]$ and call it Q. We show that Q is the desired polynomial.

Suppose the vertical line through (b,c) is one of the δ_3 fraction of vertical lines mentioned in Claim 3. Since $l > 4d^2/\delta_3^2$, we have $\delta_3 l > 2d\sqrt{l}$ and so this line has at least $2d\sqrt{l}$ great points on it from $U \times \mathrm{F} \times \mathrm{F}$. Then $Q(w,b,c,V_{b,c}(w))$ is a polynomial in $\mathrm{F}[w]$ of degree at most $(d+1)\sqrt{l}$ and at least $2d\sqrt{l}$ roots. Hence Lemma 19 shows $Q(w,b,c,V_{b,c}(w)) = 0$. This means in particular that for every great point (a,b,c) on this line,

$$Q(a,b,c,V_{b,c}(a)) = Q(a,b,c,H_a(b,c)) = Q(a,b,c,f(a,b,c)). \tag{13}$$

Since (13) is true for every great point on a δ_3 fraction of lines, we conclude that the fraction of great points satisfying (13) is at least $\delta_3 \cdot \delta_2/2 = \delta_4$. Hence polynomial Q has the desired properties. $\Box$

Claim 5 *Let $\delta_5 = \delta_4^4/(16(d+1)l^{3/2})$. Suppose $q > \frac{192(d+1)^4 l^6)}{\delta_4^2}$. Then there exists a degree d polynomial $g(w,x,y)$ such that f and g have agreement at least δ_5.*

Proof: The claim follows from the trivariate case of Lemma 22 applied to the polynomial Q and the set U of Claim 4, with the setting $\gamma = \delta_4$ and $D = 2(d+1)l^{3/2}$ which is clearly an upper bound on the total degree of Q. □

Thus we have shown that if f has d-success rate δ, then it has agreement at least δ_5 with some degree d polynomial provided

$$|\mathrm{F}| > \max\left\{\frac{64}{\delta^2}, \frac{c_0'(d+1)}{(\delta_1^2/2))^{c_1'}}, \frac{12(d+1)}{\delta_1^2}, \frac{128d^4}{\delta_3^2\delta_2^2}, \frac{192(d+1)^4l^6)}{\delta_4^2}\right\},$$

where $\delta_1, \ldots, \delta_5$ are also polynomials in $(d+1)/\delta$. Thus we find that it $|\mathrm{F}| > c_0((d+1)/\delta)^{c_1}$, then the agreement is at least $\frac{1}{c_2}(\delta/(d+1))^{c_3}$ for appropriate choice of constants c_0, c_1, c_2, c_3. □

Since we have proved the trivariate case of the low degree test, the trivariate cases of Theorem 16 and 17 now follow.

Corollary 27 *There exist constants c_0, c_1 such that if δ, d and F satisfy $|F| \geq c_0((d+1)/\delta)^{c_1}$, then if $f : F^3 \to F$ has d success rate δ, then it has agreement $\delta/2$ with some degree d polynomial.*

Corollary 28 *There exist constants c_0, c_1 such that if γ, d and F satisfy $|F| \geq c_0((d+1)/\gamma)^{c_1}$, then for every function $f : F^3 \to F$ and any d-oracle, there exists a set of at most $\frac{4}{\gamma}$ polynomials $Q_1, \ldots, Q_k$ such that the probability that the test picks a point x where $f(x)$ does not equal $Q_i(x)$ for any $i \in \{1, \ldots, k\}$ and the low-degree test accepts is at most γ.*

3.3 The Bootstrapping

The title of this section refers to the fact that we will assume the truth of Theorem 1 (as well as Theorems 16 and 17) for $m = 2, 3$, and then prove it for general m. We rely on symmetry-based arguments similar to those in [3]. These use the notion of a k-dimensional subspace of F^m.

We try to define a function $\hat{f}$ that we hope is "almost" a polynomial and has significant agreement with f. The definition of $\hat{f}$ is probabilistic; the hope is to show that with high probability this function will suffice. We pick a line l randomly and define $\hat{f}$ as a function of this line.

Definition 4 **($\hat{f}_l$)** For any line l we define a function $\hat{f}_l : \mathrm{F}^m \to \mathrm{F}$ as follows. Let $P_d^f(l)$ denote the univariate degree d polynomial that best describes f's restriction to l (see Definition 2). Now consider every plane s that contains l. (Note: since every point $x \notin l$ determines a unique plane with l, the set of planes containing l form a partition of $\mathrm{F}^m \setminus l$.) Check whether there is a bivariate polynomial, say g, that agrees with $P_d^f(l)$ on line l and that has agreement at least $(.45)\delta$ with f on plane s. If so, for every point $y \in s$, we define $\hat{f}_l(y)$ to be the value taken by g at y. If no such bivariate polynomial exists, we define $\hat{f}_l(y)$ arbitrarily in this plane (except on l).

Lemma 29 *There are constants r, s such that the following is true Let $m > 3$ and $\eta > 0$. Let $f : F^m \to F$ have d-success-rate at least δ, and $q = |F| > r\left(\frac{d+1}{\eta\delta}\right)^s$. If a line l is picked randomly, then*

$$E_l[d\text{-success-rate of } \hat{f}_l \text{ in } F^m] \geq 1 - \eta \quad (14)$$

$$E_l[\text{agreement between } f \text{ and } \hat{f}_l \text{ in } F^m] \geq \frac{\delta}{4}. \quad (15)$$

We defer the proof of this lemma for later. We first show how the lemma allows us to prove Theorem 1 almost immediately. The main idea is that it reduces the analysis of the low-degree test to the earlier studied settings of functions that have success rate close to one [30, 5]. In particular, we use the following result of [5].

Theorem 30 ([5]) *There exists $\eta_0 > 0$ and a polynomial $p : Z^+ \to Z^+$ such that for every $0 < \eta \leq \eta_0$ if $f : F^m \to F$ has d-success rate at least $1 - \eta$, then it has agreement at least $1 - 2\eta$ with some degree d polynomial, provided $|F| \geq p(d)$.*

We now show how the general case of Theorem 1 follows from Lemma 29 and Theorem 30.

Proof:(of Theorem 1; $m > 3$) We invoke Lemma 29 with $\eta = \min\{\frac{\delta\eta_0}{10}, \frac{\delta^2}{240}\}$, where η_0 is the constant given by Theorem 30. We first show that if we pick a line l at random, then with nonzero probability, we get a line such that the polynomial closest to $\hat{f}_l$ has agreement at least $\delta/24$ with f.

Using an averaging argument along with statement (14) we see that for any k,

$$\Pr_l[d\text{-success-rate of } \hat{f}_l \geq 1 - k\eta] \geq 1 - \frac{1}{k}$$

Using averaging on (15) we see that

$$\Pr_l[\text{agreement between } f \text{ and } \hat{f}_l > \frac{\delta}{8}] > \frac{\delta}{8}.$$

We let $k = 10/\delta$, and conclude that with probability $\delta/8 - \delta/10$ the following two events happen (i) d-success-rate of $\hat{f}_l > 1 - k\eta$ and (ii) the agreement between f and $\hat{f}_l$ is at least $\delta/8$. In particular, there exists at least one line l for which the two events in the preceding paragraph happen. Let l_0 be such a line. Since $k\eta < \eta_0$, we can apply Theorem 30 to conclude that $\hat{f}_{l_0}$ has agreement at least $1 - 2k\eta \geq 1 - \delta/12$ with some degree d polynomial g (provided $q \geq \text{poly}(d)$). Since $\hat{f}_{l_0}$ and f have agreement at least $\delta/8$, we conclude that f and g must have agreement of at least $\delta/8 - \delta/12 = \delta/24$ as claimed.

To now conclude the theorem it suffices to observe that the requirement on q (in particular we need to be able to apply Lemma 29) is of the form $q > c_0((d+1)/\delta)^{c_1}$, and the final agreement $\delta/24$ is also of the desired form. $\square$

Now we prove Lemma 29, which is the heart of our "bootstrapping."

Proof: (Lemma 29) The proof below uses some parameters that are functions of η and δ. We first set the values of these parameters. Let c_0, c_1 be constants given by Corollary 28 (second strong form of trivariate low-degree testing). Let $\gamma = \min\left\{\frac{\eta^2}{64}, \frac{1}{18^2}\right\}$. Let $\epsilon = \frac{\gamma\delta}{80}$. In the proof below we assume

$$q > \max\{\frac{16}{\eta}, \frac{1200}{\delta^2\eta}, \frac{48}{\delta}, \frac{7200}{\delta^3}, \frac{2(d+1)}{\epsilon}, \frac{8d}{\epsilon^2}, c_0\left(\frac{d+1}{\epsilon}\right)^{c_1}, \frac{32d}{\epsilon^2\gamma}, \frac{1600}{\epsilon\delta\gamma}, \frac{20d}{\epsilon^2\delta}\}.$$

Note that this is an assumption of the form $q > r((d+1)/\eta\delta)^s$ and thus consistent with the lemma statement.

By linearity of expectations it suffices to show that if we pick a pair of lines (l, l') randomly in F^m, then

$$\begin{aligned} & A = E_{(l,l')}[d\text{-success-rate of } \hat{f}_l \text{ on } l'] && \geq && 1 - \eta && (16) \\ \text{and} \qquad & B = E_{(l,l')}[\text{agreement of } \hat{f}_l \text{ and } f \text{ on } l'] && \geq && \frac{\delta}{4}. && (17) \end{aligned}$$

The main observation behind the "bootstrapping" is that since a pair of random lines are noncoplanar with probability at least $1 - 2/q$ (see Lemma 10), picking them is almost equivalent to picking a random cube C in F^m. Furthermore, for a random cube C, with probability at least $1 - \frac{300}{\delta^2 q}$, the d-success-rate of f in this cube is at least $\frac{9}{10}\delta$ (Lemma 12). Let us call such cubes *terrific*. Below we consider the probability of certain events when we pick two random noncoplanar lines (l, l') in C. We conclude:

$$\begin{aligned} A &\geq (1 - \frac{2}{q}) \cdot (1 - \frac{300}{\delta^2 q}) \cdot E_{C,(l,l')\in C}[d\text{-success-rate of } \hat{f}_l \text{ on } l' \mid C \text{ is terrific}] \\ &\geq E_{C,(l,l')\in C}[d\text{-success-rate of } \hat{f}_l \text{ on } l' \mid C \text{ is terrific}] - \frac{2}{q} - \frac{300}{\delta^2 q} \\ &\geq E_{C,(l,l')\in C}[d\text{-success-rate of } \hat{f}_l \text{ on } l' \mid C \text{ is terrific}] - \frac{\eta}{2}, \qquad (18) \end{aligned}$$

where the last inequality follows from $q \geq \frac{8}{\eta}, \frac{1200}{\delta^2\eta}$. A similar calculation for B using $q \geq \frac{48}{\delta}, \frac{7200}{\delta^3}$ shows that

$$B \geq E_{C,(l,l')\in C}[\text{agreement of } \hat{f}_l \text{ and } f \text{ on } l' \mid C \text{ is terrific}] - \frac{\delta}{12}.$$

Thus we see that the lemma is proved if we can prove the following (essentially trivariate) result.

Claim 0: *For a terrific cube C, the following are true.*

$$E_{(l,l'\in C)}[d\text{-success-rate of } \hat{f}_l \text{ on } l'] \geq 1 - \frac{\eta}{2} \tag{19}$$

$$E_{(l,l')\in C}[\text{agreement of } \hat{f}_l \text{ and } f \text{ on } l'] \geq \frac{\delta}{3} \tag{20}$$

We will prove Claim 0 after proving some preliminary claims below. From now on, C denotes a fixed terrific cube. By the trivariate case of Theorem 17, Corollary 27, there is a degree d trivariate polynomial that has agreement at least $\frac{6}{10}\delta$ with f on cube C. Let P_1 be one such polynomial and let $P_2, \ldots P_{k_0}$ be all the other degree d polynomials that have agreement at least $\frac{1}{2}\delta$ with f on cube C.

Let $P_{k_0+1}, \ldots, P_k$ be all the degree d polynomials whose agreement with f on cube C is between $\epsilon/2 \leq \epsilon - \frac{(d+1)}{q}$ and $\frac{1}{2}\delta$ (where the inequality follows from $q \geq 2(d+1)/\epsilon$). By Proposition 4, the set of polynomials we have identified thus far is not too big. Specifically, using $q > 8d/\epsilon^2$, we get $k_0 \leq 4/\delta$ and $k \leq 4/\epsilon$. Let $T \subseteq C$ be those points of C where f agrees with none of the polynomials. Recall, by Corollary 28, that the low degree test has low probability of succeeding at points in T, namely,

$$\Pr_{x\in C, l\ni x}[P_d^f(l) \text{ describes } f \text{ at } x \bigwedge x \in T] \leq \epsilon. \tag{21}$$

We hope to show ultimately that for "most" lines l, the function $\hat{f}_l$ has high agreement with one of $P_1, P_2, \ldots, P_{k_0}$. For any trivariate polynomial Q and line l, let $Q|_l$ denote its restriction to line l. We likewise define the restriction $Q|_s$ for a plane s. We say that line l is *nice* if the restrictions $P_1|_l, P_2|_l, \ldots, P_k|_l$ are all distinct and $P_d^f(l)$, the univariate degree d polynomial that has the highest agreement with f on l, is one of $P_1|_l, P_2|_l, \ldots, P_{k_0}|_l$.

We will use the parameter γ to denote the probability of some bad events. Recall by the definition of ϵ and q, we have

$$\gamma \geq \max\left\{\frac{32d}{\epsilon^2 q}, \frac{1600}{\epsilon\delta q}, \frac{80\epsilon}{\delta}\right\}.$$

Claim 1: *At least $1-\gamma$ fraction of the lines l in cube C are nice.*

Proof of Claim 1: For any fixed i and j, the probability the P_i agrees with P_j at a random point is at most d/q. In particular this implies that the probability with which they agree on a restriction to a random line l is also at most d/q. Thus the fraction of lines l for which $P_i|_l = P_j|_l$ for some $i \neq j$ is at most $\binom{k}{2} \times \frac{d}{q}$. Since $k \leq 4/\epsilon$, we have this probability is at most $\binom{k}{2} \times \frac{d}{q} \leq \frac{8d}{\epsilon^2 q} \leq \frac{\gamma}{4}$.

Now we estimate the fraction of lines for which $P_d^f(l)$ is not one of $P_1|_l, P_2|_l, \ldots, P_{k_0}|_l$. Such a line must satisfy one or more of the following properties.

1. $P_1|_l$ has agreement less than $(.55)\delta$ with f on line l: By Lemma 11, the fraction of such lines is at most $\frac{100}{\delta q} \leq \frac{\epsilon\gamma}{4} \leq \frac{\gamma}{4}$.

2. $P_1|_l$ has agreement $(.55)\delta$ with f on line l but one of $P_{k_0+1}|_l, \ldots, P_k|_l$ has agreement more than $(.55)\delta$: By Lemma 11, the fraction of such lines is at most $\frac{100}{\delta q} \times (k - k_0) \leq \frac{400}{\epsilon\delta q} \leq \frac{\gamma}{4}$.

3. $P_1|_l$ has agreement $(.55)\delta$ with f on line l but some univariate polynomial that is not $P_1|_l, P_2|_l, \ldots, P_k|_l$ has agreement more than $(.55)\delta$ with f on l: Note that at least $(.55)\delta - kd/q$ of this agreement happens on points in T. Since the success probability of f on points where it does not agree with $P_1|_l, \ldots, P_k|_l$ is at most ϵ (see (21)), the fraction of lines on which this success probability is more than $\beta - kd/q$ is at most $\frac{\epsilon}{(.55)\delta - kd/q}$. Using $q > \frac{2kd}{\delta}$, we get that this probability is at most $\delta \leq \frac{\epsilon}{(.05)\delta} \leq \frac{\gamma}{4}$.

Adding the probabilities of the events above, we find that the probability that a line is not nice is at most γ. □

We say that a plane s in C is *well-behaved* if (i) each of $P_1|_s, P_2|_s, \ldots, P_{k_0}|_s$ has agreement at least $(.45)\delta$ with f on s (ii) no bivariate polynomial that is not one of $P_1|_s, \ldots, P_k|_s$ has agreement more than $(.45)\delta$ with f on plane s.

Claim 2: *At least* $1-\gamma$ *fraction of planes in* C *are well-behaved.*

Proof of Claim 2: Each of $P_1, \ldots, P_{k_0}$ has agreement at least $\frac{1}{2}\delta$ with f on cube c. Picking a random plane is almost the same as picking three points x_1, x_2, x_3 at random from the cube and looking at all points of the form $S = \{(t_1x_1 + t_2x_2 + (1-t_1-t_2)x_3)|t_1, t_2 \in \mathrm{F}\}$. The only difference is that we require x_1, x_2, x_3 to be non-collinear. Along the lines of the proof of Lemma 10 we can shows that the probability that they are collinear is at most $\frac{1}{q^2} + \frac{1}{q^3} \leq \frac{1}{q}$. If we ignore the non-collinearity restriction on x_1, x_2, x_3 and consider the set of points S, then they form a pairwise independent collection of points and so we see (using Chebyshev's inequality) that:

$$\Pr_S[\exists i \text{s.t. agreement between } P_i|_S \text{ and } f \text{ on } S \text{ is } < (.45)\delta] \leq k\frac{100}{\delta q^2} < \frac{\gamma}{4}.$$

Thus, taking into account the probability of collinearity, we see that the probability that this doesn't happen for a randomly chosen plane s is at most $\frac{\gamma}{4(1-\frac{1}{q})} \leq \frac{\gamma}{2}$.

Next we bound the fraction of planes s such that some bivariate polynomial different from $P_1|_s, \ldots, P_k|_s$ has agreement at least $(.45)\delta$ with f on plane s. Let T be the set of points of C where f doesn't agree with any of $P_1, \ldots, P_k$. Consider the success rate of the low-degree test on points of T on a plane s. I.e., let α_s denote the maximum, over all d-oracles defined on lines of s, of probability that the low-degree test picks a point of T and accepts. Let α_C denote the analogous quantity over the whole cube C. Note that $\alpha_C = E_s[\alpha_s]$. (In particular, note that while the d-oracles do depend on the set T and the function f, their responses on individual lines need not depend on the plane in consideration. I.e., if s and s' are planes that contain l, then the d-oracle maximizing α_s and the d-oracle maximizing $\alpha_{s'}$ may be assumed to use the same response on the line l.)

Returning to our quest, let τ denote the fraction of planes s such that some bivariate polynomial other than $P_1|_s, \ldots, P_k|_s$ has agreement of at least $(.45)\delta$ with f. For such a plane s, since we have a polynomial that has agreement $(.45)\delta$ with f and at most $\frac{kd}{q}$ of this agreement can come from points not in T, we have $\alpha_s \geq (.45)\delta - \frac{kd}{q} \geq (.05)\delta$ (using $q \geq \frac{20d}{\epsilon^2\delta}$). Since $\alpha_C = E_s[\alpha_s]$, we get $\alpha_C \geq \tau(.05)\delta$. On the other hand, by Corollary 28, this probability is at most ϵ on the cube C. Putting them together, we get $\tau \leq \frac{20\epsilon}{\delta} \leq \gamma/2$ as desired. The claim follows. □

Claim 3: *For at least* $1-\sqrt{\gamma}$ *fraction of lines in cube* C*, at least* $1-\sqrt{\gamma}$ *fraction of the planes containing that line are well-behaved.*

Proof of Claim 3: Among all planes that contain any line l, let σ_l denote the fraction that are well-behaved. Then by symmetry we know that $E_l[\sigma_l]$ is exactly the fraction of well-behaved planes in cube C, which is at least $1-\gamma$ by Claim 2. Averaging implies that $\sigma_l \geq 1-\sqrt{\gamma}$ for at least $1-\sqrt{\gamma}$ fraction of l. □

Now call a line l *super* if it is nice and if at least $1-\sqrt{\gamma}$ fraction of the planes containing l are well-behaved. By Claims 1 and 3, at least $1-\gamma-\sqrt{\gamma}$ fraction of lines in cube c are super.

Claim 4: *If line* l *is super, then*

$$E[\textit{agreement between } \hat{f}_l \textit{ and } f \textit{ on cube } C] \geq (1-\sqrt{\gamma}) \cdot ((.45)\delta - \frac{1}{q}) \tag{22}$$

and for every line l' *that is non-coplanar with* l

$$d\textit{-success-rate of } \hat{f}_l \textit{ on } l' \geq (1-\sqrt{\gamma}-\frac{1}{q}). \tag{23}$$

Proof of Claim 4: Recall that the set of planes containing l is a partition of cube $C \setminus l$. Since l is nice, $P_d^f(l)$ is $P_i|_l$ for some $i \in [1, k_0]$ (by definition of "nice"ness). In any plane s containing

l, the bivariate polynomial used to define $\hat{f}_l$ in that plane must agree with $P_i|_l$ on l and must have agreement at least $(.45)\delta$ with f on s (by definition of $\hat{f}_l$). If s is well-behaved for a nice line l, then $P_i|_s$ is a candidate for being this polynomial (by definition of "well-behaved"-ness), and the only such candidate (by definition of "nice"-ness). Hence the agreement between $\hat{f}_l$ and f on this plane is at least $(.45)\delta$. Not counting points on l itself, this still gives at least $((.45)\delta - \frac{1}{q})q^2$ points from $C \setminus l$. Summing over all well-behaved planes containing l (and using the fact that l is super), we see that the agreement between $\hat{f}_l$ and f on the cube C is at least $(1-\sqrt{\gamma}) \cdot ((.45)\delta - \frac{1}{q})$. Now the claim in (22) follows.

Now we prove the claim in (23). Note that there are exactly $q+1$ planes containing l and at least $(1-\sqrt{\gamma})(q+1)$ are well-behaved. Since all but one of these intersects l' in exactly one point, we have that at least $\frac{(1-\sqrt{\gamma})(q+1)-1}{q} \geq 1 - \sqrt{\gamma} - \frac{1}{q}$

fraction of the planes that contain l and intersect l', are well-behaved. Thus consider picking a random point x on l' and consider the probability it passes the low-degree test when the d-oracle always answers according to P_i. With probability $(1-\sqrt{\gamma}-\frac{1}{q})$, the plane containing x and l is well-behaved and thus $\hat{f}_l(x) = P_i(x)$ and the low-degree test accepts. □

We are now ready to prove Claim 0 from which the lemma follows. We pick q so that all conditions on q are satisfied. Note that this only requires $q > r((d+1)/\eta\delta)^s$ for appropriate choices of r and s. Note that the expected d-success rate of $\hat{f}_l$ on l' (for random l, l') is at least the probability that l is super, times the expectation of this value over super lines which is lower bounded by Claim 4. Thus (19) follows from the inequalities

$$\begin{aligned} & (1-\sqrt{\gamma}-\gamma)(1-\sqrt{\gamma}-\frac{1}{q}) \\ \geq \;& 1 - 2\sqrt{\gamma} - \gamma - \frac{1}{q} \\ \geq \;& 1 - 4\sqrt{\gamma} \quad (\text{Using } q \geq \tfrac{1}{\sqrt{\gamma}}) \\ \geq \;& 1 - \frac{\eta}{2} \quad (\text{Using } \gamma \leq \tfrac{\eta^2}{64}) \end{aligned}$$

Similarly (20) follows from the inequalities

$$\begin{aligned} & (1-\sqrt{\gamma}-\gamma)(1-\sqrt{\gamma}) \cdot ((.45)\delta - \frac{1}{q}) \\ \geq \;& (1-3\sqrt{\gamma}) \cdot ((.45)\delta - \frac{\delta}{48}) \quad (\text{Using } q \geq \tfrac{48}{\delta}) \\ \geq \;& \delta(.40)(1-3\sqrt{\gamma}) \\ \geq \;& \frac{\delta}{3} \quad (\text{Using } \gamma \leq \tfrac{1}{18^2}) \end{aligned}$$

This concludes the proof of Claim 0. Hence the lemma is also proved. □

This concludes the analysis of the low-degree test in the general case.

4 Applications

We now describe two applications of the low-degree test.

4.1 Construction of constant prover 1-round proof systems

Our first application is to the task of constructing low-error constant prover one-round proof systems (MIPs) for languages in NP. Such objects are of interest due to their role in the construction of probabilistically checkable proofs and the derivation of hardness of approximation results. We formally define MIPs next.

A p-prover 1-round proof system for a language L consists of a verifier that checks membership proofs for L (a proof that a given input x is in L) in the following way. The proof consists of p oracles. (An oracle is a table that, for some $a, b > 0$, contains 2^b strings from $\{0,1\}^a$. When we supply this oracle a b-bit address, it returns the a-bit string stored at the corresponding location. We call a the *answer size* of the oracle.) The verifier is probabilistic. It uses its randomness to compute one address in each of the p oracles, reads the strings in those locations, and then computes an ACCEPT or REJECT decision. (The name "p-prover 1-round system" is a holdover from the past; we could also use "p-oracle 1-round systems.") If the string x is in L, then there must exist p oracles such that the verifier accepts on every random string. On the other hand, if the input is not in the language, then for any tuple of p oracles, the verifier accepts with probability at most e. We call e the "error" of the proof system.

To use verifier composition we also need to ensure that the verifier's ACCEPT/REJECT decision is computed in a very simple way, by evaluating a small circuit. At the start, the verifier uses its random string and the input to compute a circuit C and one location in each of the p oracles. After reading the oracles, the verifier outputs ACCEPT iff the concatenation of the strings it just read is a satisfying assignment to C. The size of C (= number of wires in it) is called the *circuit size* of the verifier.

Now we define MIP$[p, r, a, e]$, the class of languages that have such verifiers.

Definition 5 (MIP$[p, r, a, e]$ **)** *For a positive integer p, functions $r, a : \mathcal{Z}^+ \to \mathcal{Z}^+$, and $e : \mathcal{Z}^+ \to Q^+$ a language L is said to belong to MIP$[p, r, a, e]$ if there exists a probabilistic polynomial-time verifier V that on any input $x \in \{0,1\}^n$ uses $r(n)$ random bits, expects the membership proof to contain p oracles of answer size $a(n)$, and has the following behavior:*

1. *If $x \in L$, then there exist oracles $\pi_1, \ldots, \pi_p$ such that V always outputs ACCEPT (i.e., outputs ACCEPT with probability 1).*

2. *If $x \notin L$, then there for every set of oracles $\pi_1, \ldots, \pi_p$, verifier V outputs ACCEPT with probability at most $e(n)$.*

Furthermore, the circuit size of the verifier is polynomial in $a(n)$.

To construct very efficient $O(1)$-prover 1-round proof systems for SAT we use two standard techniques. First we plug our low degree test into a construction of [5] to get a proof system with 3 provers that uses $O(\log n)$ random bits but the oracles in the proof have answer size $2^{\log^\beta n}$ for some $\beta < 1$. We then extend this verifier into a verifier of concatenated and encoded proofs (see [6]), or an inner verifier in the language of [5] that takes any p-prover MIP and converts it into a $p + 3$ prover MIP with smaller answer sizes. Then we use "verifier composition," a technique from [6], to reduce the answer size to $O(\log n)$ (the number of oracles stays $O(1)$). Lemma 31 summarizes the consequence of this construction.

Lemma 31 *There exist $\epsilon > 0$ and $\alpha < \infty$ such that for every r, p, a, e the following holds: MIP$[p, r, a, e] \subset$ MIP$[p + 3, r + O(m \log |F|), O((poly \log a) d \log |F|), e^{\frac{1}{2p+2}}]$. where d, m are any positive integers and F is a any finite field satisfying the following conditions:*

- $|F| \geq poly(m, d, \log a, \frac{1}{e})$.
- $(d/m)^m \geq a^{O(1)}$.

The following theorem follows by a simple induction from this lemma.

Theorem 32 *For every $\beta < 1$, there exists a $p < \infty$ such that*

$$NP \subset MIP[p, O(\log n), O(\log n), 2^{-\Omega(\log^\beta n)}].$$

Proof: We start with the obvious containment NP $\subseteq$ MIP$[1, 0, \text{poly}(n), 0] \subseteq$ MIP$[1, 0, \text{poly}(n), 2^{-log^\beta n}]$. We then recurse $(\frac{1}{1-\beta})$-times using Lemma 31 with $|\mathrm{F}| = 2^{\log^\beta n}$, $m = \log^{1-\beta} n$ for all applications. The choice of d is set to satisfy the condition $(d/m)^m \geq a^{O(1)}$ and we pick $d = m2^{\log^{1-i(1-\beta)} n}$ in the ith application. This yields NP $\subset$ MIP$[1 + 3(\frac{1}{1-\beta} + 1), O(\log n), O(\log^\beta n \text{poly} \log \log n), 2^{-\Omega(\log^\beta n)}]$. □

4.2 Self-correction of programs

Consider a program $\mathcal{P}$ that is supposed to be computing an unknown polynomial g. Suppose $\mathcal{P}$ is correct on only some tiny δ fraction of the inputs. We can use the analysis of our low-degree test to design a testing procedure that estimates the largest δ for which the program's output agrees with the output of some polynomial, to within an additive error of $O(1/q^\epsilon)$ over a field of size q. (In order to implement such a procedure we need to simulate a d-oracle for $\mathcal{P}$. We can do this by *computing*, for any given line l, a degree d polynomial that has maximal agreement with $\mathcal{P}$ on this line.) In this section we describe a complementary result that takes such a (slightly correct) program and uses it to produce a randomized fully correct program computing g.

The task of self-correcting this program needs to be defined carefully. For starters, there can be more than one polynomial agreeing with the program $\mathcal{P}$ in δ fraction of the inputs. In fact, we can have $O(\frac{1}{\delta})$ such polynomials (Proposition 4). However, we can be expected to reconstruct $O(\frac{1}{\delta})$ (randomized) "programs", each of which computes a polynomial (and is correct on every input with high probability), such that every polynomial that has $\frac{1}{\delta}$ agreement with $\mathcal{P}$ is computed by one of the programs. This task was left as an open problem in Ar et al. [1], and no polynomial (in m, d and $\frac{1}{\delta}$) time algorithm was known for this problem. Goldreich et al. [19] solve this task when $\delta \geq 2\sqrt{d/q}$ in time exponential in d. We now describe our solution that works when $\delta \geq (md/q)^\epsilon$, for some positive ϵ, and is the first polynomial time-bounded solution for any $\delta < 1/2$.

Given a program $\mathcal{P}$, our algorithm works in two phases: First, a preprocessing phase, where we instantiate $k \leq O(\frac{1}{\delta})$ programs $\mathcal{P}_1, \ldots, \mathcal{P}_k$. In the second phase a program $\mathcal{P}_i$ takes an input $x \in \mathrm{F}^m$ and computes its output $\mathcal{P}_i(x)$. The guarantee is that at the end of the first phase, with high probability, we create k randomized programs, such that the output of each is (with high probability) a polynomial; furthermore, for every polynomial g which agrees with $\mathcal{P}$ on δ fraction of the input, one of the programs $\mathcal{P}_i$ computes g correctly with high probability on every input. The two phases are based on the analysis of the bootstrapping method described in Section 3.3 which is in turn based on the work of Arora [3]. Both phases use the algorithm of Sudan [33].

> **Preprocessing Phase:** Pick a random line l through the m-dimensional space F^m and reconstruct (using the algorithm of [33]) a list of all polynomials $p_1, \ldots, p_k$ agreeing with $\mathcal{P}$ on $\delta/4$ fraction of the inputs. For each polynomial p_i create a new program $\mathcal{P}_i$ which works as follows:
>
> **Query phase:** Let $x \in \mathrm{F}^m$ be the query. Pick two random points $r_1, r_2 \in \mathrm{F}^m$. Let D be the 4-dimensional (affine) space containing points x, r_1, r_2 and line l from the the preprocessing phase. Reconstruct a list of all 4-variate polynomials $g_1, \ldots, g_k$ that agree with $\mathcal{P}$ on at least $\delta/2$ fraction of the points in D (again using the algorithm of [33]). If there exists a unique polynomial g_j such that $g_j|_l$ equals p_i, then output the value of g_j at x, else output FAIL.

Theorem 33 *There exist constants c_0, c_1 and $\gamma < \frac{1}{4}$ such that the algorithm described above runs in time polynomial in d, $\frac{1}{\delta}$, $\log q$ and m with oracle access to $\mathcal{P}$ and, with probability at least $1 - \gamma$, produces a set of randomized oracle programs $\mathcal{P}_1, \ldots, \mathcal{P}_k$ such that for for every degree d polynomial $g : F^m \to F$ which has δ agreement with $\mathcal{P}$, there exists a randomized program $\mathcal{P}_i$ that computes g correctly on every input with probability at least $1 - \gamma$, provided $|F| > c_0 \left(\frac{d+1}{\delta}\right)^{c_1}$.*

Remark: Note the two occurrences of $1 - \gamma$ in the statement above. The first refers to the success of the preprocessing phase, and the second to the success of the algorithms in the query phase, given a successful preprocessing phase. As is usual, the error probability γ can be reduced by standard amplification techniques. In particular, we can run the preprocessing stage $\log 1/\gamma$ times and take the union of all the programs produced in the preprocessing stages.

Proof: We prove the lemma with a sequence of claims. The first sequence of claims proves the correctness of the preprocessing phase. As usual we will need a lower bound on q. Here we set $\gamma = \frac{1}{4}$, $\epsilon = min\{\frac{\delta}{8}, \frac{\gamma}{6}, \frac{\gamma\delta}{48}\}$. Let c_0', c_1' denote the constants needed to apply the remark following the proof of Theorem 17 (which shows that most of the success probability of the low-degree test comes from lines where the d-oracle responds with the restriction of some polynomial that has significant agreement with f). We will assume

$$q > \max\left\{\frac{4d}{\epsilon^2}, \frac{6d}{\epsilon^2\gamma}, \frac{24}{\epsilon\delta\gamma}, c_0'((d+1)/\epsilon)^{c_1}, \frac{27}{2\delta\gamma}, \frac{16d}{\epsilon\gamma}, \frac{288}{\gamma\delta}, \frac{16}{\gamma}\right\}.$$

Let $P_1, \ldots, P_n$ be all the degree d polynomials that agree with $\mathcal{P}$ in δ fraction of the inputs, and let $P_1, \ldots, P_{n'}$ be all the degree d polynomials that agree with $\mathcal{P}$ in ϵ fraction of the inputs. (Notice, since $q > \frac{4d}{\epsilon^2}$, we have $n \leq n' \leq 2/\epsilon$).)

Claim 6 *With probability at least $1 - \frac{\gamma}{3}$, over the choice of l, no two polynomials P_i, P_j, $i, j \in [n']$, agree on the line l.*

Proof: We prove something even stronger, i.e. that the polynomials all disagree at some randomly chosen point on r. For a fixed $i \neq j$, the probability that P_i and P_j agree at a randomly chosen point r in F^m is at most d/q. Thus the probability that there exists a pair (i, j) such that P_i and P_j agree at r is at most $\binom{n'}{2} d/q$ which is at most $2d/\epsilon^2 q) \leq \frac{\gamma}{3}$ (using $q \geq \frac{6d}{\epsilon^2\gamma}$). □

Claim 7 *With probability at least $1 - \frac{\gamma}{3}$, for every $i \in [n]$, $\mathcal{P}$ agrees with P_i on at least $\delta/4$ fraction of the points on l.*

Proof: Recall that to pick a random line l we need to pick two random points in F^m. Thus we conclude that the set of points in l is a pairwise independent sample of random points from F^m. Hence the fraction of points on the line where $\mathcal{P}$ and P_i agree is (by Chebyshev's inequality) very likely to be close to the fraction of points where they agree in F^m. Specifically, the probability that this fraction is less than $\delta/2$ is at most $\frac{4}{\delta q}$. The probability that there exists an i such that this happens is at most $\frac{4n}{\delta q} \leq \frac{8}{\epsilon\delta q} \leq \frac{\gamma}{3}$ (where the last inequality uses $q \geq \frac{24}{\epsilon\delta\gamma}$). □

Claim 8 *With probability at least $1 - \frac{\gamma}{3}$, every polynomial p_i found in the preprocessing phase is the restriction of some polynomial P_j for $j \in [n']$.*

Proof: From the remark following the proof of Theorem 17, we have that the probability that the low-degree test accepts on a line where the d-oracle responds with a polynomial that is not the restriction of $P_1, \ldots, P_{n'}$ is at most $2\epsilon \leq \frac{\gamma}{3}$. □

We say that the preprocessing phase is *successful* if all of the good events listed in Claims 6-8 occur. Notice that this phase is successful with probability at least $1 - \gamma$. We now carry out the rest of the analysis assuming the first phase is successful. We will show that for any polynomial P_i, $i \in [n]$, there exists some program $\mathcal{P}_j$ which computes it. Specifically if we let p_i be the restriction of P_i to l, the program computing P_i will be $\mathcal{P}_i$, the program associated with p_i.

The probability in all the following claims are made for a random choice of r_1 and r_2. Recall that $D, g_1, \ldots, g_{k'}$ were defined in the query phase.

Claim 9 *With probability at least $1 - \frac{\gamma}{2}$, the polynomials $P_1, \ldots, P_n$ have agreement at least $\delta/2$ with $\mathcal{P}$ when restricted to D.*

Proof: D consists of all affine combinations of x, r_1, r_2, l. Since r_1, r_2 are chosen randomly, the points of D (except for those lying in the plane containing x and l) form a pairwise independent sample from F^m. Specifically we have $q^4 - q^2 \geq \frac{3}{4}q^4$ pairwise independent points and we need to get an agreement of at least $\delta q^4/2$ points. Thus it suffices to get an agreement ratio of $\frac{2\delta}{3}$ on these points. By Chebyshev's inequality, it follows that the probability that this does not happen is at most $\frac{27}{4\delta q^4} \leq \frac{\gamma}{2}$ (using $q \geq \frac{27}{2\delta\gamma}$). □

Claim 10 *With probability at least $1 - \frac{\gamma}{2}$, every g_i is the restriction of some P_j, $j \in [n']$.*

Proof: The proof involves noticing that instead of picking r_1, r_2 uniformly at random from F^m, an almost equivalent way is the following: first we pick D randomly and then we pick two random points r_1, r_2 from D. If the line l, and the points x, r_1 and r_2 affinely span all of D, then the points r_1 and r_2 are truly random in F^m. The probability that these points don't span D is at most $\frac{2}{q}$. Let l' be the line connecting r_1, r_2.

Let p be the probability of our "bad event," namely, when we pick D, there exists a 4-variate polynomial g which is not the restriction of some P_j, but has agreement $\delta/2$ with f on D. To upperbound p, we consider the following event: there is a univariate polynomial h that has agreement at least $\delta/3$ with f on the line l', but it is not the restriction of any of $P_1, \ldots, P_{n'}$. Let τ denote the probability of this event. Note that on a line where such a polynomial exists, a random point satisfies the low-degree test with probability at least $\delta/3$. Thus with probability at least $\tau\delta/3$, the low-degree test accepts on lines where the d-oracle does not agree with any of the polynomials $P_1, \ldots, P_{n'}$. As noted earlier this probability is at most 2ϵ, giving $\tau \leq \frac{6\epsilon}{\delta}$. Next we give a lower bound on τ in terms of p.

Now consider the other way of picking l', by picking D first. When D is picked, the probability is at least p that our bad event happens. Conditioned upon this event, consider the following possibilities:

1. g equals one of $P_1, \ldots, P_{n'}$ on the line l': The probability of this event is at most $\frac{n'd}{q} \leq \frac{2d}{\epsilon q}$.

2. g has agreement less than $\delta/3$ on l': By Chebyshev's inequality this probability is at most $\frac{36}{\delta q}$.

3. The line l' is not a random line of F^m: The probability of this event is at most $\frac{2}{q}$.

Barring the three possibilities above, we find that we find a polynomial that has agreement at least $\delta/3$ with f on a random line of F^m which is not the restriction of one of the P_i's. Thus we find $\tau \geq p - \frac{2d}{\epsilon q} - \frac{36}{\delta q} - \frac{2}{q}$.

Combining the two we get

$$p \leq \frac{6\epsilon}{\delta} + \frac{2d}{\epsilon q} + \frac{36}{\delta q} + \frac{2q}{\leq}\gamma/2.$$

(The last inequality follows from $\epsilon \leq \frac{\gamma\delta}{48}$, and $q \leq \max\{\frac{16d}{\epsilon\gamma}, \frac{288}{\gamma\delta}, \frac{16}{\gamma}\}$.
□

We are now almost done. By Claim 9, for any P_i, there exists some 4-variate polynomial, say g_j, that agrees with P_i on D. Furthermore, all the g's are the restrictions of some $P_{i'}$, $i' \in [n']$ and thus the g's restricted to l are distinct. Thus the program $\mathcal{P}_i$ will output according to g_j on x which is equal to P_i on input x.

This concludes the proof of the correctness of the self-correction algorithm. □

5 Conclusions

We do not know how to reduce the number of provers in our constructions of constant prover protocols to 2. So long as we use the verifier composition idea of [6], 3 provers appears to be the best possible. Reducing the number of provers to 2 would imply the NP-hardness of approximation problems dealt with in [4].

Thanks

Sanjeev Arora thanks Laci Babai and Kati Friedl for introducing him to "symmetry-based" arguments for the low degree test in summer 1993. We thank Dick Lipton for saving us from fruitless labor on an incorrect conjecture on irreducibility (he provided a counterexample). We thank Erich Kaltofen for providing pointers to his work. We thank the anonymous referees for pointing out numerous errors in a previous version of this paper.

References

[1] S. Ar, R. Lipton, R. Rubinfeld and M. Sudan. Reconstructing algebraic functions from mixed data. *SIAM Journal on Computing*, 28(2): 487–510, April 1999.

[2] S. Arora *Unpublished, 1993.*

[3] S. Arora. *Probabilistic Checking of Proofs and Hardness of Approximation Problems.* PhD thesis, U.C. Berkeley, 1994. Available from `http://www.cs.princeton.edu/~ arora` .

[4] S. Arora, L. Babai, J. Stern and Z. Sweedyk. The hardness of approximating problems defined by linear constraints. *Journal of Computer and System Sciences*, 54(2):317–331, April 1997.

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

[6] S. Arora and S. Safra. Probabilistic checking of proofs: a new characterization of NP. *Journal of the ACM.* 45(3):501-555, January 1998. Preliminary version in *Proceedings of the* 33rd *Symposium on Foundations of Computer Science*, IEEE, 1992.

[7] L. Babai, L. Fortnow, and C. Lund. Non-deterministic exponential time has two-prover interactive protocols. *Computational Complexity*, 1:3–40, 1991.

[8] L. Babai, L. Fortnow, L. Levin, and M. Szegedy. Checking computations in polylogarithmic time. *Proceedings of the* 23rd *Annual Symposium on Theory of Computing*, ACM, 1991.

[9] M. Bellare, O. Goldreich and M. Sudan. Free bits, PCPs and non-approximability — towards tight results. *SIAM Journal on Computing*, 27(3): 804–915, June 1998.

[10] M. Bellare, S. Goldwasser, C. Lund, and A. Russell. Efficient probabilistically checkable proofs. *Proceedings of the* 25th *Annual Symposium on Theory of Computing*, ACM, 1993. (See also Errata sheet in *Proceedings of the* 26th *Annual Symposium on Theory of Computing*, ACM, 1994).

[11] M. Blum, M. Luby, and R. Rubinfeld. Self-testing/correcting with applications to numerical problems. *Journal of Computer and System Sciences*, 47:549–595, 1993.

[12] U. Feige. A threshold of ln n for Set Cover. *Journal of the ACM*, 45(4):634–652, July 1998.

[13] U. Feige, S. Goldwasser, L. Lóvasz, S. Safra and M. Szegedy. Interactive proofs and the hardness of approximating cliques. *Journal of the ACM*, 43(2):268-292, 1996.

[14] U. Feige and J. Kilian. Two prover protocols – Low error at affordable rates (preliminary version). *Proceedings of the* 26th *Annual Symposium on Theory of Computing*, ACM, 1994.

[15] U. Feige and J. Kilian. Impossibility results for recycling random bits in two-prover proof systems. *Proceedings of the* 27th *Annual Symposium on Theory of Computing*, ACM, 1995.

[16] U. Feige and L. Lóvasz. Two-prover one-round proof systems: Their power and their problems. *Proceedings of the* 24th *Annual Symposium on Theory of Computing*, ACM, 1992.

[17] K. Friedl and M. Sudan. Some improvements to low-degree tests. *Proceedings of the Third Israel Symposium on Theory and Computing Systems*, IEEE, 1995.

[18] P. Gemmell, R. Lipton, R. Rubinfeld, M. Sudan and A. Wigderson. Self-testing/correcting for polynomials and for approximate functions. *Proceedings of the* 23rd *Annual Symposium on Theory of Computing*, ACM, 1991.

[19] O. Goldreich, R. Rubinfeld and M. Sudan. Learning polynomials with queries: The highly noisy case. *SIAM Journal on Discrete Mathematics*, 13(4): 535–570, November 2000.

[20] R. Impagliazzo and A. Wigderson. P=BPP if E requires exponential size circuits: Derandomizing the XOR lemma. *Proceedings of the* 29th *Annual Symposium on Theory of Computing*, ACM, 1997.

[21] E. Kaltofen. Polynomial-time reductions from multivariate to bi- and univariate integral polynomial factorization. *SIAM Journal on Computing*, 14(2):469–489, 1985.

[22] E. Kaltofen. Effective Hilbert irreducibility. *Information and Control*, 66:123–137, 1985.

[23] E. Kaltofen. Effective Noether irreducibility forms and applications. *Journal of Computer and System Sciences*, 50(2):274-295, 1995.

[24] D. Lapidot and A. Shamir. Fully Parallelized Multi-prover protocols for NEXP-time. *Journal of Computer and System Sciences*, 54(2):215-220, April 1997.

[25] C. Lund, L. Fortnow, H. Karloff, and N. Nisan. Algebraic methods for interactive proof systems. *Journal of the ACM*, 39(4):859-868, 1992.

[26] C. Lund and M. Yannakakis. On the hardness of approximating minimization problems. *Journal of the ACM*, 41(5): 960-981(1994).

[27] A. Polishchuk and D. Spielman. Nearly Linear Sized Holographic Proofs. *Proceedings of the* 26th *Annual Symposium on Theory of Computing*, ACM, 1994.

[28] R. Raz. A parallel repetition theorem. *Proceedings of the* 27th *Annual Symposium on Theory of Computing*, ACM, 1995. *SIAM Journal on Computing*, 27(3):763-803, June 1998.

[29] R. Raz and S. Safra. A Sub-Constant Error-Probability Low-Degree Test, and a Sub-Constant Error-Probability PCP Characterization of NP. *Proceedings of the* 29th *Annual Symposium on Theory of Computing*, ACM, 1997. (Communicated to the authors in March 1996.)

[30] R. Rubinfeld and M. Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM Journal on Computing* 25:2, pp. 252–271, 1996.

[31] J. T. Schwartz. Probabilistic algorithms for verification of polynomial identities. *Journal of the ACM*, 27:701–717, 1980.

[32] A. Shamir. IP = PSPACE. *Journal of the ACM*, 39(4):869-877, 1992.

[33] M. Sudan. Decoding of Reed Solomon codes beyond the error-correction bound. *Journal of Complexity*, 13(1):180–193, March 1997.

[34] M. Sudan, L. Trevisan, and S. Vadhan. Pseudorandom generators without the XOR lemma. *Journal of Computer and System Sciences*, 62(2): 236–266, March 2001.

[35] G. Tardos. *Personal Communication*, 1993.

[36] G. Tardos. Multi-prover encoding schemes and three-prover proof systems. *Journal of Computer and System Sciences*, 53(2):251-260, October 1996.

[37] G. Tardos. *Personal Communication*, 1996.

A A version of Hilbert irreducibility

In order to analyze the low-degree test in the cases $m = 2, 3$ we needed something like the following: if a polynomial is irreducible, then its restriction on most lines does not have a "linear" factor. Now we state and prove this fact. It is a simpler version of Kaltofen's "Effective Hilbert Irreducibility" [22], in that it focusses only factors that are linear in one of the variables. However, the theorem in Kaltofen [23] considers a different substitution $y_i = a_i t + b_i z + c_i$ instead of $y_i = a_i t + b_i$, making the theorem statement incomparable. However, the proof follows from that of Kaltofen [23], and is included here mainly for completeness.

A polynomial (in this section, "polynomial" means a formal polynomial) $Q \in \mathrm{F}[z, y_1, \ldots, y_m]$ is said to be *absolutely irreducible* if it does not factor over $\overline{\mathrm{F}}$, the algebraic closure of F.

Theorem 34 *Let $Q \in F[z, y_1, y_2, \ldots, y_m]$ be a degree l polynomial that is absolutely irreducible and not of the form $c \cdot (z - f(y_1, y_2, \ldots, y_m))$ for some $c \in F$, $f \in F[(y_1, \ldots, y_m]$. Then the fraction of $(a_1, a_2, \ldots, a_m, b_1, b_2, \ldots, b_m) \in F^{2m}$ for which*

$$Q(z,\, a_1 t + b_1,\, \ldots, a_m t + b_m) \in F[z,t] \text{ has a factor of the form } z - p(t) \text{ in } \overline{F}[z,t]$$

is at most $3l^3/q$.

First some notation: We will use $\mathrm{F}(y_1, y_2, \ldots, y_k)$ and $\overline{\mathrm{F}}(y_1, y_2, \ldots, y_k)$ to denote the quotient fields of $\mathrm{F}[y_1, y_2, \ldots, y_k]$ and $\overline{\mathrm{F}}[y_1, y_2, \ldots, y_k]$ respectively. The congruence (i.e., $\equiv$) modulo $[y_1, \ldots, y_m]^{l+1}$ means that the polynomials on the two sides of the $\equiv$ are identical once we discard all terms whose total degree in y_i's is $l + 1$ or higher. We also use the shorthand $\hat{y}$ to denote the tuple $\langle y_1, \ldots, y_m \rangle$.

The main idea behind the proof (as adapted from [22]) is as follows: We apply a multivariate factorization algorithm to the polynomial Q, and simultaneously apply the algorithm to one of the polynomials $Q_{\hat{a},\hat{b}}(z, t) = Q(z, a_1 t + b_1, \ldots, a_m t + b_m)$. The algorithm fails to produce any non-trivial factors in the first case. We show that the algorithm runs in a close parallel in the latter case and hence also fails to produce any (linear) factors in this case, for most choices of $\hat{a}, \hat{b}$.

The central idea behind the factorization algorithm is that of Hensel lifting - an idea that lifts a co-prime factorization of $Q(z, 0, \ldots, 0)$ into a factorization of Q modulo $[\hat{y}]^l$ for $l = 1, 2, \ldots$. Since Q may not have a co-prime factorization at $\hat{y} = (0, \ldots, 0)$, we first shift the polynomial to a random choice of origin at $b_1, \ldots, b_m$. The first two lemmas below describe why such a shift works. The following lemmas establish the correctness and uniqueness of the Hensel-lifting procedure. This allows us to conclude with a proof of Theorem 34 at the end of this section.

The proofs use some basic algebra including Gauss's Lemma and how to solve linear equations, which can be found in any basic algebra text.

We say that a univariate polynomial $p \in \mathrm{F}[x]$ is *square-free* if it doesn't have a repeated irreducible factor in F. For such a polynomial if $f \cdot g = p$ is any factorization of p in $\overline{\mathrm{F}}[x]$, then f, g have no common factor. The following standard lemma about *discriminants* gives a necessary and sufficient condition for square-freeness.

Lemma 35 *A degree k polynomial $p = \sum_{i=0}^{d} p_i x^i \in F[x]$ is square-free iff the determinant of the following $(2d-1) \times (2d-1)$ matrix (the so-called* discriminant*) is nonzero, where $g_i = (i+1)p_{i+1}$*

for $i \leq d-1$ and $g_d = 0$.

$$\begin{pmatrix}
p_0 & 0 & 0 & \cdots & 0 & g_0 & 0 & 0 & \cdots & 0 \\
p_1 & p_0 & 0 & \cdots & 0 & g_1 & g_0 & 0 & \cdots & 0 \\
p_2 & p_1 & p_0 & \cdots & 0 & g_2 & g_1 & g_0 & \cdots & 0 \\
p_3 & p_2 & p_1 & \cdots & 0 & g_3 & g_2 & g_1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
p_{d-1} & p_{d-2} & p_{d-3} & \cdots & p_1 & g_{d-1} & g_{d-2} & g_{d-3} & \cdots & g_0 \\
p_d & p_{d-1} & p_{d-2} & \cdots & p_2 & 0 & g_{d-1} & g_{d-2} & \cdots & g_1 \\
0 & p_d & p_{d-1} & \cdots & p_3 & 0 & 0 & g_{d-1} & \cdots & g_2 \\
0 & 0 & p_d & \cdots & p_4 & 0 & 0 & 0 & \cdots & g_3 \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
0 & 0 & 0 & \cdots & p_d & 0 & 0 & 0 & \cdots & g_{d-1}
\end{pmatrix}$$

Proof: A polynomial p has a repeated irreducible factor iff p and p' (the formal derivative of p) share a factor, which happens iff there are nonzero polynomials A and B of degree at most $d-2$ and $d-1$ respectively such that $A \cdot p + B \cdot p' = 0$. (The "only if" direction is trivial: if g is a repeated factor of p, then $A = p'/g$ and $B = -p/g$ satisfy this condition. The "if" direction is also standard.) We can try to find such polynomials by representing their coefficients as unknowns and writing a (homogeneous) linear system of equations. The determinant we have written is the determinant of this system; it is zero iff a nontrivial solution (i.e., A and B) exists. $\square$

Corollary 36 *Let $Q \in F[z, y_1, \ldots, y_m]$ have degree l and be absolutely irreducible. Then there is a nonzero polynomial $\Phi_Q \in F[y_1, \ldots, y_m]$ of degree $2l^2$ such that for all $b_1, \ldots, b_m \in F$,*

$$\Phi_Q(b_1, \ldots, b_m) \neq 0 \ \Rightarrow\ Q(z, b_1, \ldots, b_m) \text{ is square-free in } F[z].$$

Proof: Write Q as $\sum_{i=0}^{l} z^i p_i(y_1, \ldots, y_m)$, that is, as a polynomial in $\mathrm{F}[y_1, \ldots, y_m][z]$. Let Φ_Q be the discriminant of this polynomial. Since the discriminant is a polynomial of degree $2l-1$ in the coefficients, and each coefficient in this case is itself a degree l polynomial in $\mathrm{F}[y_1, \ldots, y_m]$, we conclude that $\Phi_Q \in \mathrm{F}[y_1, \ldots, y_m]$ has degree at most $2l^2$. Furthermore, since Q is irreducible in $\overline{\mathrm{F}}[y_1, \ldots, y_m][z]$, Gauss' Lemma (alternatively, unique factorization in polynomial rings) implies that Q is irreducible in $\overline{\mathrm{F}}(y_1, \ldots, y_m)[z]$ as well, so the discriminant Φ_Q is non-zero. $\square$

We now move onto the factorization algorithm. For the moment we describe the algorithm as if the factorization is being lifted from $\hat{y} = (0, \ldots, 0)$.

Lemma 37 *Let $Q \in F[z, \hat{y}]$ and $g \in F[\hat{y}]$ be any polynomials. Then for any $l \geq 0$,*

$$Q(g(\hat{y}), \hat{y}) \equiv 0 \ (\bmod\ [\hat{y}]^{l+1}) \tag{24}$$

iff there is a polynomial $h \in F[z, \hat{y}]$ such that

$$Q(z, \hat{y}) \equiv (z - g(\hat{y}))h(z, \hat{y}) \ (\bmod\ [\hat{y}]^{l+1}). \tag{25}$$

Furthermore, if such an h exists, then it is unique modulo $[\hat{y}]^{l+1}$.

Proof: Let $Q(z, \hat{y}) = z \cdot h(z, \hat{y}) + f(\hat{y})$, where h and f are polynomials. Let $r(\hat{y}) = Q(g(\hat{y}), \hat{y})$. Then it is easily verified that

$$Q(z, \hat{y}) = (z - g(\hat{y}))h(z, \hat{y}) + r(\hat{y}).$$

Using the above form for Q, we get:

$$Q(g(\hat{y}), \hat{y}) \equiv 0 \ (\bmod\ [\hat{y}]^{l+1}) \ \Leftrightarrow\ r(\hat{y}) \equiv 0 \ (\bmod\ [\hat{y}]^{l+1}) \ \Leftrightarrow\ Q(z, \hat{y}) \equiv (z - g(\hat{y}))h(z, \hat{y}) \ (\bmod\ [\hat{y}]^{l+1}).$$

To establish uniqueness of h, let us write $h(z, \hat{y})$ as $h_0(\hat{y}) + zh_1(\hat{y}) + \cdots + z^k h_k(\hat{y})$. Suppose h' satisfies

$$Q(z, \hat{y}) \equiv (z - g(\hat{y}))h'(z, \hat{y}) \ (\bmod\ [\hat{y}]^{l+1}).$$

Further let $h'(z,\hat{y}) = h'_0(\hat{y}) + \cdots + z^{k'} h'_{k'}(\hat{y})$. Assume w.l.o.g. that $k < k'$. Then by comparing coefficients of $z^{k'+1}, z^{k'}, \ldots, z^1$ in the equation

$$(z - g(\hat{y}))h(z,\hat{y}) \equiv (z - g(\hat{y}))h'(z,\hat{y}) \ (\bmod\ [\hat{y}]^{l+1}),$$

we get $h_i(\hat{y}) \equiv h'_i(\hat{y})$ (mod $[\hat{y}]^{l+1}$), thus establishing uniqueness of h. □

Hensel lifting is used in factorization algorithms to "lift" univariate factors to multivariate factors. The following lemma is an instance.

Lemma 38 (A Version of Hensel Lifting) *Let $Q \in F[z,\hat{y}]$ be any polynomial and $\alpha \in \overline{F}$ be a root of multiplicity 1 of the polynomial $p(z) \overset{\text{def}}{=} Q(z,0,0,\ldots,0)$. Then for each $l \geq 1$ there exists a unique polynomial $q_l \in \overline{F}[\hat{y}]$ of total degree l such that*

$$Q(q_l(\hat{y}),\hat{y}) \equiv 0 \ (\bmod\ [\hat{y}]^{l+1}) \ \text{ and } q_l(0,0,\ldots,0) = \alpha.$$

Proof: By Lemma 37, it suffices to prove that for each $l \geq 0$, there exists a unique pair of polynomials $h_l \in \overline{\mathrm{F}}[z,\hat{y}], q_l \in \overline{\mathrm{F}}[\hat{y}]$ whose total degree in $\hat{y}$ is at most l and which satisfy:

$$Q(z,\hat{y}) \equiv (z - q_l(\hat{y})) \cdot h_l(z,\hat{y}) \ (\bmod\ [\hat{y}]^{l+1}) \ \text{ and } q_l(0,0,\ldots,0) = \alpha. \tag{26}$$

We use induction on l. The base case $l = 0$ is trivial, since $q_0 = \alpha$ and $h_0(z) = Q(z,0,\ldots,0)/(z-\alpha)$ are the only such polynomials.

Assume the statement is true up to $l \leq k$. The uniqueness property implies that $q_0, q_1, \ldots, q_k$, $h_0, \ldots, h_k$ satisfy for all $1 \leq i \leq k, 0 \leq j \leq i$:

$$q_i(\hat{y}) \equiv q_{i-j}(\hat{y}) \ (\bmod\ [\hat{y}]^{i-j+1})$$

and

$$h_i(z,\hat{y}) \equiv h_{i-j}(z,\hat{y}) \ (\bmod\ [\hat{y}]^{i-j+1}).$$

This means, for example, that each q_i is expressible as

$$q_i(\hat{y}) = q_j(\hat{y}) + \ \ \text{(terms whose degree in } \hat{y} \text{ is between } j+1 \text{ and } i).$$

A similar fact holds for the h_i's. Thus if any polynomials q_{k+1}, h_{k+1} satisfy condition (26) for $l = k+1$, then they must necessarily be of the form

$$q_{k+1}(\hat{y}) = q_k(\hat{y}) + \sum_{d_1,\ldots,d_m : \sum_i d_i = k+1} c_{d_1,\ldots,d_m} \prod_i y_i^{d_i}, \tag{27}$$

$$h_{k+1}(z,\hat{y}) = h_k(z,\hat{y}) + \sum_{d_1,\ldots,d_m : \sum_i d_i = k+1} e_{d_1,\ldots,d_m}(z) \prod_i y_i^{d_i}, \tag{28}$$

where each $c_{d_1,\ldots,d_m} \in \overline{\mathrm{F}}$ and each $e_{d_1,\ldots,d_m} \in \overline{\mathrm{F}}[z]$.

We claim that the following values are the only ones that allow q_{k+1}, h_{k+1} to have the desired properties:

$$c_{d_1,\ldots,d_m} = r_{d_1,\ldots,d_m}(\alpha)/h_0(\alpha) \tag{29}$$

and

$$e_{d_1,\ldots,d_m}(z) = \frac{1}{z-\alpha}(r_{d_1,\ldots,d_m}(z) - h_0(z)c_{d_1,\ldots,d_m}). \tag{30}$$

To see that these expressions make sense, note first that $h_0(\alpha) \neq 0$ since α is known to be a root of multiplicity 1 of $Q(z) = (z-\alpha)h_0(z)$. Furthermore, though $e_{d_1,\ldots,d_m}(z)$ is expressed as a rational function here, it is actually a polynomial because $r_{d_1,\ldots,d_m}(\alpha) - h_0(z)c_{d_1,\ldots,d_m}$ has a root at $z = \alpha$.

To see that the claim is true, let $R \in \overline{\mathrm{F}}[z, y_1, \ldots, y_m]$ be defined as

$$R(z,\hat{y}) = Q(z,\hat{y}) - (z - q_k(\hat{y}))h_k(z,\hat{y}) \ (\bmod\ [\hat{y}]^{k+2}). \tag{31}$$

Note that each term in R has degree $k+1$ in $\hat{y}$. Express it as

$$R(z,\hat{y}) = \sum_{d_1,\ldots,d_m : \sum_i d_i = k+1} r_{d_1,\ldots,d_m}(z) \prod_i y_i^{d_i}.$$

Since we desire q_{k+1}, h_{k+1} to satisfy

$$Q(z,\hat{y}) \equiv (z - q_{k+1}(\hat{y}))h_{k+1}(z,\hat{y}) \ \ (\text{ mod } [\hat{y}]^{k+2}),$$

we replace this in (31) to get

$$R(z,\hat{y}) = (z - q_{k+1}(\hat{y}))h_{k+1}(z,\hat{y}) - (z - q_k(\hat{y}))h_k(z,\hat{y}) \ (\text{ mod } [\hat{y}]^{k+2}).$$

Now replace expressions from (27) and (28) in (31), drop terms of total degree $> k+1$ in $\hat{y}$, and equate coefficients of monomials of $\hat{y}$ on both sides. We get for every tuple of degrees $(d_1,\ldots,d_m)$ satisfying $\sum_i d_i = k+1$:

$$(z-q_0)e_{d_1,\ldots,d_m}(z) + h_0(z)c_{d_1,\ldots,d_m} = r_{d_1,\ldots,d_m}(z). \tag{32}$$

Substituting $z = \alpha$ gives (29) and then (30) also follows. Clearly, these are the only solutions to (32). Thus we have proved both the existence and uniqueness of q_{k+1}, h_{k+1}, thus completing the induction.

□

Lemma 39 *Let $Q \in F[z, y_1, y_2, \ldots, y_m]$ be a degree l polynomial that is absolutely irreducible and not of the form $c \cdot (z - g(y_1,\ldots,y_m))$ for $c \in F$. Suppose $b_1,\ldots,b_m \in F$ are such that $Q(z, b_1, b_2, \ldots, b_m)$ is square-free. Then there exists a nonzero polynomial $\Psi_Q \in \overline{F}[v_1,\ldots,v_m]$ of degree l^3 such that for all $a_1,\ldots,a_m \in F$,*

$$\Psi_Q(a_1,a_2,\ldots,a_m) \neq 0 \implies f(z, a_1t+b_1,\ldots,a_mt+b_m) \textit{ has no factor like } z - p(t) \textit{ in } \overline{F}[z,t].$$

Proof: By the hypothesis, $Q(z,b_1,\ldots,b_m)$ is square-free. Define $T \in \mathrm{F}[z,y_1,\ldots,y_m]$ as

$$T(z,y_1,\ldots,y_m) = Q(z, y_1+b_1,\ldots,y_m+b_m).$$

Clearly, T is absolutely irreducible and $T(z,0,\ldots,0)$ is square-free. For each $a_1,\ldots,a_m \in \mathrm{F}$, let $T_{a_1,\ldots,a_m} \in \overline{\mathrm{F}}[z,t]$ be defined as

$$T_{a_1,\ldots,a_m}(z,t) \ = \ T(z, a_1t, a_2t, \ldots, a_mt). \tag{33}$$

We wish to give a "nice" description (namely, as roots of a low-degree polynomial Ψ_Q) of those tuples $(a_1,\ldots,a_m)$ for which

$$T_{a_1,\ldots,a_m} \text{ has a factor of the form } z - p(t), \text{ where } p \in \overline{\mathrm{F}}[t]. \tag{34}$$

Let $\alpha_1,\ldots,\alpha_k$ be all the roots of $T(z,0,\ldots,0)$. Thus $k \leq l$ and the α_i's are distinct. By Lemma 38, for each $i = 1,\ldots,k$, there is a unique degree l polynomial $g_i \in \overline{\mathrm{F}}[y_1,\ldots,y_m]$ such that

$$T(g_i(y_1,\ldots,y_m), y_1,\ldots,y_m) \equiv 0 \ (\text{ mod } [y_1,\ldots,y_m]^{l+1}) \qquad \text{and } g_i(0,\ldots,0) = \alpha_i. \tag{35}$$

Note that $g_i \neq g_j$ for $i \neq j$, since g_i and g_j differ at $(0,\ldots,0)$. Further, for each i

$$T(g_i(y_1,\ldots,y_m), y_1,\ldots,y_m) \neq 0, \tag{36}$$

since otherwise $z - g_i(y_1,\ldots,y_m)$ would be a factor of T and T is known to be absolutely irreducible.

Now let us identify tuples $(a_1,\ldots,a_m)$ for which $T_{a_1,\ldots,a_m}$ has a linear factor. For each $i = 1,\ldots,k$, think of the polynomial $g_i(a_1t,\ldots,a_mt)$ as a univariate polynomial in t. By examining (35) and the definition of $T_{a_1,\ldots,a_m}$, we see that for each $a_1,\ldots,a_m \in \mathrm{F}$,

$$T_{a_1,\ldots,a_m}(g_i(a_1t,\ldots,a_mt),t) \equiv 0 \ (\text{ mod } [t]^{l+1}) \ \text{ and } g_i(a_1t,\ldots,a_mt) \text{ is } \alpha_i \text{ at } t = 0.$$

The degree of $g_i(a_1t, \ldots, a_mt) \in \overline{\mathrm{F}}[t]$ is at most l. So we conclude from the uniqueness condition in the conclusion of Lemma 38 that $T_{a_1,\ldots,a_m}$ has a factor of the form $z - p(t)$ for $p \in \overline{\mathrm{F}}[t]$ iff that factor is $z - g_i(a_1t, \ldots, a_mt)$ for some $i \in [1..k]$. In other words, iff $T_{a_1,\ldots,a_m}(g_i(a_1t, \ldots, a_mt), t)$ is the zero polynomial. Now we show that the set of $(a_1, \ldots, a_m)$ for which $T_{a_1,\ldots,a_m}(g_i(a_1t, \ldots, a_mt), t)$ is the zero polynomial have a nice description as the roots of some polynomial Ψ_Q.

When $v_1, \ldots, v_n$ are indeterminates, then polynomial $T(g_i(v_1t, \ldots, v_mt), v_1t, \ldots, v_mt)$ is nonzero (see (36)). Write this polynomial as $\sum_j p_{ij}(v_1, \ldots, v_m)t^i$, where each $p_{ij} \in \overline{\mathrm{F}}[v_1, \ldots, v_m]$ is a degree l^2 polynomial. For each i pick a j_i such that p_{i,j_i} is nonzero. Then define Ψ_Q as

$$\Psi_Q(v_1, \ldots, v_m) = \prod_i p_{i,j_i}(v_1, \ldots, v_m). \tag{37}$$

Now consider any $(a_1, \ldots, a_m)$ such that $\Psi_Q(a_1, \ldots, a_m) \neq 0$. Then $T(g_i(a_1t, \ldots, a_mt), a_1t, \ldots, a_mt)$ is a nonzero polynomial in $\overline{\mathrm{F}}[t]$ for $i = 1, \ldots, k$. As already argued, $T_{a_1,\ldots,a_m}$ has no linear factor for such an $(a_1, \ldots, a_m)$.

□

Now we are ready to prove Theorem 34.

Proof:(of Theorem 34) Pick $(b_1, \ldots, b_m)$ randomly from F^m. With probability at least $1 - 2l^2/q$, the polynomial $Q(z, b_1, \ldots, b_m)$ is square-free and thus the polynomial $\Phi_Q = \Phi_{Q,b_1,\ldots,b_m}$ of Corollary 36 is non-zero. Pick $(a_1, \ldots, a_m)$ randomly from F^m. With a further probability $1 - l^3/q$, the polynomial Ψ_Q from Lemma 39 becomes nonzero and so $Q(z, a_1t + b_1, \ldots, a_mt + b_m)$ has no linear factor. Thus we have shown that with probability $(1 - 2l^2/q)(1 - l^3/q)) \geq 1 - 3l^3/q$ $Q(z, a_1t + b_1, \ldots, a_mt + b_m)$ has no linear factor. □

B Construction of Constant-Prover 1-Round systems

In this section we give the proof of Lemma 31 which shows how to reduce the answer sizes of MIP proof systems using three extra provers and a small penalty in the amount of randomness. Our proof is obtained by a simple modification of an "inner verifier" given by Arora et al. [5, Section 7]. The main difference is in the proof, where we employ the low-degree test given by Theorem 1 of this paper. A second difference is in the ingredients we use. While we use essentially the same starting point as [5], which was a probabilisitically checkable proof system given by Arora and Safra [6], we need to use slightly different properties of this verifier. In the following sections, we describe these differences and then conclude with a proof of Lemma 31.

B.1 A reconstruction procedure for polynomials

We start by describing an algebraic procedure that allows a verifier to reconstruct "many" values of a polynomial using only 3 queries. Our exposition closely follows the exposition in [3], Chapter 3. The only difference is a tremendous performance gain due to our new analysis of the low-degree test.

To set the context for how this algebraic procedure is used, we recall that many standard MIP verifiers rely on the fact that a satisfying assignment can be encoded as a degree d polynomial, for some appropriate d. The verifier expects the proof of satisfiability to contain such a polynomial, represented *by value*. This means that the proof contains some oracle $f : \mathrm{F}^m \to \mathrm{F}$ (the encoding is such that $|\mathrm{F}|^m$, the size of this oracle, is polynomial in the size of the assignment being encoded). Using the low degree test the the verifier checks that f has reasonable agreement with a degree d polynomial. Next, to check satisfiability, the verifier picks in some way (note: we're omitting many details here) k points $z_1, z_2, \ldots, z_k \in \mathrm{F}^m$ and then has to reconstruct the values of P at those points, where P is any polynomial that has significant agreement with f. Now we describe a procedure from [5] that allows the verifier to do this reconstruction, provided the proof contains additional information. The most important property of this procedure is that the verifier reads only 3 entries from the oracles provided to it, even though k might be pretty large (and not a constant).

Goals of the Reconstruction Procedure:

Given: A function $f : \mathrm{F}^m \to \mathrm{F}$, an integer d, a fraction $\delta, \rho > 0$, and k points $z_1, \ldots, z_k \in \mathrm{F}^m$.

Desired: One of the k-tuples $(P_1(z_1), \ldots, P_1(z_k)), \ldots$ or $(P_r(z_1), \ldots, P_r(z_k))$, where $P_1, \ldots, P_r$ are all the the polynomials that have agreement at least δ with f.

Auxiliary Information: Two tables T and T_1 in which each entry has $\mathrm{poly}(dk \log q)$ bits.

*Procedure's Properties***:** The procedure outputs either REJECT or a k-tuple.

- If f is a degree d polynomial, then there exist tables T, T_1 such that

$$\Pr[\text{procedure outputs } f(z_1), \ldots, f(z_k)] = 1.$$

- For every f, $\Pr[\text{procedure doesn't output one of the desired tuples}] \leq \rho$.

Procedure's Complexity: Uses $O(m \log q)$ random bits. Reads 1 entry from each of T and T_1 and one value of f.

Now we describe the procedure, which works correctly provided q is large enough and δ is small enough (compared to ρ). Recall that a degree-k curve is a set of points with a parametric representation like $\{(c_1(t), \ldots, c_m(t)) : t \in \mathrm{F}\}$, where each c_i is a degree k univariate polynomial. Note that the restriction of a degree d polynomial to this curve is a univariate polynomial of degree kd.

Below, we talk about a *random* degree k curve that passes through $z_1, \ldots, z_k$. We can pick such a curve by choosing a random point $y \in \mathrm{F}_m$ and identifying (using interpolation) m degree k univariate polynomials $c_1(t), \ldots, c_m(t)$ such that

$$\forall 1 \leq i \leq k \qquad (c_1(i), c_2(i), \ldots, c_m(i)) = z_i \tag{38}$$

$$(c_1(k+1), c_2(k+1), \ldots, c_m(k+1)) = y \tag{39}$$

Here we are using the integers $1, 2, \ldots, |\mathrm{F}|$ to also denote field elements. Note that by choosing the $k+1$th point of the curve randomly from F^m, we have ensured that the the last $|\mathrm{F}| - k$ points on the curve are randomly (though not independently) distributed in F^m. This will be important.

Now we describe the procedure. Note that part of the procedure (involving a random line) just consists in doing the low degree test at a point $C(a)$ on the curve C.

Description of Reconstruction Procedure:

INPUTS: Function $f : \mathrm{F}^m \to \mathrm{F}$, two oracles T_1, T_2, and k points $z_1, \ldots, z_k \in \mathrm{F}^m$. T_1 contains a sequence of univariate degree d polynomials, one for each line in F^m. Oracle T_2 contains, for each degree-$(k+1)$ curve in F^m that passes through $z_1, \ldots, z_k$, a univariate degree $(k+1)d$ polynomial.

PROCEDURE:

1. Randomly pick a degree k curve $C(t)$ in F^m whose first k points are $z_1, \ldots, z_k$. Pick a random $a \in \mathrm{F}$, and compute the point $C(a) \in \mathrm{F}^m$. Pick a random line l that passes through $C(a)$.
2. Read the value of f at $C(a)$. Read the polynomial given for curve C in oracle T_2; say it is $g_C(t)$. Read the polynomial given for line l in oracle T_1; say it is $h_l(t)$.
3. If $g_C(t)$ and $h_l(t)$ produce the value $f(C(a))$ at point $C(a)$, then output $(g_C(1), g_C(2), \ldots, g_C(k))$, the values of g_C at $1, 2, \ldots, k \in \mathrm{F}$. Otherwise output REJECT.

Complexity: The procedure runs in time $\mathrm{poly}(m + d + \log|\mathrm{F}| + k)$. Randomness is required only to generate $O(1)$ elements of F^m, so only $O(m \log|\mathrm{F}|)$ random bits are needed. Whenever we use this procedure, the function f is supposed to represent an assignment to n variables. The field size, the degree and the number of variables have been carefully chosen so that $|\mathrm{F}|^m = \mathrm{poly}(n)$. Thus the procedure requires $O(m \log|\mathrm{F}|) = O(\log n)$ random bits. Also, $d > m$, so the running time and the size of the oracle entries are $\mathrm{poly}(d + k)$.

Now we prove the correctness of the procedure. We are only interested in two cases. In the first case, the oracle-constructor is trying to help the verifier. Then it is clear that by just taking f to be a degree d polynomial and constructing oracles T_1, T_2 appropriately, it can make the verifier accept with probability 1. Now suppose the oracle constructor is malicious. Let c_0, c_1 be constants of the same name that appeared in Theorem 17. Let $P_1, \ldots P_r$ be all degree d polynomials that have agreement at least δ with f. We say that the procedure *makes a mistake* if it outputs a k-tuple that isn't one of $(P_1(z_1), \ldots, P_1(z_k))$, $(P_2(z_1), \ldots, P_2(z_k)), \ldots$ or $(P_r(z_1), \ldots, P_r(z_k))$.

Lemma 40 *Let c_0, c_1 be the constants given by Theorem 17. If $q \geq c_0(2(d+1)/\delta)^{c_1}$, then*

$$\Pr[\textit{procedure makes a mistake}] \leq \frac{2kd}{\delta q} + 2\delta + \frac{k}{q}.$$

Proof: Let us try to identify characteristics of any curve C, point $C(a)$ and line l that causes the procedure to make a mistake. It must be that (i) For each polynomial P_i there is some point among $z_1, \ldots, z_k$ at which P_i and g_C disagree (since otherwise the procedure would output $(P_i(z_1), \ldots, P_i(z_k))$, and thus not make a mistake). In other words, the univariate polynomial g_C differs from each of the restrictions $P_1|_C, \ldots P_r|_C$. (ii) f passes the low degree test using line l (iii) The curve polynomial g_C produces the value $f(C(a))$ at $C(a)$ (since otherwise the procedure would output REJECT).

We upperbound the probability of making a mistake as follows. Suppose curve C satisfies condition (i). Since two univariate degree kd polynomials can agree at at most kd points, we conclude that on such a curve, $1 - dkr/q$ fraction of $a \in \mathrm{F}$ are such that g_C does not agree with any of $P_1|_C, \ldots P_r|_C$ at $C(a)$. Thus conditions (ii) and (iii) become difficult to satisfy: if $g_C(a) = f(C(a))$ for "many" a — as required by condition (iii) — then on most such points f must disagree with all of $P_1, \ldots, P_r$, in which case the low degree test is very unlikely to succeed.

Now we formalize this. Let $S \subseteq \mathrm{F}^m$ be the set of points where f doesn't agree with any of $P_1, P_2, \ldots, P_r$. With each point $x \in \mathrm{F}^m$ let us associate a number ρ_x as follows: if $x \notin S$ then $\rho_x = 0$

and otherwise ρ_x is the success probability of the low degree test at x. Applying Theorem 17 with $\gamma = 2\delta$, we get

$$E_{x \in \mathrm{F}^m}[\rho_x] \leq 2\delta.$$

Now if C is a curve, we denote by $Y_C \in [0,1]$ the average of ρ_x among all points $x \in C$. When the test picks a random curve C, then the last $|\mathrm{F}| - k$ points of the curve are randomly distributed in F^m. Hence by linearity of expectations $E_C[Y_C] \leq 2\delta + \frac{k}{q}$. On any curve C that satisfies condition (i),

$$\Pr_{a,l}[\text{made a mistake on } C(a) \text{ using line } l] \leq \frac{dkr}{q} + E_{a \in \mathrm{F}}[\rho_{C(a)}] = \frac{dkr}{q} + Y_C.$$

Hence

$$\Pr_{C,a,l}[\text{made a mistake on } C(a) \text{ using line } l] \leq \frac{dkr}{q} + E_C[Y_C] \leq \frac{dkr}{q} + 2\delta + \frac{k}{q}.$$

Now the lemma follows by noticing that r, the number of polynomials with agreement at least δ with f, is at most $\frac{2}{\delta}$ by Proposition 4. □

B.2 The Composition

We now move on to showing how the reconstruction procedure of the previous section can be substituted into the inner verifier of [5, Section 7]. This verifier, in turn, builds upon a PCP verifier of [6, Proof of Theorem 3.5]. Below we recapitulate the properties of the PCP verifier of [6, Proof of Theorem 3.5]. This verifier verifies the satisfiability of a circuit C by the concatenation of p-strings given in encoded form, encoded by multivariate polynomial extensions. While the verifier is exactly the one given by [6], the properties extracted are slightly different. On the one hand, we wish to save on the number of queries, and we wish to have small error. On the other hand, we have available a strong low-degree test, so we only focus on the soundness condition of their test when the proferred proofs are low-degree functions. Later we will use our low-degree test and the reconstruction procedure of the previous section to reduce to this case, without losing too much in the error.

For any integer $p \geq 1$, the PCP verifier given by Arora and Safra behaves as follows: Given an input circuit C, a prime power q, and integers m and d satisfying $(d/m)^m \geq |C|$ and $q > d$, the verifier makes queries to $p+1$ oracles $f_1, \ldots, f_{p+1}$ with $f_i : \mathrm{F}^m \to \mathrm{F}$. (Strictly speaking some oracles may use more variables than the others, but we pad the number of variables to the maximum amount for notational simplicity.) The verifier computes projection functions $\rho_i : \mathrm{F}^m \to \mathrm{F}^m$, for $i \in \{1, \ldots, p\}$. It tosses a random string R of length $O(\log |C|)$ and makes at most $\mathrm{poly}(m, d, \log |C|)$ queries to the oracles $f_1, \ldots, f_{p+1}$, and then computes a Boolean verdict. This verdict satisfies the following properties:

Completeness: If $a_1 \cdots a_p$ satisfy C, and $f_1, \ldots, f_p$ are low-degree extensions of $a_1, \ldots, a_p$ respectively, then there exists an oracle f_{p+1} satisfying (1) For every $i \in \{1, \ldots, p\}$ and $x \in \mathrm{F}^m$ $f_i(x) = f_{p+1}(\rho_i(x))$. (2) The verifier accepts $f_1, \ldots, f_{p+1}$ with probability 1.

Soundness: If f_{p+1} is a polynomial of total degree at most d and f_i is given by $f_i(x) = f_{p+1}(\rho_i(x))$ for every i and x, and further f_i is a low-degree extension of a_i such that $C(a_1, \ldots, a_p) = 0$ then the verifier accepts with probability at most $\mathrm{poly}(m, d, \log |C|)/|\mathrm{F}|$.

Note: the soundness condition is incomparable with the one given in [6]. They analyze the soundness for arbitrary f_{p+1} while we only mention the case when f_{p+1} is a low-degree polynomial. However our soundness error is lower. To see how this error bound is achieved, note that the analysis in [6] just applies the Schwartz-Zippel lemma a polynomial number of times in the parameters d, m and $\log |C|$ and each application leads to an error of $\mathrm{poly}(m, d)/|\mathrm{F}|$.

Armed with this tool, we are ready to show the main lemma of this section, namely, that MIP with p provers and answer size a is contained in MIP with $p+3$ provers and answer size that is roughly $\mathrm{poly} \log a$.

Proof of Lemma 31: Let V_1 be the verifier that places NP in MIP$[p, r, a, e]$. Let V_2 be the p-prover PCP verifier for circuit satisfiability from [6] as described above, and let $e_2 = \text{poly}(m, d, \log |C|)/|\text{F}|$ denote the acceptance probability of the verifier V_2 on inputs that don't satisfy the circuit C.

As in [5], we convert the verifier V_2 into a verifier for a constant prover proof system. We then use the technique of recursive proof checking [6] to combine the verifiers V_1 and the converted version of V_2 to get the required verifier. We directly describe the combined proof system.

Let $\pi_1, \ldots, \pi_p$ be the provers for V_1. The composed verifier, denoted V, works with $p+3$ provers $\pi'_1, \ldots, \pi'_p, \pi_f, \pi_{\text{lines}}, \pi_{\text{curves}}$. For $i \in \{1, \ldots, p\}$, the prover π'_i takes as queries (q_i, x) where q_i ranges over queries of V_1 to π_i and $x \in \text{F}^m$. π'_i is supposed to encode the response a_i of π_p into $f_{i,q_i} : \text{F}^m \to \text{F}$ - a degree d polynomial extension of a_i over F^m and respond with $f_{i,q_i}(x)$. A query to π_f is a pair (R, x) where R ranges over random strings used by V_1 and $x \in \text{F}^m$. π_f responds with $f_R(x)$ where $f_R : \text{F}^m \to \text{F}$ is supposed to be the degree d extension of the concatenation of $a_1, \ldots, a_p$ and the proof that $a_1, \ldots, a_p$ satisfy $\mathcal{C}$ (where $a_1, \ldots, a_p$ are the responses of $\pi_1, \ldots, \pi_p$ for queries raised by V_1 on random string R). π_{lines} takes as query (R, x, y) where $R \in \{0,1\}^r$ and $x, y \in \text{F}^m$ and reponds with (the coefficients of) a degree d univariate polynomial. This is supposed to be the restriction of f_R to the line $\{x + ty | t \in \text{F}\}$. Lastly the prover π_{curves} takes as queries (R, k, C), where $R \in \{0,1\}^r$, $k \in \mathcal{Z}^+$ and C is a "degree k curve" through F^m, i.e., $C = (C_1, \ldots, C_m)$ where $C_j : \text{F} \to \text{F}$ is a degree k univariate polynomial. π_{curves} responds with a degree kd univariate polynomial which is supposed to be f_R restricted to the curve C.

We now describe the verifier V.

1. Pick random string $R \in \{0,1\}^r$ and generate questions $q_1, \ldots, q_p$ and circuit $\mathcal{C}$ of size $c = a^{O(1)}$ according to the verifier V_1.
2. Pick random points $x_1, \ldots, x_p, x, y \in \text{F}^m$ and $t_1 \in \text{F} - \{1, \ldots, a_2 + p\}$.
3. Pick random string $R_2 \in \{0,1\}^{r_2}$ and generate queries $y_1, \ldots, y_{a_2} \in \text{F}^m$ according to V_2.
4. Generate curve C of degree $k \overset{\text{def}}{=} a_2 + p$ such that $C(1) = y_1, \ldots, C(a_2) = y_{a_2}$ and $C(a_2+1) = x_1, \ldots, C(k) = x_p$, and $C(t_1) = x + t_2 y$. (happens then V accepts with asking any questions.)
5. Sends questions (q_i, x_i) to π'_i, for $i \in \{1, \ldots, k\}$. Sends question $(R, x + t_2 y)$ to π_f, (R, x, y) to π_{lines} and (R, k, C) to π_{curves}.
6. Denote the responses of the provers $\pi'_1, \ldots, \pi'_p, \pi_f, \pi_{\text{lines}}, \pi_{\text{curves}}$ by $\alpha_1, \ldots, \alpha_p, \alpha, h$ and g respectively. V accepts if the following conditions hold:
   (a) For every i, $\alpha_i = g(a_2 + i)$.
   (b) $\alpha = g(t_1)$.
   (c) $\alpha = h(t_2)$.
   (d) If V_2 accepts $(g(1), \ldots, g(a_2))$ on input $\mathcal{C}$ and random string R_2.

Based on the description of the provers above, it is clear that if V_1 accepts with probability 1, then there exist provers $\pi'_1, \ldots, \pi'_p, \pi_f, \pi_{\text{lines}}, \pi_{\text{curves}}$ such that V accepts with probability 1. It also clear that V uses $p+3$ provers, tosses $r + r_2 + (p+2)m \log |\text{F}| = r + O(m \log |\text{F}|)$ coins, and gets responses of length at most $kd \log |\text{F}| = (\text{poly} \log a) d \log |\text{F}|$ bits.

To conclude, we need to show that the acceptance probability of V is at most $O(e^{\frac{1}{2p+2}})$. Fix provers $\pi'_1, \ldots, \pi'_p, \pi_f, \pi_{\text{lines}}, \pi_{\text{curves}}$. Let ρ be a parameter to be fixed later. We first construct provers $\pi_1, \ldots, \pi_p$ for V_1 based on $\pi'_1, \ldots, \pi'_p$ as follows: For every i and q_i, let $f_i^{(1)}, \ldots, f_i^{(l_i)}$ be all the degree d polynomials with agreement ρ^2 with f_{i,q_i} and let $a_i^{(1)}, \ldots, a_i^{(l_i)}$ be their corresponding decodings. If $l_i > 0$, π_i picks a random $j \in \{1, \ldots, l_i\}$ and picks $a_i^{(j)}$ as its response to q_i. Notice that $0 \le l_i \le \frac{2}{\rho^2}$. We now show that for an appropriate choise of ρ, the expected acceptance probability of the verifier V_1 on provers $\pi_1, \ldots, \pi_p$ is at least $O(e^{\frac{1}{2p+2}})$.

Fix a random string R. This fixes questions $q_1, \ldots, q_p$. Let $f_1, \ldots, f_p : \text{F}^m \to \text{F}$ denote the functions $\pi'_1(q_1, \cdot), \ldots, \pi'_p(q_p, \cdot)$. of the provers $\pi'_1, \ldots, \pi'_p$ respectively. Let f denote the function

$\pi_f(R,\cdot)$. For a random choice of $R_2, x_1, \ldots, x_p, x, y, t_1$ and t_2, let g denote the response of π_{curves} and h denote the response of π_{lines}. Let the expected acceptance probability of V_1 on $\pi_1, \ldots, \pi_p$ on this string be denoted e_R. Let $f^{(1)}, \ldots, f^{(l)}$ be polynomials with agreement $\rho/2$ to f. Let $f_i^{(1)}, \ldots, f_i^{(l_i)}$ be polynomials with agreement ρ^2 to f_i.

Case: $f(x+t_1y) \neq f^{(j)}(x+t_1y)$ for any j: In this case the low degree test accepts with probability at most ρ, provided $q \geq c_0((d+1)/\rho)^{c_1}$, where c_0, c_1 are as given by Theorem 17.

Case: $f(x+t_1y) = f^{(j)}(x+t_1y)$ for some j: We argue the remaining cases by fixing a $j \in \{1, \ldots, l\}$ and assuming $f = f^{(j)}$. We now consider the remaining subcases:

Case: C not consistent with $f^{(j)}$: In this case the acceptance probability is upper bounded by the probability that $C(t_1) = f^{(j)}(x+t_1y)$ which is at most $\frac{kd}{|\mathrm{F}|-k}$.

Case: C consistent with $f^{(j)}$ and there exists an i such that $f^{(j)}$ not an extension of any $f_i^{(j_i)}$: The acceptance probability is bounded by the probability that $f^{(j)}(x_i) = f_i(x_i)$ which is at most ρ^2.

Case: C consistent with $f^{(j)}$ and $f^{(j)}$ does not satisfy $\mathcal{C}$: The probability of acceptance by V_2 is upper bounded by e_2.

Case: C consistent with $f^{(j)}$; $\forall i$, $\exists j_i$ s.t. $f^{(j)}$ extends $f_i^{(j_i)}$; and $f^{(j)}$ satisfies $\mathcal{C}$: In this case $\pi_1, \ldots, \pi_p$ cause V_1 to accept with probability at least $(\rho^2)^p$. Inverting this probability we find that this case occurs with probability at most $(\frac{2}{\rho^2})^p e_R$.

Thus summing up over all j we find that the acceptance probability in this case is at most $(\frac{4}{\rho})\left(\frac{kd}{|\mathrm{F}|-k} + \rho^2 + e_2 + (\frac{2}{\rho^2})^p e_R\right)$.

Thus summing up over all cases and taking expectations over all R we find that the expected acceptance probability of the verifier V is bounded by $\rho + \frac{4}{\rho}\left(\frac{kd}{|\mathrm{F}|-k} + \rho^2 + e_2 + e(\frac{2}{\rho^2})^p\right)$. Assuming $q \geq \frac{kd}{\rho^2} + k$, and $e_2 \leq \rho$ (which again translates into a condition of the form $q > \mathrm{poly}(m, d, \log|C|)/\rho^2$), this quantity simplifies to $13\rho + e(\frac{2}{\rho^2})^p$. We pick $\rho = e^{\frac{1}{2p+2}}$ and this bounds the error probability of V by $O(e^{\frac{1}{2p+2}})$. Notice that under this setting of ρ, the conditions on q translate to a condition of the form $q > \mathrm{poly}(m, d, \log|C|, \frac{1}{e})$ as promised. □